14TH International Joint Symposium
on Artificial Intelligence
and Natural Language Processing

iSAI-NLP 2019

PROCEEDINGS

October 30 - November 1, 2019
Chiang Mai, Thailand
https://isai-nlp2019.aiat.or.th

The Processding

The 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019)

About this Publication

Title:	The Processding of the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019)
Editor-in-chief:	Thanaruk Theeramunkong
Editors:	Hashimoto Kiyota, Thepchai Supnithi, Mahasak Ketcham, Narumol Chumuang, Pokpong Songmuang, Supakrit Sukjarern, Rachada Kongkachandra, Juntima Donjuntai Sumate Lipirodjanapong, Amonrada Rongtong, Jureebhorn Kaewjunda
Production Assistants:	Thodsaporn Chay-intr, Benjaphan Sommana, Uraiwan Buatoom, Rachasak Somyanonthanakul
Cover Designer:	Jiragorn Chalerndit, Nawarat Wittayakhom
Organizers:	Artificial Intelligence Association of Thailand (AIAT), Thailand Muban Chom Bueng Rajabhat University (MCRU), Thailand Mahidol University (MU), Thailand Sirindhorn International Institute of Technology, Thammasat University (SIIT, TU), Thailand National Electronics and Computer Technology Center (NECTEC), Thailand Rungsit University Chiang Mai University
Publisher:	Artificial Intelligence Association of Thailand (AIAT), Thailand
Date Published:	November 2019

Welcome Message from the iSAI-NLP 2019 General Chairs

Welcome all presenters, participants, and contributors to the Fourteenth International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019) held at Chiang Mai, Thailand from 30

October 2019 to 1 November 2019. We would like to show our sincere gratitude to your great corporation and contribution to iSAI-NLP 2019.

iSAI-NLP, formerly called SNLP until 2016, has survived for 26 years in this fast developing research area of Natural Language Processing and Artificial Intelligence in general. The First SNLP was held in Bangkok, hosted by Chulalongkorn University, in 1993, aiming to promote and improve Thai, though not limited to it, research on natural language processing. Since then, subsequent symposia were held by Kasetsart University in Bangkok (1995), Asian Institute of Technology in Phuket (1997), King Mongkut's University of Technology Thonburi in Chiang Mai (2000),SIIT,Thammasat University in Hua Hin(2002), Chulalongkorn University in Chiang Rai(2005),Kasetsart University in Pattaya(2007), Dhurakij Pundit University in Bangkok(2009), King Mongkut's Institute of Technology Ladkrabang in Bangkok(2011), SIIT in Phuket(2013), Thammasat University in Ayutthaya(2016), King Mongkut's University of Technology Thonburi and Rangsit University in Hua Hin(2017), and Mahidol University in Pattaya(2018). Following this long tradition, iSAI-NLP2019 is hosted by Muban Chombueng Rajabhat University, the Center of Excellence in Community Health Informatics, Chiang Mai University, Sirindhorn International Institute of Technology, Thammasat University, Thammasat University, Mahidol University, National Electronics

and Computer Technology Center (NECTEC), and Artificial Intelligence Association of Thailand (AIAT).

According to the recent flourishment of a variety of research on AI and NLP, iSAI-NLP 2019 holds five tracks: Natural Language Processing; Robotics, IoT and Embedded System; Data Analytics and Machine Learning; Signal, Image and Speech Processing; and Smart Industrial Technologies, together with the co-located International Exhibition of Inventions of Thailand (IEIT2019). Thus iSAI-NLP 2019 covers all important topics in AI and NLP.

We would like to express our sincere appreciation to our sponsors and supporters for their valuable supports, particularly Phetchaburi Rajabhat University, Electricity Generation Authority of Thailand, iApp Technology, and IEEE Thailand Section; three distinguished keynote speakers for their invaluable talks, Addoc. Prof. Dr. CharturongTantibundhit, Dr. Ryota Yamanaka, and Prof. Dr. Hiroaki Oagata; all organizing committee members for their hard work; and most importantly all presenters and participants for their research presentations and discussions.

We hope that iSAI-NLP 2019 will the place again to know the frontier of AI and NLP research, to meet new acquaintances, to reunite old friends, and to enjoy active research discussion in a famous cultural city of Chiang Mai, Thailand. See you all there.

October 2019

The iSAI-NLP 2019 General Co-Chairs

ThanarukTheeramunkong

(SIIT, Thammasat University, Thailand)

ThepchaiSupnithi

(National Electronics and Computer Technology Center, Thailand)

Kiyota Hashimoto

(Prince of Songkla University, Thailand)

Welcome Message from the iSAI-NLP 2019 Program Chairs

On behalf of the Technical Program Committee, we are pleased to welcome you to the 2019 edition of International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). iSAI-NLP 2019 is the fourteenth conference in the series of the international symposium on NLP (SNLP) started in 1993. This year, it is held in Chiang Mai, Thailand from October 30th to November 1st in conjunction with the 1st International Exhibition of Inventions of Thailand 2019 (IEIT 2019). We hope the participants of both events, iSAI-NLP and IEIT, to have a great time for intellectual exchange in Chiang Mai.

iSAI-NLP 2019 has 5 tracks namely NLP; Robotic, IoT and Embedded System; Data Analytic and Machine Learning; Signal, Image and Speech Processing; Smart Industrial Technologies. We have received 79 submissions from various countries including Thailand, India, The Philippines, USA, China, Bangladesh, and Japan. All the submissions are rigorously reviewed by at least three anonymous reviewers, and finally we have accepted 69 submissions; 49 regular papers and 20 short papers. This builds up to 88% acceptance rate. To this end, we would like to express our gratitude to all reviewers that spent their valuable time reviewing and evaluating the submitted papers.

This year, we are pleased to have three interesting keynote speakers namely Associate Professor CharturongTantibundhit, Dr. Ryota Yamanaka and Professor Hiroaki Ogata. We are honored by their presence. We would also like to express our sincere gratitude to our honorary co-chairs Professor Vilas Wuwongse, Professor Yoshinori Sagisaka, Assistant Professor ChairitSiladech and Professor Nicolai Petkov. The organization of this conference would be different

without their invaluable advice and support. Furthermore, we would like to show our appreciation to the iSAI-NLP team that worked hard to organize this conference. This includes our general co-chairs, Professor ThanarukTheeramunkong, Dr. ThepchaiSupnithi, and Dr. Kiyota Hashimoto; publication co-chairs, ThodsapornChay-intr, JiragornChalermdit, and NawaratWittayakhom; technical programs chairs of all tracks; the organizing committee members, international committee members, financial co-chairs, and secretary generals. Moreover, we would like to thank our hosts, co-hosts, and local-host: Artificial Intelligence Association of Thailand (AIAT), MubanChombuengRajabhat University (MCRU, Thailand), Thammasat University (TU, Thailand), Sirindhorn International Institute of Technology (SIIT, TU, Thailand), Mahidol University (MU, Thailand), National Electronics and Computer Technology Center (NECTEC, Thailand), and Center of Excellence in Community Health Informatics, Chiang Mai University (CMU, Thailand). In addition, we would like to thank all sponsors for their generous support.

Last but not least, we would like to express our greatest gratitude to all the authors who have submitted their valuable works, to those who join the conference and to all other iSAI-NLP participants. Thank you for your effort, for your time, and kindness.

October 2019

The iSAI-NLP 2019 Program Co-Chairs

SanparithMarukatat

(National Electronics and Computer Technology Center, Thailand)

ItsuoKumazawa

(Tokyo Institute of Technology, Japan)

Welcome Message from the iSAI-NLP 2019 Organizing Committees

It is our great honor that the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019) at the Kantary Hills Hotel, Chiang Mai, Thailand during October 30-November 1, 2019. The iSAI-NLP 2019 will trigger an opportunity for professors, researchers, practitioners, junior researchers, and students to exchange ideas, methods, insights, and current research progresses. Hosted by MubanChombuengRajabhat University with great support from Artificial Intelligence Association of Thailand (AIAT).

The response to our initial call for paper was overwhelming with more than 80 papers submitted from 12 countries. After rigorous review of all the papers by at least three expert reviewers each, 48 papers were accepted for presentation with an acceptance ratio of about 60 percent. We have organized the program into a number of technical sessions, keynote presentations, and tutorials. The technical sessions cover a wide range of topics related to natural language processing, robotics, IoTand embedded system, data analytics and machine learning, signal, image and speech processing and smart industrial technologies.

The conference will be preceded by four major tutorials on October 30. The first tutorial will be given by Prof.Dr. Hiroaki Ogata, on "Connecting formal and informal learning through learning evidence and analytics framework." while the second tutorial will be given by Dr. Ryota Yamanaka, from Oracle Corporation, Thailand, on "DNA sequence analysis and its reproducible platform, Network analysis and semantic technologies".The Third is given by Prof.Dr.Itsuo Kamazawa,

on –Evaluation of deep learning techniques for detecting lesion areas in MRI and X ray images,‖and the last will be given by Assistant ProfessorDr.CharturongTantibundhit, on "Artificial Intelligence for Medical Screening: Research and Innovation in Low-resource Settings".

As chair of the organizing committee, we would like to express our sincere appreciation to all committeemembers, presenters, reviewers, organizing teams, including secretariesand staffs, for their dedication and hard work behind the scene tomake this a truly successful international conference. We sincerely hopethat all of you enjoy this remarkable event. We look forward to seeingand discussing with you at the iSAI-NLP 2019.

October 2019

The iSAI-NLP 2019 Organizing Co-Chairs

ChairitSiladesh

(MubanChombuengRajabhat University)

ThepchaiSupnithi

(National Electronics and Computer Technology Center, Thailand)

iSAI-NLP 2019 Organization Committees

Honorary Co-Chair

Vilas Wuwongse
SakonNakhonRajabhat University, Thailand

Yoshinori Sagisaka
Waseda University, Japan

Chairit Siladech
MubanChombuengRajabhat University, Thailand

Nicolai Petkov
University of Groningen, Netherlands

Conference-Chair

Thanaruk Theeramunkong
SIIT, Thammasat University, Thailand

Thepchai Supnithi
National Electronics and Computer Technology Center, Thailand

Kiyota Hashimoto
Prince of Songkla University, Thailand

Program Co-Chair

Sanparith Marukatat
National Electronics and Computer Technology Center, Thailand

Itsuo Kumazawa
Tokyo Institute of Technology, Japan

Martin Hobelsberger
the Munich University of Applied Sciences, Germany

Technical Program Chairs

Track 1

Prachya Boonkwan
National Electronics and Computer Technology Center, Thailand

Aw Ai Ti
Institute for Infocomm Research, Singapore

Rachada Kongkachandra
Thammasat University, Thailand

Ye Kyaw Thu
National Electronics and Computer Technology Center, Thailand

Track 2

Mahasak Ketcham
King Mongkut's University of Technology North Bangkok, Thailand

Michael Pecht
University of Maryland, United States of America

Thaweesak Yingthawornsuk
King Mongkut's University of Technology Thonburi, Thailand

Track 3

Worawut Yimyam
PhetchaburiRajabhat University, Thailand

Natsuda Kaothanthong
SIIT, Thammasat University, Thailand

OlarikSurinta
Mahasarakham University, Thailand

Track 4

Narit Hnoohom
Mahidol University, Thailand

Anuchit Jitpattanakul
King Mongkut 's University of Technology North Bangkok, Thailand

Ngoc Hong Tran
University College Dublin, Ireland

Sakorn Mekruksavanich
University of Phayao, Thailand

Thach-Thao Nguyen Duong
University of Burgundy, France

Track 5

Narumol Chumuang
MubanChombuengRajabhat University, Thailand

Sumate Lipirodjanapong
MubanChombuengRajabhat University, Thailand

Supakrit Sukjarern
MubanChombuengRajabhat University, Thailand

Advisory Committee

Aijun An
York University, Canada

Organizing Committee

Chutima Beokhaimook
Rangsit University, Thailand

Choochart Haruechaiyasak
National Electronics and Computer Technology Center, Thailand

Konlakorn Wongpatikaseree
Mahidol University, Thailand

Kritchana Wongrat
PhetchaburiRajabhat University, Thailand

Kritsada Sriphaew
Rangsit University, Thailand

Pokpong Songmuang
Thammasat University, Thailand

Nattapong Tongtep
Prince of Songkla University, Thailand

Sasiporn Usanavasin
SIIT, Thammasat University, Thailand

Sumeth Yuenyong
Mahidol University, Thailand

Thaweesak Yingthawornsuk
King Mongkut's University of Technology Thonburi, Thailand

Wimol San-Um
Thai-Nichi University, Thailand

Wiwit Suksangaram
PhetchaburiRajabhat University, Thailand

International Committee

Atsuo Yoshikata
Japan Advanced Institute of Science and Technology, Japan

Colin De La Higuera
University of Nantes, France

Cristina Tirnauca
University of Cantabria, Spain

Philippe Lenca
IMT Atlantique, France

Trung-Hieu Huynh
Vietnamese-German University, Vietnam

The-Bao Pham
Vietnam National University Ho Chi Minh City, Vietnam

Publication Co-Chair

Thodsaporn Chay-intr
SIIT, Thammasat University, Thailand

Jiragorn Chalermdit
MubanChombuengRajabhat University, Thailand

Nawarat Wittayakhom
MubanChombuengRajabhat University, Thailand

Secretary Generals

Narumol Chumuang
MubanChombuengRajabhat University, Thailand

Choermath Hongakkaraphan
Artificial Intelligence Association of Thailand, Thailand

Local Organizing Committee

Ekkarat Boonchaimg
Chiang Mai University, Thailand

Rachasak Somyanontanakul
Rangsit University, Thailand

Rungphailin Sukriwanat
MubanChombuengRajabhat University, Thailand

Thanaree Numklin
MubanChombuengRajabhat University, Thailand

Uraiwan Buatoom
Burapha University, Thailand

Vorapon Luantangsrisuk
Thammasat University, Thailand

Financial Co-Chair

Choermath Hongakkaraphan
Artificial Intelligence Association of Thailand, Thailand

Wirat Chinnan
Artificial Intelligence Association of Thailand, Thailand

Webmaster

Suchathit Boonnag
Artificial Intelligence Association of Thailand, Thailand

Phalakron Nilkhet
Artificial Intelligence
Association of Thailand,
Thailand

Host and Co-Host

Artificial Intelligence Association of Thailand AIAT, Thailand

MubanChombueng RajabhatUniversity
MCRU, Thailand

Center of Excellence in Community Health Informatics, Chiang Mai University
CMU, Thailand

Sirindhorn International Institute of Technology
SIIT, Thammasat University, Thailand

Thammasat University
TU, Thailand

Mahidol University
MU, Thailand

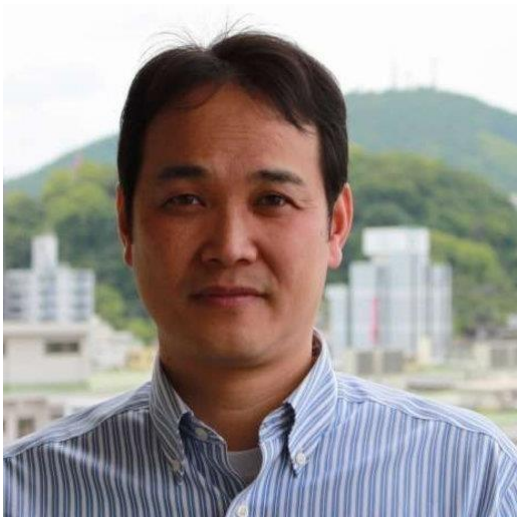
National Electronics and Computer Technology Center
NECTEC, Thailand

Supporters

IEEE Thailand Section
Thailand

Keynote/Invited/Guest Speakers

iSAI-NLP 2019keynote Speaker



Professor. Dr. Hiroaki Ogata

Professor, Academic Center for Computing and Media Studies, and the Graduate School of Informatics, Kyoto University

Title of the talk : Connecting formal and informal learning through learning evidence and analytics framework

Abstract:

The multi-disciplinary research approach of Learning Analytics (LA) has been providing methods to understand learning logs collected during varied teaching-learning activities and potentially enrich such experiences. However, LA is mainly focusing on formal learning in classrooms. This talk will explain how technology can help to entwine formal and informal learning and to extract evidence of effective teaching-learning practices by applying LA and developing novel techniques. It focuses discussions on realizing a technology enhanced evidence-based education and learning (TEEL) system. This talk will propose the Learning Evidence Analytics Framework (LEAF) and draw a research roadmap of an educational big data driven evidence-

based education system. Teachers can refine their instructional practices, learners can enhance learning experiences and researchers can study the dynamics of the teaching-learning process with it. While LA platforms gather and analysis the data, there is the lack of a specific design framework to capture the technology enhanced teaching-learning practices. Finally, this talk will present the research challenges to smart evidence-based education.

Research Interests : Learning Analytics, Educational data science, evidence-driven education

Bibliography :

Hiroaki Ogata is a Professor at the Academic Center for Computing and Media Studies, and the Graduate School of Informatics, Kyoto University, Japan. His research includes Computer Supported Ubiquitous and Mobile Learning, CSCL, CSCW, CALL, and Learning Analytics. He has published more than 300 peer-reviewed papers including SSCI Journals and international conferences. He has received several Best Paper Award and gave keynote lectures in several conferences.

Contact Information:

Prof. Dr. Hiroaki Ogata

Academic Center for Computing and Media Studies,

Graduate School of Informatics, Kyoto University

Phone: +81-75-753-9052

Fax: +81-75-753-9053

E-mail: hiroaki.ogata@gmail.com

iSAI-NLP 2019 keynote Speaker

Ryota Yamanaka
Senior Solutions Consultant (Big Data & Analytics),
Oracle Corporation Thailand

Title of the talk : Graph Database for AI

Abstract:

Graph database is one of the emerging database management systems, whose data model is based on mathematical graph, consists of nodes and edges. Because of its capability to represent complex data, graph database is expected to become a new data management platform in AI fields. In machine learning, since graph data is a rich data source to train predictive models, new methods are recently proposed to take graph data as input of machine learning. Graph database provides graph data for such algorithms as well as generates graph-based features for conventional algorithms. In semantic web, more information has become available in the form of knowledge graphs. Graph database is suitable to populate such graph data keeping its semantic information, and it enables us to run high-response queries

as well as graph-based algorithms to analyze the data. In this talk, I will present the introduction of graph database and its usage in AI systems. I will also discuss its actual use cases from industrial perspective.

Keywords: Graph Database, Machine Learning, Semantic Web

Research Interests : DNA sequence analysis and its reproducible platform, Network analysis and semantic technologies.

Bibliography:

Dr. Ryota Yamanaka is a senior solutions consultant in big data & analytics at Oracle Corporation Thailand. He received his bachelor's degree in computer science from Tokyo Institute of Technology (2007), his master's degree in bioinformatics from King's College London (2011), and his PhD in genome science from The University of Tokyo (2015). He also worked as a database consultant for Oracle Corporation Japan, during 2007 - 2010. His interest includes bioinformatics, database, and semantic web.

Contact Information:

Dr. Ryota Yamanaka
Solutions Consultant, Big Data and Analytics
Oracle Corporation Thailand
E-mail: ryota.yamanaka@oracle.com
Phone: +66-6-3373-5925

iSAI-NLP 2019 keynote Speaker

Professor. Dr. Itsuo Kamazawa
Laboratory for Future Interdisciplinary Research of Science and Technology,
Institute of Innovative Research Tokyo Institute of Technology

Title of the talk : Advanced in Human-Computer Interface

Abstract:

This talk is related to technologies for human interface with a few demonstrations. Thesetechnologies are applied to gaming including serious games, virtual reality and any situation where human and machine interaction is needed. Talks and demonstration with collaborating students in the ultrafast image sensing to minimize latency in generating feedback for virtual reality system. Compared to the original method, the innovative device for image sensing that combines the ultrafast optical image sensor that is used in the computer mouse and the Leap Motion to detect the very quick motion of the human hand and the fingers. This technique is essential to improve the reality and usability of virtual reality applications. The talk also includes brain machine interface to control machines or

systems by brain activities, and haptic technologies with haptic sensing and feedback devices to provide realistic information. Several demos are provided to facilitate fruitful discussions on multimodal human-computer interfaces in gaming context.

Keywords: Artificial Neural Networks, Human Interface, Pattern Recognition, and Image Processing

Bibliography:

ItsuoKumazawa received Bachelor of Engineering in Electrical and Electronic Engineering from Tokyo Institute of Technology (TIT) in 1981, Master of Engineering, and Doctor of Engineering in Computer Science from Tokyo Institute of Technology (TIT) in 1983, and 1986, respectively. In 1986, he started his academic career as Assistant Professor in TIT, where he became Associate Professor in 1990. Currently, he is a Professor in the Institute of Innovative Research at TIT. Dr. Kumazawa has published research papers in the fields of Artificial Neural Networks, Human Interface, Pattern Recognition, and Image Processing. He received grants from JSPS, JST and a number of private funds. He is a member of IEICE, IPSJ, ITE and IEEE and received awards from these academic societies such as the “best demo award” in the IEEE virtual reality conference in 2013.

Contact Information:

Prof. Dr. ItsuoKamazawa

Imaging Science and Engineering Research Center

Information Processing Department

Tokyo Institute of Technology

E-mail: kumazawa.i.aa@m.titech.ac.jp

iSAI-NLP 2019 keynote Speaker

Charturong Tantibundhit

Associate Professor (Department of Electrical and Computer Engineering), Thammasat University

Title of the talk : Artificial Intelligence for Medical Screening: Research and Innovation in Low-resource Settings

Abstract:

Artificial Intelligence (AI) has been applied to a lot of medical applications especially for disease screening and diagnosis. One objective is to support medical staffs especially in developing countries, where medical experts and resources are very limited. As a result, patients can have medical screening and can receive medical treatment on time reducing disability and loss of life. Our research group, emphasized in AI in medicine, have collaborated on interdisciplinary research and development of medical innovations using resources available in Thailand. Our ultimate goal is to facilitate physicians, public health officers, and people in medical screening

and diagnosis, focusing on the major diseases, e.g., stroke, Alzheimer's, learning disability, diabetic retinopathy, aged-macular degeneration, glaucoma, cytomegalovirus retinitis, cervical cancer, skin cancer, lung cancer, and tuberculosis that are affecting majority people around the world. Based on our continuous dedication for more than 10 years, we have developed a lot of innovative medical products with world class quality. These innovations have high impacts resulting in the better quality of life of Thai people. Our research work and innovations have been perceived as one of the best research groups in Thailand as shown by awards received nationally and internationally. Moreover, our research group is only one that won the Grand Prize in International Exhibition of Inventions of Geneva, Switzerland recognized as the world's largest exhibition of inventions. Finally, our research work and innovations have been published in the world leading international journals.

Research Interests: Speech Enhancement, Tonal-speech Perception, Signal 6Processing for Cochlear Implants, Pattern Recognition and Machine Learning

Bibliography :

CHARTURONG TANTIBUNDHIT received the B.E. degree in electrical engineering from Kasetsart University, Bangkok, Thailand, in 1996, and the M.S. degree in information science and Ph.D. degree in electrical engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 2001 and 2006, respectively. Since 2006, he has been with Thammasat University, Thailand, where he is currently an Associate Professor with the Department of Electrical and Computer Engineering and the Head of the Speech and Language Technology Cluster, Center of Excellence in Intelligence Informatics, Speech and Language Technology, and Service Innovation. From 2007 to 2008, he was a Post-Doctoral Researcher with the Signal Processing and

Speech Communication Laboratory, Graz University of Technology, Graz, Austria. He was an IEEE ICASSP Student Paper Contest Winner in 2006. He led a team to win the Grand Prix of the 45th International Exhibition of Inventions of Geneva in 2017. His research interests include handcrafted machine learning and deep learning in medicine, biomedical signal processing, and speech processing.

Contact Information :

Associate Prof. Dr. CharturongTantibundhit

Department of Electrical and Computer Engineering,

Faculty of Engineering, Thammasat University Rangsit Campus, KhlongLuang, PathumThani 12120

Phone: +66-2-564-3213

Fax: +66-2-564-3010

E-mail: tchartur@engr.tu.ac.th

conference Overall Schedule

iSAI-NLP-2019

Wednesday, October 30, 2019

Time	DoiSuthep 1	DoiSuthep 2	Doi Nua
08:00-09:00	Registration		
09:00-10:20	**Tutorial 1 (NLP)** When 1 + 1 > 2: Joint Neural NLP Models Demystified Prachya Boonkwan and Ye Kyaw Thu (NECTEC, Thailand)	**R3-1** Data Analytics and Machine Learning (5 papers : R3-07, R3-15,R3-16, R3-17,R3-20) (Chair: Worawut Yimyam, Phetchaburi Rajabhat University, Thailand)	
10:20-10:40	Break		
10:40-12:00	**Tutorial 1 (NLP)** When 1 + 1 > 2: Joint Neural NLP Models Demystified Prachya Boonkwan and Ye Kyaw Thu (NECTEC, Thailand)	**R3-2** Data Analytics and Machine Learning (5 papers : R3-24, R3-32,R3-36, R3-41,R3-42,) (Chair: Itsuo Kamazawa, Tokyo Institute of Technology, Japan)	
12:00-13:30	Lunch		
13:30-15:10	**WS-1 (13:00-16:00) NLP Workshop** (Chair: Ye Kyaw Thu, Thepchai Supnithi, NECTEC, Thailand)	**R4-1** Signal, Image and Speech Processing (6 papers : R4-64,R4-65,R4-66, R4-68,R4-71,R4-79) (Chair: Narit Hnoohom, Mahidol University, Thailand)	IEIT Committee Meeting
15:10-15:30	Break		
15:30-18:00	**Tutorial 2 (16:00-17:30)** How to improve your research: A writing-as-research method *Kiyota Hashimoto (Prince of Songkla University, Thailand)*	**R3-3** Data Analytics and Machine Learning (7 papers : R3-45, R3-51,R3-53,R3-54, R3-55,R3-58,R3-61) (Chair: Olarik Surinta, Mahasarakham University, Thailand)	IEIT Committee Meeting
18:00-21:00	iSAI-NLP 2019 + AIAT Committee Dinner		

Thursday, October 31, 2019

Time	DoiSuthep 1+2		
08:00-09:00	Registration		
09:00-09:20	**Opening Ceremony** (AIAT, MCRU, CMU) (MC: Rachada Kongkachan & Prachya Boonkwan) (08:40-09:00) (Photo Session: 09:20)		
09:20-10:10	**Keynote I** Artificial Intelligence for Medical Screening: Research and Innovation in Low-resource Settings Charturong Tantibundhit (Thammasat University) (Chair: Pokpong Songmuang)		
10:10-10:40	Break		
Room	DoiSuthep 1	DoiSuthep 2	Foyer
10:40-12:00	**R1-1** Natural Language Processing (5 papers : R1-01, R1-04,R1-08, R1-14, R1-19) (Chair: Mikifumi Shikida, KUT, Japan)	**R4-2** Signal, Image and Speech Processing (5 papers : R4-05, R4-11,R4-12, R4-38,R4-59) (Chair: Sakorn Mekruksavanich, University of Phayao, Thailand)	The 1st International Exhibition of Inventions of Thailand 2019 (IEIT 2019) (Chair: Patiyuth Pramkeaw, KMUTT, Thailand)
12:00-13:30	Lunch		
Room	DoiSuthep 2		
13:30-14:00	**Keynote II** Connecting formal and informal learning through learning evidence and analytics framework Hiroaki Ogata (Kyoto University) (Chair: Thepchai Supnithi, NECTEC, Thailand)		
Room	DoiSuthep 1	DoiSuthep 2	Foyer
14:00-15:20	**R1-2** Natural Language Processing (5 papers : R1-27, R1-30,R1-34, R1-37,R1-40) (Chair: Kiyota Hashimoto, PSU,Thailand)	**R5-1** Smart Industrial Technologies (5 papers : R5-13, R5-21,R5-22, R5-23,R5-29) (Chair: Sumate Lipirodjanapong, MCRU, Thailand)	The 1st International Exhibition of Inventions of Thailand 2019 (IEIT 2019) (Chair: Patiyuth Pramkeaw, KMUTT, Thailand)
15:20-15:30	Break		
15:30-17:30	**R1-3** Natural Language Processing (8 papers : R1-49, R1-52,R1-56,R1-57, R1-60,R1-63, R1-72,R1-74) (Chair: Ye Kyaw Thu, NECTEC, Thailand)	**R5-2** Smart Industrial Technologies (7 papers : R5-31, R5-35,R5-43,R5-44, R5-67,R5-70,R5-76) (Chair: Supakit Sukcharoen, MCRU, Thailand)	The 1st International Exhibition of Inventions of Thailand 2019 (IEIT 2019) (Chair: Patiyuth Pramkeaw, KMUTT, Thailand)
18:00-21:00	Banquet (Ballroom B) Conference Report & Paper Awards & Presentation Awards & (Local Chairs & Session Chairs) Contribution Awards & iSAI-NLP 2020 Announcement (18:00-21:00) (MC: Narumol Chumuang and Pokpong Songmuang) (Award Committee Secretary: Chutima Beokhaimook)		

Friday, November 1, 2019

Time	DoiSuthep 1+2		
08:00-09:00	Registration		
09:00-09:40	**Keynote III** Graph Database for AI Ryota Yamanaka (Oracle Corporation Thailand) (Chair: Prachya Boonkwan, NECTEC, Thailand)		
09:40-10:20	**Keynote IV** Advanced in Human-Computer Interface Itsuo Kamazawa (Tokyo Institute of Technology) (Chair: Kiyota Hashimoto, PSU, Thailand)		
10:20-10:40	Break		
Room	DoiSuthep 1	DoiSuthep 2	Foyer
10:40-12:00	**R3- 4** Data Analytics and Machine Learning (5 papers : R3-62, R3-73,R3-77, R3-78,R3-80) (Chair: Worawut Yimyam, Phetchaburi Rajabhat University, Thailand)	**R2-1** Robotics, IoT and Embedded System (6 papers : R2-09, R2-10,R2-25,R2-26, R2-46,R2-47) (Chair: Thaweesak Yingthawornsuk, KMUTT. Thailand)	The 1st International Exhibition of Inventions of Thailand 2019 (IEIT 2019) (Chair: Patiyuth Pramkeaw, KMUTT, Thailand)
12:00-13:30	Lunch		
Room	DoiSuthep 1+2		
13:30-14:30	Awarding Ceremony (IEIT 2019) (MC: Pokpong Songmuang, TU, Thailand)		

Annotation R1 = Track 1 : Natural Language Processing.
R2 = Track 2 : Robotics, IoT and Embedded System.
R3 = Track 3 : Data Analytics and Machine Learning.
R4 = Track 4 : Signal, Image and Speech Processing.
R5 = Track 5 : Smart Industrial Technologies.
01-80 = paper ID.

Track 1 : Natural Language Processing.

No.	Session	Paper ID: Paper Title
1	R1 group 1	R1-01 The First Wikipedia Questions and Factoid Answers Corpus in the Thai Language
2	R1 group 1	R1-04 Descriptive Feedback on Interns' Performance using a text mining approach
3	R1 group 1	R1-08 Semantic Enhancement and Multi-level Label Embedding for Chinese News Headline Classification
4	R1 group 1	R1-14 Dialogue Breakdown Detection for Understanding Comics with Deep Learning
5	R1 group 1	R1-19 Hierarchical Attention Model for Acquiring Relationships Among Sentences
6	R1 group 2	R1-27 Thai Keyword Extraction using TextRank Algorithm
7	R1 group 2	R1-30 Thai ↔ English Translation Performance of Transformer Neural Machine Translation
8	R1 group 2	R1-34 PLATOOL: Annotation-tool for creating Thai plagiarism corpus
9	R1 group 2	R1-37 An Effect of Using Deep Learning in Thai-English Machine Translation Processes
10	R1 group 2	R1-40 Natural Language Processing
11	R1 group 3	R1-49 Creating Awareness of Incorrect English Pronunciation in Thai Elementary School Students using the Detect Me English,
12	R1 group 3	R1-52 Review Rating Prediction with Gaussian Process Classification
13	R1 group 3	R1-56 Parsing Thai Social Data: A New Challenge for Thai NLP
14	R1 group 3	R1-57 Using Noise Filtering and Ensemble Method for Sentiment Analysis on Thai Social Data
15	R1 group 3	R1-60 A Framework of Computer-Based Learning System Based on Self-Regulated Model in English Writing
16	R1 group 3	R1-63 Using Conceptual Graph to Represent Semantic Relation of Thai Facebook Posts in Marketing
17	R1 group 3	R1-72 Statistical Machine Translation between Kachin and Rawang
18	R1 group 3	R1-74 A Supportive Environment for Knowledge Construction based on Semantic Web Technology: A Case Study in a Cultural Domain

Track 2 : Robotics, IoT and Embedded System.

No.	Session	Paper ID: Paper Title
1	R2 group 1	R2-09 Sumo Based Dynamic Traffic Simulation for Intelligence Traffic Management System
2	R2 group 1	R2-10 R-Cane: A Mobility Aid for Visually Impaired
3	R2 group 1	R2-25 Design and Fabrication of an Affordable SCARA 4-DOF Robotic Manipulator for Pick and Place Objects
4	R2 group 1	R2-26 Design and Implementation of an Indigenous Solar Powered 4-DOF Robotic Manipulator Controlled Unmanned Ground Ve
5	R2 group 1	R2-46 Classification of Depressed Speech Samples using Spectral Energy Ratios as Depression Indicator
6	R2 group 1	R2-47 Characterizing Depressive Speech with MFCC

Track 3 : Data Analytics and Machine Learning.

No.	Session	Paper ID: Paper Title
1	R3 group 1	R3-07 Comparing Effectiveness of Six Text Classifiers for Predicting Stock Price Direction of SET
2	R3 group 1	R3-15 Deep Neural Network Pretrained by a Support Vector Machine
3	R3 group 1	R3-16 the novel index of the similarity between hand-drawn sketches for machine learning
4	R3 group 1	R3-17 Optical-based Limit Order Book Modelling using Deep Neural Networks
5	R3 group 1	R3-20 Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO
6	R3 group 2	R3-24 Parameterized Minutiae Analysis for Generating Secured Fingerprint Template
7	R3 group 2	R3-32 Query-by-Example Word Spotting with Fuzzy Word Sizes
8	R3 group 2	R3-36 TVis: A Light-weight Traffic Visualization System for DDoS Detection
9	R3 group 2	R3-41 Combining Extreme Multi-label Classification and Principal Label Space Transformation for Cold Start Thread Recommenc
10	R3 group 2	R3-42 Analysis of Detecting and Interpreting Warning Signs for Distance of Cars using Analyzing the License Plate
11	R3 group 3	R3-45 Predicting Chance of Success on Epiretinal Membrane Surgery using Deep Learning
12	R3 group 3	R3-51 Effective face verification systems based on histogram of oriented gradients and deep learning techniques
13	R3 group 3	R3-53 Predictive Analytics of Various Factors Influencing Gold Prices in Thailand using ARIMA Model on R
14	R3 group 3	R3-54 The development of an Alerting System for Spread of Brown Plant hoppers in paddy Using Unmanned Aerial Vehicle and Ir
15	R3 group 3	R3-55 Bandit Multiclass Linear Classification for the Group Linear Separable Case
16	R3 group 3	R3-58 A Classification Model for Thai Statement Sentiments by deep learning techniques
17	R4 group 3	R3-59 The analysis for quantitative evaluation of palpation skills in maternity nursing
18	R3 group 4	R3-61 Predicting Drug Sale Quantity using Machine Learning
19	R3 group 4	R3-62 Analyzing behavior in nursing training toward grasping trainee's situation remotely
20	R3 group 4	R3-73 Predicting business alliance factors that affect community enterprise performance
21	R3 group 4	R3-77 Gender Recognition from Facial Images using Local Gradient Feature Descriptors
22	R3 group 4	R3-78 Develop the Framework Conception for Hybrid Indoor Navigation for Monitoring inside Building using Quadcopter
23	R3 group 4	R3-80 Document Clustering for Oil and Gas News Articles

Track 4 : Signal, Image and Speech Processing.

No.	Session	Paper ID: Paper Title
1	R4 group 1	R4-05 Improving Voice Activity Detection by using Denoising-Based Techniques with Convolutional LSTM
2	R4 group 1	R4-11 RoadWay: Lane Detection for Autonomous Driving Vehicles via Deep Learning
3	R4 group 1	R4-12 A combined method for detecting seven segment digit detection on medical devices
4	R4 group 1	R4-38 Thai Vowels Speech Recognition using Convolutional Neural Networks
5	R4 group 1	R4-64 Object Distance Estimation with machine learning algorithms for Stereo Vision
6	R4 group 2	R4-65 Automatic Football Match Event Detection from the Scoreboard using a Single-Shot MultiBox Detector
7	R4 group 2	R4-66 A Light-Weight Deep Convolutional Neural Network for Speech Emotion Recognition using Mel-Spectrograms
8	R4 group 2	R4-68 Design and Implementation of A Smart Shopping Basket Based on IoT Technology
9	R4 group 2	R4-71 Adaptive e-Learning Recommendation Model Based on Multiple Intelligence
10	R4 group 2	R4-79 DDOS Attack Detection & Prevention in SDN using OpenFlow Statistics

Track 5 : Smart Industrial Technologies.

No.	Session	Paper ID: Paper Title
1	R5 group 1	R5-13 Synchronization Control for Microgrid Seamless Reconnection
2	R5 group 1	R5-21 An Efficiency Comparison for Predicting of Educational Achievement Based on LMT
3	R5 group 1	R5-22 Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier
4	R5 group 1	R5-23 The Development of a Model to Predict Marbling Score for fattening Kamphaeng Saen Beef Breed Using Data Mining
5	R5 group 1	R5-29 Intelligent Credit Service Risk Predicting System Based on Customer's Behavior By Using Machine Learning
6	R5 group 2	R5-31 Development of Fun Hint Game Applications for Special Children on Smart Devices
7	R5 group 2	R5-35 The Development of Intelligent Models for Health Classification
8	R5 group 2	R5-43 Automatic Envelope Sorting Using the Template Matching Technique
9	R5 group 2	R5-44 Smart Industrial Technologies Interactive LED Table
10	R5 group 2	R5-67 The Development of Eyes Tracking System in Smartphone for Disabled Arm Person
11	R5 group 2	R5-70 Short-circuit and Over Current Notification in Sub-transmission Line by Messege Cellular Network
12	R5 group 2	R5-76 Voltage Failure Warning Device for 3-Phase Transformer

conference Overall Schedule

Day 1 WEDNESDAY, OCTOBER 30

Time		ROOM
[08:00-09:00]	**Registration**	**DOISUTHEP 1**

Tutorial 1 (NLP)

Time	[09:00-10:20]	DOISUTHEP 1
Title	**When 1 + 1 > 2 : Joint Neural NLP Models Demystified**	
Session Chair:	**Prachya Boonkwan and Ye Kyaw Thu** NECTEC, Thailand	

TRACK 3 : Data Analytics and Machine Learning.

GROUP 1	4 papers	DOISUTHEP 2
Session Chair:	**Worawut Yimyam** Phetchaburi Rajabhat University,Thailand	

Time	Paper ID	Title and Auther
[09:00-09:15]	R3-15	Deep Neural Network Pretrained by a Support Vector Machine *Hironori Yamamoto and Naoki Mori*
[09:15-09:30]	R3-16	The novel index of the similarity between hand-drawn sketches for machine learning *Ryosuke Fujii, Naoki Mori and Makoto Okada*
[09:30-09:45]	R3-17	Optical-based Limit Order Book Modelling using Deep Neural Networks *Pak Laowatanachai and Poj Tangamchit*
[09:45-10:00]	R3-20	Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO *Kwanruan Rusmee and Narumol Chumuang*

Day 1 WEDNESDAY, OCTOBER 30

Tutorial 1 (NLP)

Time	[10:40-12:00]	**DOISUTHEP 1**
Title	**When 1 + 1 > 2 : Joint Neural NLP Models Demystified**	
Session Chair:	**Prachya Boonkwan and Ye Kyaw Thu** NECTEC, Thailand	

TRACK 3 : Data Analytics and Machine Learning.

GROUP 2	**5 papers**		**DOISUTHEP 2**
Session Chair:	**Itsuo Kamazawa** Tokyo Institute of Technology, Japan		
Time	**Paper ID**	**Title and Auther**	
[10:40-10:55]	R3-24	Parameterized Minutiae Analysis for Generating Secured Fingerprint Template *Md. Mijanur Rahman and Tanjarul Islam Mishu*	
[10:55-11:10]	R3-32	Query-by-Example Word Spotting with Fuzzy Word Sizes *Amornchai Wiwatcharee and Tanasanee Phienthrakul*	
[11:10-11:25]	R3-36	TVis: A Light-weight Traffic Visualization System for DDoS Detection *Abhishek Kalwar, Monowar Bhuyan, Dhruba Bhattacharyya, Jugal Kalita, Youki Kadobayashi and Erik Elmroth*	
[11:25-11:40]	R3-41	Combining Extreme Multi-label Classification and Principal Label Space Transformation for Cold Start Thread Recommendation *Kantarakorn Jitharn and Eakasit Pacharawongsakda*	
[11:40-11:55]	R3-42	Analysis of Detecting and Interpreting Warning Signs for Distance of Cars using Analyzing the License Plate *Patiyuth Pramkeaw, Mahasak Ketcham, Warissara Limpornchitwilai and Narumol Chumuang*	

NLP Workshop

Time	[13:00-16:00]	**DOISUTHEP 1**
Title	**Workshop**	
Session Chair:	**Ye Kyaw Thu, Thepchai Supnithi** NECTEC, Thailand	

Day 1 WEDNESDAY, OCTOBER 30

TRACK 4 : Signal, Image and Speech Processing.

GROUP 1	**6 papers**	**DOISUTHEP 2**
Session Chair:	**Narit Hnoohom** Mahidol University, Thailand	

Time	**Paper ID**	**Title and Auther**
[13:30-13:45]	R4-64	Object Distance Estimation with machine learning algorithms for Stereo Vision *Pawarit Akepitaktam and Narit Hnoohom*
[13:45-14:00]	R4-65	Automatic Football Match Event Detection from the Scoreboard using a Single-Shot MultiBox Detector *Rungroj Somwong and Narit Hnoohom*
[14:00-14:15]	R4-66	A Light-Weight Deep Convolutional Neural Network for Speech Emotion Recognition using Mel-Spectrograms *Kamin Atsavasirilert, Thanaruk Theeramunkong, Sasiporn Usanavasin, Anocha Rugchatjaroen, Surasak Boonkla, Jessada Karnjana, Suthum Keerativittayanun and Manabu Okumura*
[14:15-14:30]	R4-68	Design and Implementation of A Smart Shopping Basket Based on IoT Technology *Sakorn Mekruksavanich*
[14:30-14:45]	R4-71	Adaptive e-Learning Recommendation Model Based on Multiple Intelligence *Sakorn Mekruksavanich and Thaksaorn Jommanop*
[14:45-15:00]	R4-79	DDOS Attack Detection & Prevention in SDN using OpenFlow Statistics *Nisha Ahuja and Gaurav Singal*

Day 1 WEDNESDAY, OCTOBER 30

TRACK 3 : Data Analytics and Machine Learning.

GROUP 3	**7 papers**	**DOISUTHEP 2**
Session Chair:	**Olarik Surinta** Mahasarakham University, Thailand	

Time	**Paper ID**	**Title and Auther**
[15:30-15:45]	R3-45	Predicting Chance of Success on Epiretinal Membrane Surgery using Deep Learning *Suvimol Reintragulchai, Thanaruk Theeramunkong, Paisan Ruamviboonsuk, Natsuda Kaothanthong and Vorarit Jinaratana*
[15:45-16:00]	R3-51	Effective face verification systems based on histogram of oriented gradients and deep learning techniques *Sawitree Khunthai, Pichada Saichua and Olarik Surinta*
[16:00-16:15]	R3-53	Predictive Analytics of Various Factors Influencing Gold Prices in Thailand using ARIMA Model on R *Nuntouchapron Prateepausanont, Chidchanok Choksuchat and Sureena Matayong*
[16:15-16:30]	R3-54	The development of an Alerting System for Spread of Brown Plant hoppers in paddy Using Unmanned Aerial Vehicle and Image Processing Technique *Worawut Yimyam and Mahasak Ketcham*
[16:30-16:45]	R3-55	Bandit Multiclass Linear Classification for the Group Linear Separable Case *Jittat Fakcharooenphol and Chayutpong Prompak*
[16:45-17:00]	R3-58	A Classification Model for Thai Statement Sentiments by deep learning techniques *Pakawan Pugsee and Nitikorn Ongsirimongkol*
[17:00-17:15]	R3-61	Predicting Drug Sale Quantity using Machine Learning *Warayut Saena and Vasin Suttichaya*

Tutorial 2 (NLP)

Time	[16:00-17:30]	**DOISUTHEP 1**
Title	**How to improve your research: A writing-asresearch method**	
Session Chair:	**Kiyota Hashimoto** Prince of Songkla University, Thailand	

Day 2 THURSDAY, OCTOBER 31

Time		ROOM
[08:00-09:00]	**Registration**	**DOISUTHEP 1**

Opening Ceremony

Time	[09:00-10:20]	
MC	**Rachada Kongkachan & Prachya Boonkwan**	**DOISUTHEP 1**
Photo Session	**[09:20]**	

Keynote I

Time	[09:20-10:10]	
Title	**Artificial Intelligence for Medical Screening: Research and Innovation in Low-resource Settings**	**DOISUTHEP 1**
Keynote:	**Charturong Tantibundhit** Thammasat University	
Chair:	**Pokpong Songmuang**	

TRACK 1 : Natural Language Processing.

GROUP 1	**5 papers**	**DOISUTHEP 1**
Session Chair:	**Mikifumi Shikida** KUT, Japan	

Time	Paper ID	Title and Auther
[10:40-10:55]	R1-01	The First Wikipedia Questions and Factoid Answers Corpus in the Thai Language *Kanokorn Trakultaweekoon, Santipong Thaiprayoon, Pornpimon Palingoon and Anocha Rugchatjaroen*
[10:55-11:10]	R1-04	Descriptive Feedback on Interns' Performance using a text mining approach *Geraldine Mangmang, Larmie Feliscuzo and Elmer Maravillas*
[11:10-11:25]	R1-08	Semantic Enhancement and Multi-level Label Embedding for Chinese News Headline Classification *Jiangnan Qi, Yuan Rao, Ling Sun and Xiong Yang*
[11:25-11:40]	R1-14	Dialogue Breakdown Detection for Understanding Comics with Deep Learning *Ryo Iwasaki and Naoki Mori*
[11:40-11:55]	R1-19	Hierarchical Attention Model for Acquiring Relationships Among Sentences *Hiroki Teranishi, Makoto Okada and Naoki Mori*

Day 2 THURSDAY, OCTOBER 31

TRACK 4 : Signal, Image and Speech Processing.

GROUP 2	**4 papers**	**DOISUTHEP 2**
Session Chair:	**Sakorn Mekruksavanich** University of Phayao, Thailand	

Time	**Paper ID**	**Title and Auther**
[10:40-10:55]	R4-05	Improving Voice Activity Detection by using Denoising-Based Techniques with Convolutional LSTM *Nattapong Kurpukdee, Surasak Boonkla, Vataya Chunwijitra, Phuttapong Sertsi and Sawit Kasuriya*
[10:55-11:10]	R4-11	RoadWay: Lane Detection for Autonomous Driving Vehicles via Deep Learning *Gaurav Singal, Riti Kushwaha and Himanshu Singhal*
[11:10-11:25]	R4-12	A combined method for detecting seven segment digit detection on medical devices *Noppakun Boonsim and Saranya Kanjaruek*
[11:25-11:40]	R4-38	Thai Vowels Speech Recognition using Convolutional Neural Networks *Niyada Rukwong and Sunee Pongpinigpinyo*

IEIT 2019

Time	[10:40-17:30]	**FOYER**
Title	**The 1st International Exhibition of Inventions of Thailand 2019**	
Session Chair:	**Patiyuth Pramkeaw,** KMUTT, Thailand	

Keynote II

Time	[13:30-14:00]	**DOISUTHEP 1**
Title	**Connecting formal and informal learning through learning evidence and analytics framework**	
Keynote:	**Hiroaki Ogata** Kyoto University	
Chair:	**Thepchai Supnithi** NECTEC, Thailand	

Day 2 THURSDAY, OCTOBER 31

TRACK 1 : Natural Language Processing.

GROUP 2	**5 papers**	**DOISUTHEP 1**
Session Chair:	**Kiyota Hashimoto** PSU,Thailand	

Time	**Paper ID**	**Title and Auther**
[14:00-14:15]	R1-27	Thai Keyword Extraction using TextRank Algorithm *Rattapoom Kedtiwerasak, Ekkarat Adsawinnawanawa, Pimolluck Jirakunkanok and Rachada Kongkachandra*
[14:15-14:30]	R1-30	Thai ↔ English Translation Performance of Transformer Neural Machine Translation *Kanchana Saengthongpattana, Kanyanut Kriengket, Peerachet Porkaew and Thepchai Supnithi*
[14:30-14:45]	R1-34	PLATOOL: Annotation-tool for creating Thai plagiarism corpus *Supon Klaithin, Pornpimon Palingoon, Kanokorn Trakultaweekoon and Santipong Thaiprayoon*
[14:45-15:00]	R1-37	An Effect of Using Deep Learning in Thai-English Machine Translation Processes *Prasert Luekhong, Peerat Limkonchotiwat and Taneth Ruangrajitpakorn*
[15:00-15:15]	R1-40	Natural Language Processing *Brijesh Bhatt*

Day 2 THURSDAY, OCTOBER 31

TRACK 5 : Smart Industrial Technologies.

GROUP 1	**5 papers**	**DOISUTHEP 2**
Session Chair:	**Sumate Lipirodjanapong** MCRU, Thailand	

Time	**Paper ID**	**Title and Auther**
[14:00-14:15]	R5-13	Synchronization Control for Microgrid Seamless Reconnection *Thakul Uten, Chalie Charoenlarpnopparut and Prapun Suksompong*
[14:15-14:30]	R5-21	An Efficiency Comparison for Predicting of Educational Achievement Based on LMT *Sudarat Thiennoi, Narumol Chumuang and Chairit Siladech*
[14:30-14:45]	R5-22	Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier *Kumnung Vongprechakorn, Narumol Chumuang and Adil Farooq*
[14:45-15:00]	R5-23	The Development of a Model to Predict Marbling Score for fattening Kamphaeng Saen Beef Breed Using Data Mining *Watchara Ninphet, Narumol Chumuang, Chairit Siladech and Mahasak Ketcham*
[15:00-15:15]	R5-29	Intelligent Credit Service Risk Predicting System Based on Customer's Behavior By Using Machine Learning *Jittimaporn Chaisuwan and Narumol Chumuang*

Day 2 THURSDAY, OCTOBER 31

TRACK 1 : Natural Language Processing.

GROUP 3	**8 papers**	**DOISUTHEP 1**
Session Chair:	**Ye Kyaw Thu** NECTEC, Thailand	

Time	**Paper ID**	**Title and Auther**
[15:30-15:45]	R1-49	Creating Awareness of Incorrect English Pronunciation in Thai Elementary School Students using the Detect Me English, Natural Language Processing, Application *Ma'Ayan Grace and Jeerapan Phomprasert*
[15:45-16:00]	R1-52	Review Rating Prediction with Gaussian Process Classification *Hidekazu Yanagimoto and Kiyota Hashimoto*
[16:00-16:15]	R1-56	Parsing Thai Social Data: A New Challenge for Thai NLP *Sattaya Singkul, Borirat Khampingyot, Nattasit Maharattamalai, Supawat Taerungruang and Tawunrat Chalothorn*
[16:15-16:30]	R1-57	Using Noise Filtering and Ensemble Method for Sentiment Analysis on Thai Social Data *Chayanont Eamwiwat, Pongpisit Thanasutives, Chanatip Saetia and Tawunrat Chalothorn*
[16:30-16:45]	R1-60	A Framework of Computer-Based Learning System Based on Self-Regulated Model in English Writing *Kanyalag Phodong, Thepchai Supnithi and Rachada Kongkachandra*
[16:45-17:00]	R1-63	Using Conceptual Graph to Represent Semantic Relation of Thai Facebook Posts in Marketing *Kwanrutai Saelim and Rachada Kongkachandra*
[17:00-17:15]	R1-72	Statistical Machine Translation between Kachin and Rawang *Ye Kyaw Thu, Manar Hti Seng, Thazin Myint Oo, Dee Wom, Hpau Myang Thint Nu, Seng Mai, Thepchai Supnithi and Khin Mar Soe*
[17:15-17:30]	R1-74	A Supportive Environment for Knowledge Construction based on Semantic Web Technology: A Case Study in a Cultural Domain *Akkharawoot Takhom, Dhanon Leenoi, Sasiporn Usanavasin, Prachya Boonkwan and Thepchai Supnithi*

Day 2 THURSDAY, OCTOBER 31

TRACK 5 : Smart Industrial Technologies.

GROUP 2	**6 papers**	**DOISUTHEP 2**
Session Chair:	**Supakit Sukcharoen** MCRU, Thailand	

Time	**Paper ID**	**Title and Auther**
[15:30-15:45]	R5-31	Development of Fun Hint Game Applications for Special Children on Smart Devices *Burin Narin, Titichaya Sribun and Tanniga Yoongrum*
[15:45-16:00]	R5-35	The Development of Intelligent Models for Health Classification *Wattanapong On-Num, Narumol Chumuang and Chairit Siladech*
[16:00-16:15]	R5-44	Smart Industrial Technologies Interactive LED Table *Sirimonpak Suwannakhun*
[16:15-16:30]	R5-67	The Development of Eyes Tracking System in Smartphone for Disabled Arm Person *Thidarat Pinthong and Mahasak Ketcham*
[16:30-16:45]	R5-70	Short-circuit and Over Current Notification in Sub-transmission Line by Messege Cellular Network *Sumate Lipirodjanapong, Pumpat Uthaisiritanon and Pitipol Duangjinda*
[16:45-17:00]	R5-76	Voltage Failure Warning Device for 3-Phase Transformer *Asst.Prof.Dr. Luechai Promratrak*

BANQUET

Time	[18:00-21:00]	**BALLROOM B**
Title	**Conference Report & Paper Awards & Presentation Awards & (Local Chairs & Session Chairs) Contribution Awards & iSAI-NLP 2020 Announcement**	
MC	**Narumol Chumuang and Pokpong Songmuang**	

Day 3 FRIDAY, NOVEMBER 1

Time		ROOM
[08:00-09:00]	**Registration**	**DOISUTHEP 1**

Keynote III

Time	[09:00-09:40]	
Title	**Graph Database for AI**	**DOISUTHEP 1**
Keynote:	**Ryota Yamanaka** Oracle Corporation Thailand	
Chair:	**Prachya Boonkwan** NECTEC, Thailand	

Keynote IV

Time	[09:40-10:20]	
Title	**Advanced in Human-Computer Interface**	**DOISUTHEP 1**
Keynote:	**Itsuo Kamazawa** Tokyo Institute of Technology	
Chair:	**Kiyota Hashimoto** PSU, Thailand	

Day 3 FRIDAY, NOVEMBER 1

TRACK 3 : Data Analytics and Machine Learning.

GROUP 4	**5 papers**	**DOISUTHEP 1**
Session Chair:	**Worawut Yimyam** Phetchaburi Rajabhat University, Thailand	

Time	**Paper ID**	**Title and Auther**
[10:40-10:55]	R3-62	Analyzing behavior in nursing training toward grasping trainee's situation remotely *Yuki Kodera, Miwa Saito, Sumika Yoshimura, Kyoko Yamawaki, Kunimasa Yagi and Mikifumi Shikida*
[10:55-11:10]	R3-73	Predicting business alliance factors that affect community enterprise performance *Wiwit Suksangaram, Waratta Hemtong and Sopaporn Klamsakul*
[11:10-11:25]	R3-77	Gender Recognition from Facial Images using Local Gradient Feature Descriptors *Olarik Surinta and Thananchai Khamket*
[11:25-11:40]	R3-78	Develop the Framework Conception for Hybrid Indoor Navigation for Monitoring inside Building using Quadcopter *Sanya Khruahong and Olarik Surinta*
[11:40-11:55]	R3-59	The analysis for quantitative evaluation of palpation skills in maternity nursing *Shunya Inoue, Sumika Yoshimura, Miwa Saito, Kyoko Yamawaki and Mikifumi Shikida*

Day 3 FRIDAY, NOVEMBER 1

TRACK 2 : Robotics, IoT and Embedded System.

GROUP 1	**5 papers**	**DOISUTHEP 2**
Session Chair:	**Thaweesak Yingthawornsuk** KMUTT. Thailand	

Time	**Paper ID**	**Title and Auther**
[10:40-10:55]	R2-10	R-Cane: A Mobility Aid for Visually Impaired *Kanak Manjari, Madhushi Verma and Gaurav Singal*
[10:55-11:10]	R2-25	Design and Fabrication of an Affordable SCARA 4-DOF Robotic Manipulator for Pick and Place Objects *Sara Fatima Noshahi, Adil Farooq, Muhammad Irfan and Narumol Chumuang*
[11:10-11:25]	R2-26	Design and Implementation of an Indigenous Solar Powered 4-DOF Robotic Manipulator Controlled Unmanned Ground Vehicle *Adil Farooq, Sundas Arshad, Tayyaba Ansar, Muhammad Irfan and Narumol Chumuang*
[11:25-11:40]	R2-46	Classification of Depressed Speech Samples using Spectral Energy Ratios as Depression Indicator *Thaweesak Yingthawornsuk and Thaweewong Akkaralaertsest*
[11:40-11:55]	R2-47	Characterizing Depressive Speech with MFCC *Thaweesak Yingthawornsuk and Sirimonpak Suwannakhun*

IEIT 2019

Time	[10:40-12:00]	**FOYER**
Title	**The 1st International Exhibition of Inventions of Thailand 2019**	
Session Chair:	**Patiyuth Pramkeaw,** KMUTT, Thailand	

Awarding Ceremony (IEIT 2019)

Time	[13:30-14:30]	**DOISUTHEP 1**
MC	**Pokpong Songmuang** TU, Thailand	

Contents

The First Wikipedia Questions and Factoid Answers Corpus in the Thai Language

Kanokorn Trakultaweekoon
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
Pathumthani, Thailand
kanokorn.tra@nectec.or.th

Santipong Thaiprayoon
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
Pathumthani, Thailand
santipong.tha@nectec.or.th

Pornpimon Palingoon
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
Pathumthani, Thailand
pornpimon.pal@nectec.or.th

Anocha Rugchatjaroen
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
Pathumthani, Thailand
anocha.rugchatjaroen@nectec.or.th

***Abstract*— This article introduces a Thai questions-answers corpus for a question-answering task which was extracted from Thai Wikipedia which was downloaded on 17 December 2017. The answers comprise 5,000 annotated factoids. The corresponding questions are exact phrases/sentences that contain the answer, but are replaced by a question word, or synthetic questions acquired from phrases and/or sentences on the wiki page. A question must contain only one of a set of 7 specific question words and a complex question must be avoided. Fifteen annotators used an annotation system specifically designed for this task. Acceptance, rejection, and revision processes were monitored by a language specialist. The final set was divided into 4,000 pairs for a training set and 1,000 pairs for a validation set. A baseline evaluation was conducted and an F1 score of 27.25 was obtained from document readers and 71.24 from document retrievals.**

***Keywords*— Thai questions-answers corpus, Thai Question-Answering system**

I. INTRODUCTION

A Thai Question-Answering (QA) system is a challenging task especially in the Thai language. Nowadays, English QA corpuses are available for research and benchmarking. In 2016, Microsoft research introduced WikiQA, an English QA corpus gathered from Bing, Microsoft search engine, and query logs. It contains 3,047 pairs of questions and answers based on user clicks on the Wikipedia page [1]. In 2017, Joshiy M. et al. announced TriviaQA which is the largest QA corpus which contains 650K of question-answer evidence triples. The evidence documents are the third important element which were collected and checked for redundancy from a Web search of results and Wikipedia pages [2]. Recently, a Google research team introduced "Natural Questions: a Benchmark for Question Answering Research", in which the questions are from queries issued to the Google search engine. In total, it contains 307, 373 training examples with single annotations. Its development set contains 7,830 examples from 5-way annotations, which is a technique of answer collection that used a nonnull answer that is seen at least once in the 5 annotations (see Section 5 of this paper for more details) and a further 7,842 examples of sequestered test data [3].

In Thai, asking for a meaning of a word from Thai Wikipedia is the easiest type of question-answering task in NLP, because the system can search for its result by searching through a list of page titles. However, searching for a specific answer on a page is another challenging NLP task. The traditional Thai information retrieval system used a well-structured knowledge graph. This stores knowledge at the roots of knowledge trees that have well-designed paths to the root [4]. To construct a tree, the researcher has to work with Thai texts which do not show syllables, words, phrases, or punctuation at the end of sentences. Hence, the construction has always been done manually, which caused difficulties to create an open-domain QA system in the Thai language.

A process of information extraction from both questions and answers has to be well defined before establishing a corpus. Since this work planned to use the information from Thai Wikipedia, the first definition of this work was "the answer must be an exact word or number found in TH-Wiki". Then the corresponding question has to be a combination of phrases found in the same paragraph.

A web-based annotation system has been established. It has two modes, annotator mode and linguistic administration mode. For annotator mode, it allows a user to search for a topic, create, edit, revise, and delete QA pairs, whilst administration mode allows one to correct, comment, accept, or reject pairs. The content of this paper is organized as follows: Section 2 explains the difficulties of working with the Thai language for a QA task, Section 3 shows the proposed system which includes all the corpus design and the annotation restrictions, Section 4 analyzes the collected QA pairs, Section 5 describes the implementation of a QA baseline system with its accuracies and Section 6 concludes this paper.

II. Related works and Problems

In the English QA system, an unstructured knowledge-based approach has been developed using a deep learning approach such as memory networks from Jason Weston and Sainbayar Sukhabaatar (2015), which uses an attention mechanism for the information retrieval (IR) process. However, an IR implementation for an unstructured-knowledge source requires a good annotated corpus for training. Recent years, there are a few numbers of QA corpuses created from Wikipedia contents, WikiQA [1], SQuAD [5] and TriviaQA [2] mentioned in section 1, they also used Wikipedia as one of their major resources. WikiQA

from Microsoft research influenced our construction of the very first Thai Wiki-QA set.

Research in Thai information retrieval started in 1996 when Asst. Prof. Somchai Prasitjutrakul and his student Paramin Jindavimonlert established a Thai text retrieval system using PAT trees [4]. It used a hashed indexing tree to store and retrieve text. However, for the Thai language, Thai Wikipedia has also been a favorite resource for Thai NLP researchers, but Thai QAs corpuses has not been established yet. This work then tagged factoids and formed their corresponding questions, then published them for research use.

A factoid can be a unit of a word or a phrase or a date or a number or even an equation. This work focused on a unit of a single word only, hence single words were extracted manually by 7 annotators. Although the Thai writing system does not show syllables, words, phrases, or punctuation at the end of sentences, this corpus does not provide any of them. Therefore, the information extraction process needs to work automatically without them.

III. CORPUS CREATION

This work proposed a set of Thai QAs extracted from Thai Wikipedia. The answers are factoids which are single words extracted from the source. The corresponding question is assembled from words and phrases from the same page with an additional question word.

There are three main components in the corpus creation of this work. They are a resource text which is from Thai Wikipedia, annotators, and an annotation system. Thai Wikipedia, the resource, was downloaded on 12 December 2017. It contained 120,764 articles from 304,693 registered users, some of which were authors. The annotators were 15 native Thai speakers with different kinds of expertise. They were undergraduate students, post-graduate students, and computer scientists. Fourteen of them are female, and one of them is male. They were 25 years of age on average.

The annotation system was a web application written in Javascript, PHP, and it used MySQL as a database. Three tables in a structure of a relational database system stored information of: created questions, answers, a character count of the position where the answer begins, a character count of the position where the answer ends, related Wiki content ID, a section of answers in the Wiki content ID, and annotator details. The frontend consisted of two modes, which were for the annotators and for the administrators.

When an annotator login to the system, the interface provides a topic search tool for choosing a Wiki article, then the user chooses an article and reads it. Afterwards, the user can annotate an exact answer (as found in the text), then creates a corresponding question. There were 2 types of question which are an exact question and a modified question, respectively. An exact question is a question formed by the phrase or sentence that contains the answer, but replaces the answer words with a question word and a modified question is a question formed by multiple phrases in the text plus an additional question word for the answer. All QA pairs were controlled by a conductor, who was a specialist in charge of setting the scope of the different questions and responsible for issuing annotation guidelines. The guidelines contain a preliminary scope of the Thai questions answers system. It was created by an experienced Thai semantics expert who is called a conductor. This corpus has been involved with only basic questions at present, therefore a question has to be appropriate in order to provide an answer. All constraints have to be carefully set. They are shown in the next few paragraphs.

The details of the annotation guidelines are divided into three parts: the definition of a question in this work, the constraints for question creation, and the language restrictions. This work defines the meaning of a question as "The simplest question one can ask for an annotated answer formed using phrases and words on the same page plus an additional question word." The interrogative words used are

- "อะไร" ('what')
- "ใคร" ('who')
- "ไหน" ('which or where')
- "เมื่อไหร่" ('when')
- "ใด" ('which or where')
- "กี่" ('how much/many or when')
- "เท่าไร" ('how much/many')

This corpus was created under 7 constraints when forming a question which are:

(1) It must not form a complex question or use "why" or "how" as the question word.

(2) It must use formal language with a formal question structure which means it must contain a subject + verb + object.

(3) Questions and answers must be semantically clear.

(4) Every question must contain a question word from the defined set only.

(5) A question must contain symbols, e.g. ":", ",", ";", which they must be used precisely as appears in the text, not from annotator insertion.

(6) The question length must be less than 2 lines.

(7) Typos from Wiki should be avoided in questions and answers.

In addition, there is a simple language restriction, which is that all questions must be polite, although they can be in informal language.

IV. CORPUS CHARACTERISTICS

The corpus contains 5,000 QA pairs extracted from 2,923 Thai Wikipedia articles out of 120,764 which were downloaded. Table 4.1 shows the numbers of QA pairs separated by the two types of question as described in the previous section.

TABLE 1. NUMBER OF QA PAIRS IN TRAINING AND VALIDATION SETS

Set of data	QA pairs	Number of Question type	
		Exact	Modified
Train	4,000	573	3,427
Validation	1,000	147	853

There was no control for balancing the numbers of each question type, because it was not easy to find a well-structured

sentence or phrase for forming an “Exact” type of question. Therefore, the figures show that the annotators synthesized proper questions rather than editing a sentence/phrase.

The average position of answers on a page is 22.06% of page, which was calculated by finding the average location of all the answers then converting them to percentages based on overall page lengths, which we call a normalized location in this paper. Moreover, answer lengths are about 12.56 characters on average. The distributions of answers, normalized locations and lengths are shown in Fig.1.

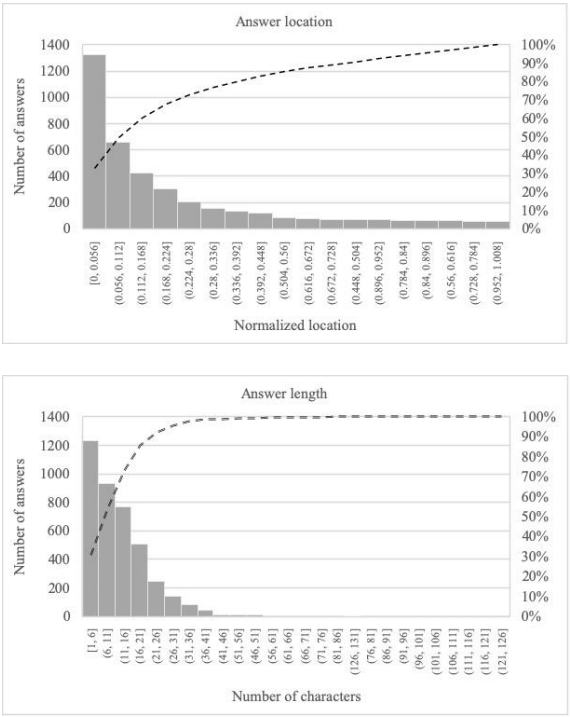


Fig. 1. Histograms of answer locations (top) and answer lengths (bottom). Dashed lines represent accumulative percentages of numbers in each graph.

The statistics for the use of each interrogative word are shown in Table 2. Some of the question words have common spelling variations, but they have the same meaning, so they are grouped in the table.

TABLE 2. NUMBER OF QA PAIRS SEPARATED BY AN INTERROGATIVE WORD.

Interrogative word		% of usage in corpus
Defined	Variation	
อะไร *(what)*	อะไร	31.10
	ว่าอย่างไร	0.04
Interrogative word		% of usage in corpus
Defined	Variation	
ใคร *(who)*	ใคร	16.11
ไหน *(which or where)*	ไหน	0.58
เมื่อไร *(when)*	เมื่อไหร่	n/a
	เมื่อไร	0.52
	เมื่อใด	0.92
กี่ *(how much/many or when)*	กี่	4.83
เท่าไร *(how much/many)*	เท่าไหร่	0.12
	เท่าไร	8.13
	เท่าใด	1.04
ใด *(which or where)*	ใด	36.65

A total of 5,000 factoids were annotated from the pages. The answers corresponding to the questions tended to be straightforward. However, this corpus contains noise data. Seven questions are noise. They contain typos or no question word, or neither. The questions with feigned noise are listed in Table 3.

TABLE 2. SEVEN NOISE QUESTIONS IN THE CORPUS.

Question	Error type(s)
ซิริล เฮย์คอค เกิดในตระกูลที่มีบิดาสืบเชื้อสายมาจากขุนนางเก่าแก่ บิดามีชื่อเรียกว่าอะไ Wha was Ceril Heycock's father name who was born under old hierarchies of nobility?	Spelling mistake
เอ ศุภชัย ได้รักชวน เจมส์ มาร์ เข้าสู่วงการ ด้วยการเจอกันที่ไค Whare did A Supphachai persuate James Mars to join his angency?	Spelling mistake
นุติ เขมะโยธิน เกิดเมื่อวันที่ ~~When~~ was Nuti Khemayothin born?	NoQword
ชลาศัย ขวัญชิติ หรือหม่อมลูกปลา เกิดเมื่อวันที่ ~~What~~ was Chalasai Kwanthiti, called Mhom Lookpla, birth date?	NoQword
ในปี 2012 เด็กอายุต่ำกว่า 15 ปี ที่ได้รับการวินิจฉัยว่าเ In 2012, Children who was under the age of 15 and diagnosed as having .... disease	NoQword + Spelling mistake
นวนิยายแนวลึกลับและสืบสวนชื่อเรื่องว่า รหัสลับดาวินชี ประพันธ ~~Who~~ wrot the Da Vinci Code, a mystery thriller novel?	NoQword + Spelling mistake
Question	Error type(s)
ใครเป็นผู้ชนะเลิศการแข่งขันรายการเอเชียเน็กซต์ท็อปโมเดล ฤดูกาลที่ 3 Wno was the winner of Asian Next Top Model competition in season 3?	Spelling mistake

V. BASELINE ACCURACY

This research evaluated the corpus using a baseline system. The test flow is as shown in Fig. 2

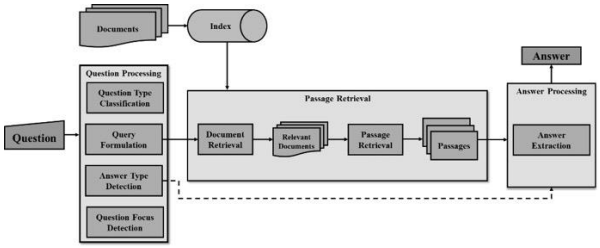


Fig. 2. Baseline system used for finding a preliminary accuracy one can implement using the proposed corpus

The objective of a QA system is to automatically generate a correct answer to a question by finding the relevant documents containing the answer strings. The baseline system consists of three main components: (1) question processing (2) passage retrieval, and (3) answer processing.

Question processing is the first process as shown on the left of Fig.2. It detects a question type, a question answer type, and a question focus word/phrase, then uses the components as query constraints. The second process in the middle of the figure is the passage retrieval. It aims to retrieve a set of text passages that might contain answer strings, so it tries to match a query with documents in the corpus and retrieves/selects only the top 5 passages which have the highest matched

scores. Then the pas-sages are passed to the answer processing as shown on the right of Figure 5.1, which extract an answer by parsing all the passages to a name-entity tagger using a pattern extraction approach. Finally, the system re-turns an answer that matches the answer type detected at the beginning of the process.

The corpus was evaluated using the baseline system. The experimental setting was conducted for two tasks: (1) document retrieval task and (2) document reader task. For the task of document retrieval, we used an F-measure that is commonly used to evaluate performance in information retrieval. The experimental result shows that our technique for the document retrieval task achieved 71.24% at 1-best accuracy. For the task of the document reader, we use the exact match (EM) metric that computes common substrings at word level between the predicted answers and the gold answers. We can achieve 25.92% of F1 scores at 1-best accuracy.

VI. SUMMARY

This paper presents a Thai Wikipedia QA Corpus whose answers are factoids extracted from Thai Wikipedia articles. There are 2 types of questions, which are Exact or Modified. Exact is a type of question that is formed by replacing the answer with a question word. Modified is a type that is synthesized from phrases around the answer plus a question word. In total, the corpus contains 4,000 training QA pairs, and 1,000 validation pairs. It should be noted that they can be mixed together and divided in different proportions for use as a training set, a validation set, or a testing set. A baseline system gives a preliminary baseline performance which can be applied to the proposed corpus. The experiment was conducted using another set of sequestered test data. The F1 score achieved 25.92% at 1-best accuracy. The corpus is free for research use at http://aiforthai.in.th/corpus.

ACKNOWLEDGMENT

The authors would like to thank all Thai Wikipedia authors and the Wikimedia Foundation, Inc. for supporting the Wikipedia platform in many languages.

REFERENCES

[1] G Y. Yang, W.-t. Yih and C. Meek, "WikiQA: A challenge dataset for open-domain question answering," in Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015.

[2] M. Joshi, E. Choi, D. Weld and L. Zettlemoyer, "Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension," in Proceedings of the 55th Annual Meeting of the Association for Compuational Linguistics (Volume 1: Long Papers), Vancouver, 207.

[3] T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, M. Kelcey, J. Devlin, K. Lee, K. N. Toutanova and L. J. a. M, "Natural Questions: a Benchmark for Question Answering Research," Transactions of the Association of Computational Linguistics, 2019.

[4] P. Jindavimonlert, "A Thai Text Retrieval System using the PAT tree: M.Sc. Thesis," Department of Computer Engineering Chulalongkorn University, 1996, 1996.

[5] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, 2016.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy

Descriptive Feedback on Interns' Performance using a text mining approach

Geraldine B. Mangmang
Cebu Institute of Technology University
Cebu City, Philippines
geraldine_mangmang@yahoo.com

Larmie Feliscuzo
Cebu Institute of Technology University
Cebu City, Philippines
larmie.feliscuzo@gmail.com

Elmer A. Maravillas
Cebu Institute of Technology University
Cebu City, Philippines
elmer.maravillas@gmail.com

***Abstract*— Descriptive feedback is a powerful tool to identify the strength and the areas that need improvement of a certain program. Text analysis using text mining approach is the most common application in the text processing field. This study aimed to determine the potentials and the imperfections of the IT interns while taking the internship program using text mining applications. Text dataset from the internship program of the IT curriculum under the College of Computer Studies and Information Technology of Southern Leyte State University was the input of the study. This study applied the text mining application using the Naïve Bayes text classification to determine the performance of the IT interns. Results show that the IT interns of the program were proficient technically but had a problem with their communication skills. The model has obtained an acceptable accuracy rating in predicting the performance of the students. Based on the results, the IT interns were efficient in the IT-related task. However, there is a need to improve interns' communication skills. The management should think for a possible bridging program that will help the student to develop their communication skills.**

Keywords*— *Natural Language Processing, Naïve Bayes algorithms, text classification, interns' performance, predictive model

I. INTRODUCTION

The internship program of the IT curriculum aims to provide students the knowledge and experience in working with the IT-based environment. The program, as mandated by CHED, is envisioned to train student interns in the actual workplace setting [1]. Thus, giving them descriptive feedback on the performance of the interns is important to help the program identify the strength and the areas that need improvement. Text mining application is one of the recommended approaches for text processing documents like the information provided by the intern's supervisor. This study aimed to determine the IT student interns' performance and qualifications after taking the internship period.

Feedbacking has a significant effect on individuals' motivation towards attaining a certain goal [2][3][4]. Performance evaluation is one of the major tools that can help in running a business organization effectively [5][6]. Techniques applied in the performance analysis are necessary to determine the qualifications in the individual being evaluated [7][8]. Text mining approach is commonly known application for text processing field that can help in the improvement of a certain program or system [9][10][11][12][13]. Text mining application is a statistical technique that is effective in the text analytics system [14][15]. Naïve Bayes algorithm is known to be simple and effective in text classification application and also known as one of the popular data mining techniques in the research field [16][17].

Descriptive feedback from the interns' supervisor is one of the requirements after the completion of the internship period. However, the functional relationship of the gathered attributes to the program being studied is not taken into consideration as to how these significantly affect the development of the program.

The IT curriculum of the university aimed to produce graduates that can effectively communicate to the real work environment and become an IT professionalized individual. Thus, shaping the students with adequate knowledge and skills to compete globally in the Information and Communication Technology is the major goals of the program. Significant information, considering the strength and the areas that need improvement of the IT interns will be determined in the final conduct of the study.

II. METHODOLOGY

The study applied the text mining approach through R programming. The data were uploaded to the R environment. Fig. 1 shows the processes involve in the study.

A. Data

The input of the study was the descriptive feedback of 5 consecutive school years (SY 2014-2015 to 2018-2019) given by the intern's supervisor after the completion of the internship period of the Bachelor of Science in Information Technology under the College of Computer Studies and Information Technology of Southern Leyte State University. The data were cleaned by removing unwanted words like tags, stop words, punctuations, number, and whitespaces. The cleaned data were stem using the stemming function in R to group the features according to the root word.

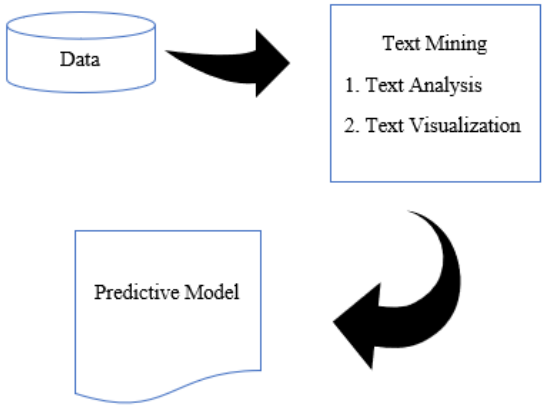


Fig. 1. The steps which involve the text mining application.

B. *Text Mining*

The study applied the text mining application through text analysis and text visualization to determine the strength and the areas that need improvement of the student interns under the program.

C. *Predictive Model*

Given the significance of the Bayes Theorem (1), the study applied the Naïve Bayes text classification wherein the main assumption of the Naïve Bayes classifier is that each feature (word) is independent of each other.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \qquad (1)$$

Bayes Theorem application is effective in analyzing time series dataset and forecasting outcomes [18]. A study also proved that Bayes Theorem application can boost the performance of the developed model [19].

The predictive model implemented the Naïve Bayes classifier algorithm in the text classification. 80% of the total number of observations served as the training set while 20% of the total number of observations served as the test data set.

III. RESULTS AND DISCUSSION

The study applied the text mining approach in R programming.

A. *Text Mining*

The study presented the frequency histogram (Fig. 2) and the study word cloud (Fig. 3) to show the graphical presentation of the results.

1) Text Analysis: The frequency histogram (Fig. 2) shows that the top identified strength are the word "proficient" which has 185 occurrences and the word "technically" with 180 occurrences whereas, the top identified areas that need improvement are the word "skill" with 318 occurrences and the word "communication" with 218 occurrences.

2) Text Visualization: Fig. 3 presents the visual representation of the strength and the areas that need improvement provided by the immediate supervisor of the interns.

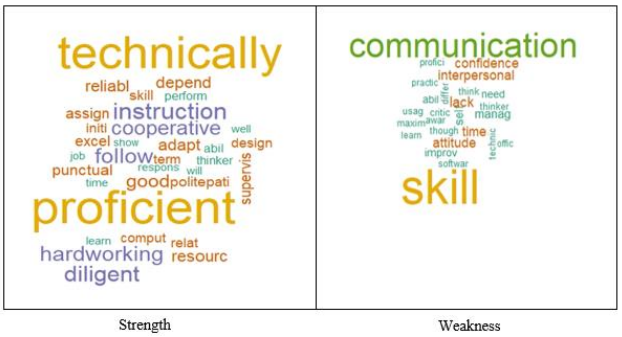


Fig. 3. The Study Word Cloud.

Based on the result, the IT interns were technically proficient. However, there is a need to improve the communication skills of IT students.

Performance feedbacking is necessary to improve the curriculum and surely be refined taking consideration in the areas that needed improvement [20][21].

B. *Predictive Model*

The predictive model was built using the "naïve_bayes" function in R. Table 1 presents the confusion matrix of the study based on the training and test dataset.

The tabular presentation shows that in the training set, 494 and 482 are correct predictions whereas, 1 is the wrong prediction against the actual value with the accuracy rating of 99.89%. In the test data set, the correct predictions are 120 and 97 whereas, 36 is the wrong prediction with the accuracy rating of 85.77%.

Based on the result, the model has an acceptable accuracy rating. Previous studies proved the remarkable indication of the Bayes Theorem application in data mining [22][23][24].

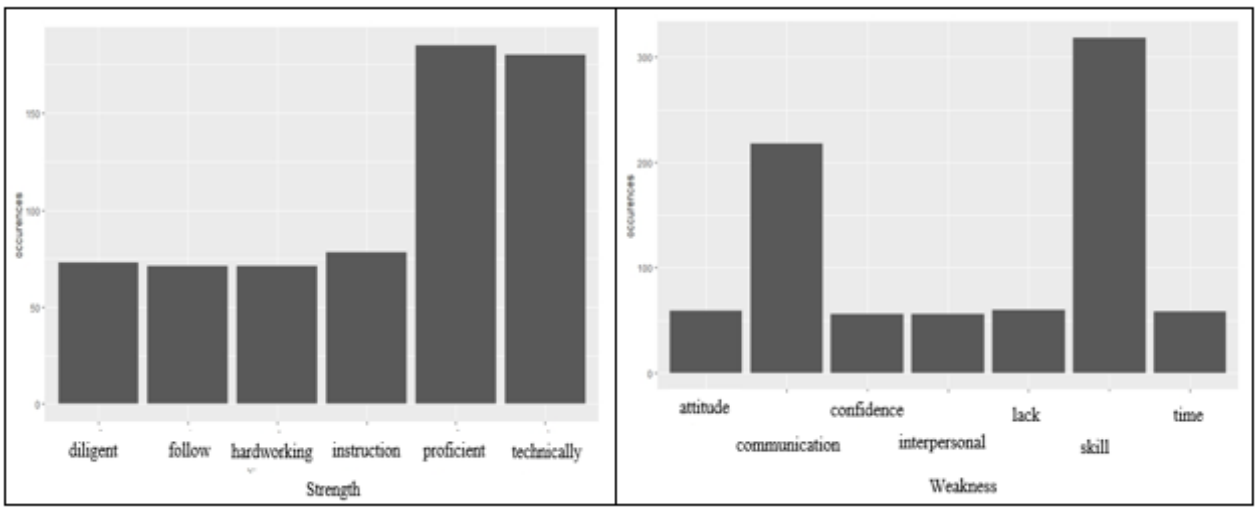


Fig. 2. The Frequency Histogram.

TABLE I. THE CONFUSION MATRIX

Actual	Predictions			
	Strength	***Weak***	***Strength***	***Weak***
Strength	494	0	120	36
Weak	1	482	0	97

IV. CONCLUSION

This study aimed to determine the potentials and the areas that need improvement of the IT students under the program. The findings of the study can help the program to achieve the desired program outcome.

Results revealed that the IT students are proficient in terms of the technical task but not in communication skills. Thus, giving an implication that the IT interns can integrate the IT-based solutions into the user environment effectively. However, there is a need to improve the students' communication skills for them to compete effectively in the labor market.

Based on the results, the IT interns are capable of designing, implementing and evaluating computer-based systems and processes to meet the desired needs of the clients. Furthermore, the communication skills of the IT students should be developed for the students to compete globally as IT professionals that can develop and advance Information and Communication Technology.

REFERENCES

[1] CHED MEMORANDUM ORDER (CMO) No. 104 Series 2017. Revised Guidelines for Student Internship Program in the Philippines.

[2] C. Burgers, A. Eden, M.D. van Engelenburg and S. Buningh, "How feedback boosts motivation and play in a brain-training game," Computers in Human Behavior, 48, 2015, pp.94-103.

[3] S. M. Brookhart, "How to give effective feedback to your students," ASCD; 2017 Mar 10.

[4] J. Carpentier and G.A. Mageau, "Predicting sport experience during training: The role of change-oriented feedback in athletes' motivation, self-confidence and needs satisfaction fluctuations," Journal of sport and exercise psychology. 2016 Feb 1;38(1):pp 45-58.

[5] M. Alam "Internship Report on Compliance Management Policies And Practices Of Designtex Fashions Limited: An Evaluation," (Doctoral dissertation, Daffodil International University).

[6] A. Taye, D. Gurciullo, B. A. Miles, A. Gupta, R. P. Owen, W. B. Inabnet III, J. N. Beyda, and J. L. Marti, "Clinical performance of a next-generation sequencing assay (ThyroSeq v2) in the evaluation of indeterminate thyroid nodules. Surgery.," 2018 Jan 1;163(1):pp 97-103.

[7] L. K. John and L. Eeckhout, "Performance evaluation and benchmarking," CRC Press; 2018 Oct 3.

[8] C. Sides and A. Mrvica, "Internships: Theory and practice," Routledge; 2017 Mar 2.

[9] R. Krishna, Z. Yu, A. Agrawal, M. Dominguez and D. Wolf, "The'BigSE'Project: Lessons Learned from Validating Industrial Text Mining," In2016 IEEE/ACM 2nd International Workshop on Big Data Software Engineering (BIGDSE) 2016 May 16 (pp. 65-71), IEEE.

[10] S. ElAtia, D. Ipperciel, O. R. Zaïane, and editors, "Data mining and learning analytics: Applications in educational research," John Wiley & Sons; 2016 Sep 26.

[11] U. Kursuncu, M. Gaur, U. Lokala, K. Thirunarayan, A. Sheth, I. B. Arpinar, "Predictive Analysis on Twitter: Techniques and Applications," InEmerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining 2019 (pp. 67-104), Springer, Cham.

[12] K. Denecke, "Sentiment Analysis from Medical Texts," InHealth Web Science 2015 (pp. 83-98), Springer, Cham.

[13] W. W. Fleuren and W. Alkema, "Application of text mining in the biomedical domain," *Methods*, *74*, pp.97-106, 2015, Elsevier.

[14] R. Karam, R. Puri, S.Bhunia, "Energy-efficient adaptive hardware accelerator for text mining application kernels," IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2016 May 11;24(12): pp 3526-37.

[15] S. Vijayarani, M. J. Ilamathi, M. Nithya, "Preprocessing techniques for text mining-an overview," International Journal of Computer Science & Communication Networks, 2015 Feb;5(1):pp 7-16.

[16] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification," Engineering Applications of Artificial Intelligence, *52*, 2016, pp.26-39.

[17] B. Tang, H. He, P. M. Baggenstoss, and S. Kay, "A Bayesian classification approach using class-specific features for text categorization," IEEE Transactions on Knowledge and Data Engineering, 28(6), 2016, pp.1602-1606.

[18] A. Pole, M. West, J. Harrison, "Applied Bayesian forecasting and time series analysis," Chapman and Hall/CRC; 2018 Oct 9.

[19] J. Zhao, K. Yang, X. Wei, Y. Ding, L. Hu, and G. Xu. "A heuristic clustering-based task deployment approach for load balancing using Bayes theorem in cloud environment," *IEEE Transactions on Parallel and Distributed Systems* 27, no. 2, 2015, pp 305-316.

[20] J. Zhao, K. Yang, X. Wei, Y. Ding, L. Hu, and G. Xu, "A heuristic clustering-based task deployment approach for load balancing using Bayes theorem in cloud environment," *IEEE Transactions on Parallel and Distributed Systems* 27, no. 2, 2015, pp 305-316.

[21] I. Singh, A. Sai Sabitha, and A. Bansal, "Student performance analysis using clustering algorithm," In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence*)*, pp. 294-299, IEEE, 2016.

[22] B. Tang, H. He, P. M. Baggenstoss, and S. Kay, "A Bayesian classification approach using class-specific features for text categorization," IEEE Transactions on Knowledge and Data Engineering 28, no. 6, 2016, pp 1602-1606.

[23] T. Mahboob, S. Irfan, and A. Karamat, "A machine learning approach for student assessment in E-learning using Quinlan's C4. 5, Naive Bayes and Random Forest algorithms," In 2016 19th International Multi-Topic Conference (INMIC*)*, pp. 1-8, IEEE, 2016.

[24] P. Chandrasekar and K. Qian, "The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier," In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC*)*, vol. 2, pp. 618-619, IEEE, 2016.

Improving Voice Activity Detection by using Denoising-Based Techniques with Convolutional LSTM

Nattapong Kurpukdee, Surasak Boonkla, Vataya Chunwijitras, Phuttapong Sertsi, and Sawit Kasuriya
National Electronics and Computer Technology Center (NECTEC),
National Science and Technology Development Agency (NSTDA),
112 Pahonyothin Road, Pathumthani, 12120, Thailand
E-mail: {nattapong.kurpukdee, surasak.boonkla, vataya.chunwijitra,
phuttapong.sertsi, sawit.kasuriya}@nectec.or.th

***Abstract*—The performance of voice activity detection (VAD) is drastically degraded when observed speech signals are from unseen noisy environments. In this paper, we propose denoising-based VAD to cope with the unseen noises. The proposed VAD system mainly consists of two stages for denoising and speech/non-speech classification. In the first stage, either log-magnitude spectral estimator (LSA) or convolutional long short-term memory neural network autoencoder (CLAE) is applied to eliminate the noises. The convolutional bidirectional long-short-term memory deep neural network (CBLDNN) is employed for the speech/non-speech classification. The results showed that the proposed VAD was better than the baseline. Furthermore, our CLAE tends to outperform the LSA in denoising algorithms when the signal-to-noise ratio is 5dB.**

***Index Terms*—Voice Activity Detection, Convolutional LSTM, DNN, Bidirectional LSTM, Convolutional Autoencoder**

I. Introduction

Voice activity detection (VAD) plays an important role in speech signal processing, since several applications processing speech signals need VAD to classify speech/non-speech segments in an utterance. The performance of the applications such as, automatic speech recognition (ASR) usually depend on the VAD accuracy. However, the VAD performance will be degraded in noisy environments especially in unseen noises. The noises can be stationary or non-stationary and correlated or uncorrelated to speech signals. Then, improving the performance of VAD is still an interesting topic in speech processing.

To tackle with the unpredictable noises, deep neural network (DNN) is frequently employed [1], [2] because the system can handle with several noises simultaneously. There are two types of DNN-based VAD systems, the system with denoise and without denoise stage. The system without denoise stage is usually trained by using noisy speech data like robust speech recognition [1], [2]. On the other hand, it is ordinarily applying the eliminated noise in training data. The denoise stage can be accomplished by using parameter-based or DNN-based techniques. For example, log-magnitude spectral estimator (LSA) is the parameter-based approach [3], [4], and an autoencoder is a DNN-based appraoch [5], [6], [7], [8]. The main benefit of LSA is to reduce uncorrelated noise in speech signal, however its performance will be dramatically dropped the quality of voice is poor (signal-to-noise ratio is near 0 dB).

One benefit of the autoencoder is that it can handle several kinds of noises simultaneously. Recently, convolutional DNN autoencoder is a well-known noise reduction technique, which was proposed in the medical image processing [9], for automatic speech recognition (ASR) systems [10], VAD [11], and gender identification system [12]. There are two types of inputs to train the DNN autoencoder: spectrogram and raw speech data [13], [10]. The results show that the raw speech outperforms the spectrogram [13] [14].

In this paper, we proposed some processes to improve the performance of denoising-based VAD. Firstly, we add convolutional layers in both the autoencoder and the speech/non-speech classifier so that the raw speech data can be used as input. Since long-short term memory (LSTM) is very good at a time-series sequence modelling, it shoud be helpful in both the autoencoder and classifier. In addition, the paper [15] shows that the bidirectional LSTM (BLSTM) is better technique than LSTM for speech recognition. Therefore, our VAD system will employs convolutional layers and LSTM layer so that it can learn raw speech for denoise. It also contains convolutional layers and BLSTM in the speech/non-speech classifier to learn denoised, time-series raw speech.

The remaining of the paper is organized as follows. Section 2 describes the noise reduction and raw waveform feature architectures. The experiments and evaluation results are reported in Section 3. The discussion and conclusion are Sections 4 and 5, respectively.

II. The Proposed VAD System

Our proposed VAD system consists of two parts: denoising and speech/non-speech classification stages, as shown in Fig. 1. Their details are as follows.

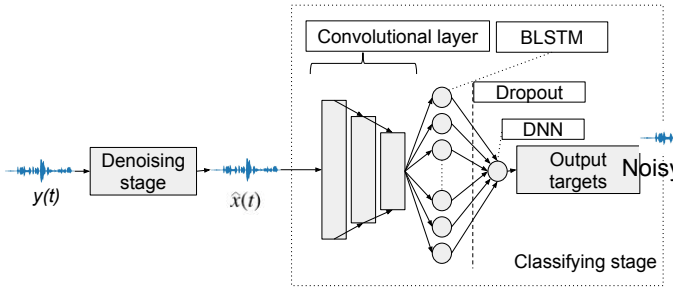
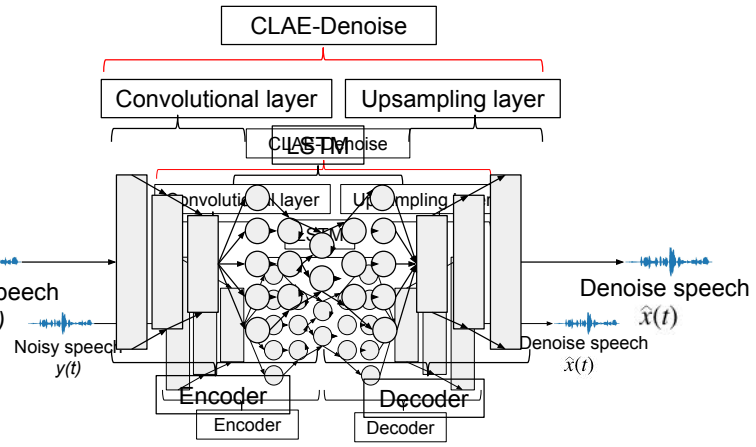
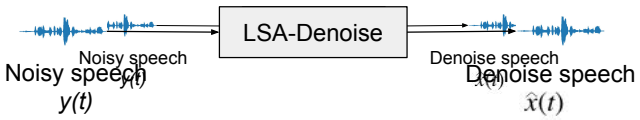


Fig. 1. Modules of raw waveform VAD classifying stage.



(a) CLAE modules



(b) LSA modules

Fig. 2. Modules of raw waveform denoising structure.

A. Denoising

An autoencoder is an artificial neural network that obtains representation of the data after training. The representation is then used in reconstructing the data. The representation is usually what we want such as parameters of speech rather than noises. The training data can be speech features, spectrogram, and raw speech.

Our autoencoder receives noisy speech and outputs the enhanced speech. It contains convolutional layers and LSTM (CLAE) for learning raw speech and reducing noises. The CLAE has mainly three layers: convolutional layer, LSTM layer, and upsampling layer as shown in Fig. 2(a). The first layer functions as an encoder that extracts information and learns from raw waveforms by convolutional and max pooling operations. The second layer employs the LSTM neural network to compress and decompress the information from the previous layer because the LSTM is very helpful in learning a time domain sequences. The final layer is the deconvolution that upsamples the output of LSTM to waveforms again.

The configuration of CLAE is shown in Table I. We used the filter size 3x1 and hyperbolic tangent activation function in the convolution layers of the encoder and decoder. A max pooling layer size of 2x1 and tanh activation function were added between the convolution layers in the encoder. The same parameters of convolutional layers were used for the upsampling layers as well.

TABLE I
DETAILS OF CLAE MODEL FOR CBLDNN MODEL.

CLAE model details	
Encoder	
Convolution layer	3 layers
# features map	$160 \rightarrow 80 \rightarrow 40$
LSTM layer	5 layers
# hidden units	$32 \rightarrow 16 \rightarrow 8 \rightarrow 16 \rightarrow 32$
Decoder	
Convolution layer	3 layers
# features map	$40 \rightarrow 80 \rightarrow 160$
Output layer raw waveform	320-dimension
Total number of parameters	122,513

Besides the autoencoder, we also used the log spectral amplitude estimator (LSA) to reduce noise in our VAD system for comparison to the autoencoder. Assume that the short time Fourier transform of a noisy speech signal is defined as

$$Y(\omega, l) = X(\omega, l) + D(\omega, l), \tag{1}$$

where $Y(\omega, l)$, $X(\omega, l)$, and $D(\omega, l)$ are Fourier transform of a noisy speech $y(t)$, a clean speech $x(t)$, and an additive noise $d(t)$. ω and l are frequency index and frame index. Let $\hat{x}(t)$ be the estimated $x(t)$ from $y(t)$. The magnitude spectrum of $\hat{x}(t)$, $|\hat{X}(\omega, l)|$, is calculated by

$$|\hat{X}(\omega, l)| = G(\omega, l)|Y(\omega, l)|, \tag{2}$$

where $G(\omega, l)$ is the adaptive gain function that depends on the probability of speech presence and absence [3]. The LSA is good at reducing uncorrelated noise to a certain extent of SNR. However, when the SNR is low and noise is correlated to speech, the enhanced speech is considerably distorted.

B. Speech and non-speech classification

The DNN-based VAD is used here as a speech/non-speech classifier. It has LSTM layers and a DNN layer. There are two types of network: feed forward (CLDNN) and bidirectional (CBLDNN) ones. We also add convolutional layers to both of them to learn from raw speech waveforms. The CLDNN is set as a baseline in this paper. The details of the networks are as follows.

The CLDNN consists of 3 convolution layers, one LSTM hidden layers with 256 hidden units, and one DNN layer. The convolution layers and LSTM layers employ the tanh function as the activation function. To reduce the raw waveform size, we did not use a pooling layer in the convolutional layers. However, the convolution layer is done with a stride of 2 to reduce the amount of raw waveform size [16]. To avoid the model over fitting, the output of LSTM is set the dropout rate at 50%. For decide the speech/non-speech, we added one DNN

layer with 1 hidden unit, a sigmoid activation function. The input of this CLDNN is a 320-dimensional raw waveform.

The CBLDNN differs from the CLDNN only at the LSTM layer that use bidirectional network instead of the feed forward network. Their details including the number of parameters are summarized in Table II. Although, the number of parameters of CBLDNN is around two times of the CLDNN, the calculation time does not increase much as the number of parameters, which is shown later in the results.

TABLE II
THE CONFIGURATION OF CLDNN AND CBLDNN MODEL FOR RAW SPEECH WAVEFORM.

Speech/Non-speech classification model details		
Layers	CLDNN	CBLDNN
Convolution layer	3 layer	
# features map	$40 \rightarrow 20 \rightarrow 10$	
RNN layer	LSTM	BLSTM
# hidden units	256	
DNN layer	1 layer	
# hidden units	1	
Total number of parameters	278,935	552,599

III. EXPERIMENTAL DETAILS

A. Experiments and Results

We used both standard clean speech corpus TIMIT [17] for developing the VAD system and our own specific corpus for testing the accuracy. TIMIT corpus contains both training and testing data sets. Ninety percent of the training set was used for training the system and the remaining 10% was used for validation. Both training and testing of TIMIT data were augmented by adding pink, white, bike, crowd, and car noises [18][19] with SNR equal to 30, 15, and 5 dB. Therefor, the training, validation, and testing data were long 13, 2.04, and 4.72 hours, respectively.

There were two sets of the testing data, one was from the augmented TIMIT testing data, which we name as seen noise condition. Another testing data was our corpus that contained noisy speech with unseen noise conditions. It was recorded from a meeting room, classrooms with and without reverberation, outdoor environments with bird singing sound, street environments with car noise, and pink noise. It was 0.5 hours long, 64% of which was speech and the remaining was non-speech.

The ground truth of speech and non-speech classification of the testing data was obtained by using phoneme of TIMIT and manually labeling. In case of TIMIT corpus, the phonemes h#, pau, epi, bcl, dcl, gcl, pcl, tck, and kcl were regarded as non-speech and the others phonemes were regarded as speech. In case of our corpus, we labeled speech/non-speech sections manually by listening and labeling.

In training the system, we train the whole system at once (not separate training of the autoencoder and classifier). The neural networks of both autoencoder and classifier stages was implemented by using Keras toolkit [20] with Tensorflow backend [21].

All speech data in this paper were recorded in 16 KHz sampling rate, 16 bits, and mono channel. They were divided into frames with the frame size 20 ms and frame shift 10 ms. There are mainly two evaluations: evaluation of denoising stage and evaluation of the VAD. The evaluation of VAD has six conditions which are (1) CLDNN : convolutional LSTM without denoising stage, (2) CBLDNN : convolutional BLSTM without denoising stage, (3) CLAE-CLDNN : CLDNN with convolutional LSTM autoencoder denoising alogrithm, (4) CLAE-CBLDNN : CBLDNN with convolutional LSTM autoencoder denoising algorithm, (5) LSA-CLDNN : CLDNN with log spectral amplitude estimator denoising algorithm, and (6) LSA-CBLDNN : CBLDNN with log spectral amplitude estimator denoising algorithm. The performance of our VAD system is measured in term of accuracy (Acc) defined as.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where TP is the number of non-speech frames that are classified as non-speech. TN is the number of speech frames that are classified as speech. FP is the number of non-speech frames that are classified as speech. FN is the number of speech frames that are classified as non-speech frames.

B. Evaluation of deep neural networks

After training the networks, the performance of CLDNN and CBLDNN (without denoise) are shown in Table III. The results of CBLDNN is slightly better than those of CLDNN in both seen and unseen noise condition data. In addition, the maximum improvement is at 5 dB, where the relative accuracy improvement by using the CBLDNN are 0.51% and 0.88% in seen and unseen conditions, respectively. Whereas the accuracy by using the CBLDNN is slightly better than that of CLDNN at 30 and 15 dB.

TABLE III
THE PERFORMANCE OF CLDNN MODEL COMPARED WITH CBLDNN MODEL FOR VAD.

		Acc(%)	
Algorithms	SNR	Seen noise condition	Unseen noise condition
CLDNN	30dB	93.68	84.03
	15dB	92.95	72.79
	5dB	92.01	67.62
	Avg	92.93	77.34
CBLDNN	30dB	93.80	84.34
	15dB	93.28	73.03
	5dB	92.52	68.50
	Avg	**93.24**	**77.74**

C. Evaluation of denoising algorithm

Before evaluation of our denoising-based VAD system, we evaluated the performance of the denoise stage. In training the autoencoder, the learning loss rate was reduced as less as possible close to zero. The best learning iteration was chosen among the last ten iterations when there was no further reduction of the loss rate. We evaluated the performance of denoising by enhancing 500 noisy audio files of seen noisy

condition (noisy TIMIT corpus). The techniques that we use to investigate the performance of denoise algorithm consist of Perceptual Evaluation of Speech Quality (PESQ) [22], speech distortion index, and noise reduction index [23]. The speech distortion or noise reduction index is calculated by

$$I = \frac{1}{N}\sum_{n=1}^{N}\frac{|\hat{x}_n(t) - s_{n,ref}(t)|}{|s_{n,ref}(t)|} * 100, \quad (4)$$

where N is the total number of speech signals, and $s_{n,ref}(t)$ is the corresponding referenced speech signal. $s_{n,ref}(t)$ is the noisy speech signal $y_n(t)$ when we calculated the noise reduction index, whereas $s_{n,ref}(t)$ is the clean speech signal $x_n(t)$ when we calculated the speech distortion index. The evaluation was the comparison between our CLAE and the LSA. The results are shown in Table IV. We can see that the PESQ is slightly increased after noise reduction by using both CLAE and LSA. When SNR is 30 and 15 dB, the LSA is efficient in noise reduction because of less speech distortion and high noise reduction. The CLAE performs higher percentage in noise reduction, but introducing more distortion. However, when SNR is 5 dB, CLAE is better than the LSA.

TABLE IV
THE PERFORMANCE OF DENOISING ALGORITHM.

SNR		Noisy speech	CLAE	LSA
30dB	Reduction(%)	-	15.28	3.99
	Distortion (%)	-	14.81	4.05
	PESQ	3.62	3.70	3.79
15dB	Reduction(%)	-	28.54	21.64
	Distortion (%)	-	20.51	15.45
	PESQ	2.66	3.02	3.05
5dB	Reduction(%)	-	**58.45**	52.56
	Distortion (%)	-	**35.45**	36.26
	PESQ	1.94	**2.48**	2.36

D. Evaluation of the proposed VAD

We evaluated the performance of our denoising-based VAD system by using both LSA and CLAE. The results are shown in Table V. The best average accuracy is 92.42% when LSA-CBLDNN was used in case of seen noise condition data. The LSA-CBLDNN also gave the best accuracy in case of unseen noise condition data, 80.97%. Without denoising, CBLDNN (without denoising) yielded the accuracy 93.24%, which is 0.82% higher than the LSA-CBLDNN (after denoising) in case of seen noise condition data. However, the LSA-CBLDNN gave 3.23% higher than the CBLDNN in case of unseen noise condition data. This indicates the success in our goal.

E. Evaluation of the VAD processing time

We have seen that the number of parameters of the CBLDNN is as twice as that of the CLDNN. We have investigated the processing time of the system by using all examples in unseen noise condition dataset. We used a personal computer having CPU speed 3.40GHz and 8G of RAM running on Ubuntu operating systems. Table VI shows the

TABLE V
ACCURACY OF THE DIFFERENT MODELS WITH RAW WAVEFORM FEATURE.

		Acc(%)	
Algorithms	SNR	Seen noise condition	Unseen noise condition
CLAE-CLDNN	30dB	93.31	85.59
	15dB	92.11	73.49
	5dB	90.32	67.64
	Avg	92.00	78.32
CLAE-CBLDNN	30dB	93.21	85.18
	15dB	92.24	73.50
	5dB	90.79	68.28
	Avg	92.15	78.25
LSA-CLDNN	30dB	93.60	86.75
	15dB	92.72	77.33
	5dB	89.44	69.73
	Avg	92.12	80.45
LSA-CBLDNN	30dB	93.89	87.60
	15dB	92.84	77.60
	5dB	90.25	69.84
	Avg	**92.42**	**80.97**

average processing time corresponsing to six conditions. The input is the total duration of the testing data.

The number of parameters of CBLDNN is 1.98 times larger than that of the CLDNN. But the processing time of CBLDNN is 1.49 times slower than CLDNN. Although the processing time by using the CLAE is 1.51 times is longer than that of the LSA, it is worthy using the CLAE because it can add more noise training data to handle more noises than the LSA. One of the example is shown in Fig. 3 that the CLEA can restore speech when the SNR is very low.

TABLE VI
THE PROCESSING TIME OF OUR VADS SYSTEMS.

		Average processing time (s)	
Algorithms	Input	Denoise	Classification
CLDNN	1855.32	-	67.13
CBLDNN		-	100.63
CLAE-CLDNN		151.82	68.73
CLAE-CBLDNN			101.65
LSA-CLDNN		100.03	68.45
LSA-CBLDNN			101.44

IV. DISCUSSION

According to the results in Table III, modifying the LSTM to bidirectional LSTM could improve the speech/non-speech classification for both seen and unseen noise condition data. In Table V, applying denoising algorithms to VAD system could improve the accuracy of VAD by 1%-3%, compared with Table III, in case of unseen noise condition data.

After denoising, the average accuracy by using the LSA-CBLDNN is hither than those by using the CLAE-CLDNN, CLAE-CBLDNN and LSA-CLDNN in both seen and unseen noise condition data. This mean that using the LSA is superior to our CLAE. We investigated the performance of noise reduction as shown in Table IV. The performance of our CLAE is superior to the LSA when the SNR is 5 dB in case of

seen noise condition data. This corresponds to the accuracy in Table V at 5 dB that the CLAE-CBLDNN is superior to the LSA-CBLDNN in case of seen noise condition data. Figure 3, emphasizes the performance of our CLAE that it can restore the speech signal when the SNR is low, compared with the LSA. We predict that insufficient training data causes the performance of CLAE-CBLDNN lower than that of LSA-CBLDNN when the testing data is unseen. This problem could be easily solved in the future and we can construct the high accuracy VAD system by using the CLAE-CBLDNN.

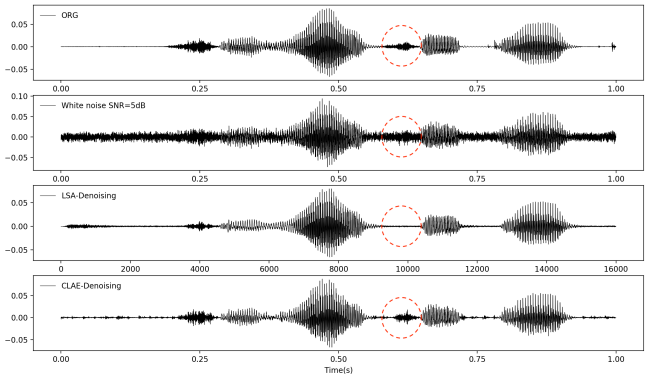


Fig. 3. The performance of noise reduction in a seen noise condition data by added white noise at SNR equal to 5dB.

V. Conclusion

In this paper, we presented the denoising-based VAD system trained with raw waveform for unseen environments. Our system had a denoising stage and a speech/non-speech classifying state. Two algorithms namely, convolutional long-short term memory autoencoder and log spectral amplitude estimator were employed for denoising. The classifying stage was convolutional bidirectional long-short term memory deep neural network. The evaluation results showed that the bidirectional long-short term memory based VAD was better than the normal long-short term memory one. The accuracy after denoising by using our autoencoder yielded better performance than using the log magnitude spectrum estimator when the signal to noise ratio was low in case of seen noise condition data.

References

[1] Y. Tachioka, "Dnn-based voice activity detection using auxiliary speech models in noisy environments," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 5529–5533.

[2] Ivan Tashev and Seyedmahdad Mirsamadi, "Dnn-based causal voice activity detector," February 2016, University of California - San Diego.

[3] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113–116, April 2002.

[4] C. H. You, B. Ma, and C. Ni, "Modification on lsa speech enhancement for speech recognition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 5475–5479.

[5] Y. Xu, J. Du, L. Dai, and C. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65–68, Jan 2014.

[6] Y. Zhao, D. Wang, I. Merks, and T. Zhang, "Dnn-based enhancement of noisy and reverberant speech," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6525–6529.

[7] O. Plchot, L. Burget, H. Aronowitz, and P. Matjka, "Audio enhancing with dnn autoencoder for speaker recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 5090–5094.

[8] K. H. Lee, S. J. Kang, W. H. Kang, and N. S. Kim, "Two-stage noise aware training using asymmetric deep denoising autoencoder," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 5765–5769.

[9] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, Dec 2016, pp. 241–246.

[10] X. Feng, Y. Zhang, and J. Glass, "Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1759–1763.

[11] X. Zhang and J. Wu, "Denoising deep neural networks based voice activity detection," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 853–857.

[12] Jilt Sebastian, Manoj Kumar, Pavan Kumar D. S., Mathew Magimai-Doss, Hema A. Murthy, and Shrikanth Narayanan, "Denoising and raw-waveform networks for weakly-supervised gender identification on noisy speech," in *INTERSPEECH*. 2018, pp. 292–296, ISCA.

[13] S. Fu, Y. Tsao, X. Lu, and H. Kawai, "Raw waveform-based speech enhancement by fully convolutional networks," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec 2017, pp. 006–012.

[14] Rubén Zazo, Tara N. Sainath, Gabor Simko, and Carolina Parada, "Feature learning with raw-waveform cldnns for voice activity detection," in *INTERSPEECH*. 2016, pp. 3668–3672, ISCA.

[15] D. Neiberg, K. Elenius, and K. Laskowski, "Improving latency-controlled blstm acoustic models for online speech recognition," in *INTERSPEECH*. 2006, pp. 809–812, ISCA.

[16] A. Sehgal and N. Kehtarnavaz, "A convolutional neural network smartphone app for real-time voice activity detection," *IEEE Access*, vol. 6, pp. 9017–9026, 2018.

[17] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "Darpa timit acoustic phonetic continuous speech corpus cdrom," 1993.

[18] Pete Warden, "Speech commands: A public dataset for single-word speech recognition.," *Dataset available from http://download.tensorflow.org/data/ speech_commands_v0.01.tar.gz*, 2017.

[19] Andrew Varga and Herman J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, jul 1993.

[20] François Chollet et al., "Keras," https://keras.io, 2015.

[21] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, Software available from tensorflow.org.

[22] Philipos C. Loizou, *Speech Enhancement: Theory and Practice*, Signal Processing and Communications, 2007.

[23] Xugang Lu, Yu Tsao, Shigeki Matsuda, and Chiori Hori, "Speech enhancement based on deep denoising autoencoder," in *INTERSPEECH*, 2013.

Semantic Enhancement and Multi-level Label Embedding for Chinese News Headline Classification

Jiangnan Qi
Lab of Social Intelligence and Complex Data Processing
School of Software Engineering Xi'an Jiao Tong University
xi'an, China
jiangnanqi2@163.com

Yuan Rao
Lab of Social Intelligence and Complex Data Processing
School of Software Engineering Xi'an Jiao Tong University
xi'an, China
yuanrao@163.com

Ling Sun
Lab of Social Intelligence and Complex Data Processing
School of Software Engineering Xi'an Jiao Tong University
xi'an, China
sunling@stu.xjtu.edu.cn

Xiong Yang
Lab of Social Intelligence and Complex Data Processing
School of Software Engineering Xi'an Jiao Tong University
xi'an, China
youngpanda@stu.xjtu.edu.cn

***Abstract*—News headline classification is a specific example of short text classification, which aims to extract semantic information from the short text and classify it accurately. It can provide a fast classification method for data of various kinds of news media, thus arousing the common concern of academia and industry. Most short text classification methods are based on the semantic expansion of external knowledge, which is unable to expansion dynamically in real time and make full use of label information. To overcome these problems, we propose a novel method which consists of three parts: semantic enhancement, multi-dimensional feature fusion network and multi-level label embedding. Firstly, the word-level semantic information are embedded into the character encoding from pre-train model to enhance semantic features. Secondly, both of Bi-GRU and multi-scale CNN are used to extract sequence and local features of text to enhance the semantic representation of the sentence. Furthermore, the multi-level label embedding is used to filter textual vector and assist classification in the word and sentence level respectively. Experimental results on NLPCC 2017 Chinese news headline classification task show that our model achieves 84.74% of accuracy and 84.75% of F1, improves over the best baseline model by 1.5% and 1.6%, respectively, and reaches the state-of-the-art performance.**

***Index Terms*—News headlines classification, multi-level label embedding, semantic enhance, multi-dimensional feature fusion.**

I. Introduction

News headline classification is a special task of short text classification, its purpose is to quickly and accurately determine the categories of news from the short headline expressions which distills the full text semantic information. However, news headlines are characterized by inherent sparsity, wide range, strong real-time, various styles and non-standard features, which seriously affect the performance of text classification. How to improve the performance of news headline classification according to the semantic features existing in news headlines not only can be widely used in important fields such as vertical search, news recommendation and public opinion analysis, but also attracted wide and high attention from academia and industry [1].Many researchers adopt external knowledge to expand semantic features of short text, such as semantic expansion based on the existing databases [2], [3]. But it needs to be constructed artificially and cannot be updated timely, so unable to meet the needs of real-time performance. Besides of this, traditional machine learning algorithms are used to get knowledge to expand text semantic representation [4]–[6].Due to the sparse features of short texts, such methods always introduce noise information and cannot extract information accurately.

In order to satisfy the classification requirement of high real-time and accuracy, some researchers have adopted some Deep Learning algorithms, such as Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN), which have made some breakthroughs by mining the deep semantic feature of text itself without introducing external knowledge and achieve some state-of-arts performances. Based on the shortcoming of RNN's long-term dependence, Baoxin Wang [7] proposed a kind of disconnected recurrent neural network to extract the sequence information of text, but too large or small block size will seriously influence the result of text semantic extraction. In addition, Conneau et al. [8] proposed a classification method of very deep convolutional networks to extract the local information on the text. Because of the disadvantage of complex network and time-consuming training, Jonson et al. [9] further proposed a pyramid convolutional neural network. But, both of these two kinds of methods only extract the partial features of text from the sequential information or local information. Furthermore, Vaswani et al. [10] adopted the attention mechanism to create a Transformer

Network, which can extract a sort of scale and granularity feature from the words or characters of the sentence. In addition, Wang et al. [11] proposed a strategy of attention mechanism with label embedding to enhance the semantic representation of the text, but this method has been not applied into the final text classification strategy. To solve this problem, Kim et al. [12] and Demirel et al. [13] use all classification tags and sentence representation to calculate similarity, but it convert a multi-classification problem into many binary classification problems, which aggravate the complexity of classifier design and implementation.

Based on the above considerations, we propose a new Semantic Enhancement and Multi-level Label Embedding Model (SEMLE), which provides an end-to-end solution without any external knowledge. Firstly, the relevant corpus is used to fine-tune the parameters of the BERT model [14] for satisfying the field of the news. Then, we utilize traditional text representation methods to expand BERT's text information encoding for semantic fusion both of characters and words. Besides, multi-scale CNN and Bi-GRU are adopted to enhance the high-level and deep semantic representation of text. Moreover, the labels containing semantic information will replace the traditional One-Hot text encoding method, which can filter the text representation and assist the classification decision at the word level and sentence level respectively. Finally, the Softmax activation function is used to classify the text. Experimental results on NLPCC 2017 Chinese news headline classification task show our model achieved 84.74% of accuracy and 84.75% of F1, reaching the state-of-the-art performance.

II. RELATED WORK

A. Semantic Representation

In general, Word2Vec. [15] and Glove [16] provided great contribution to represent the deep semantic of words. But both of these two methods use shallow neural network, which cannot be adjusted during the process of training. Furthermore, Peters et al. [17] introduced a pre-trained deep bidirectional language model, named ELMO, for supervised NLP tasks, which build a fundamental work for BERT Model. Yet, for Chinese text, BERT only considered the contextual meaning of characters, but did not get word-level semantic information. Kim et al. [12] illustrated a novel method to train characters in word with the Bi-LSTM model to generate word vector, and then stitched them with the original word vector together as one to expand the semantic of words. This method enhances the semantic representation of text at the same granularity level, when we add the word vector into the character vector to realize the disambiguation of word's semantic information at different context.

Due to the sequence features will lost over long distance in RNN model, Wang et al. [7] proposed a DRNN stacking method for improvement in the semantic representation of sentence-level text. In addition, a multi-layer Bi-LSTM method [12] and deep convolutional neural network [8] have been introduced into the text classification task. Johnson et al. [9] provided a novel pyramid network to extract sentence-level semantics based on deep convolutional network. Moreover, the methods above-mentioned only use the optimized model based on CNN or RNN to extract local features or sequence features, respectively. In this paper, we firstly construct a framework to merge the multi-scale features based on local and sequence features of text into the sentence vector representation, and then the label vector is used as a complemental feature to realize accuracy classification of the news.

B. Label Embedding

In text classification task, One-Hot vector is usually used to encode label [7]–[9], [18], but this method neglects the semantic information of the label, which can help us to screen out the irrelevant information and to assist classification of the text. So, how to utilize the semantic information of label to build a novel classification model has become a critical research work in recent years. Xu et al. [18] suggested to adjust the weight of the LSTM network used these labels and make the network enlightening. Wang et al. [11] embedded text and label into the same vector space, and calculated the weight of text through the correlation between text and label, so as to distinguish the importance of text in the classification task. But in above-mentioned methods, the label only serves as an auxiliary text representation. Kim et al. [12] and Zhang et al. [19] extracted the relationship between sentences and label by similarity calculation and then calculated the similarity score for classification. The classification strategy is changed in above methods, but there is still a lot of noise at the text representation stage. In this paper, we attempt to construct a multi-level label embedding strategy to represent as a word-level vector and sentence-level text for news classification.

C. Pre-train Model and Domain Adopt

Before the pre-training model was proposed, task-specific neural network architecture dominated the research and application in natural language processing. With the increasing of computing capability and development of deep learning, pre-train language representation models greatly improved the performance of various tasks. Bert [14] is one of the key example of contextualized representation learning, which adopts a fine-tuning approach that doesn't require specific architecture of every terminal task. It learns from data, such as Wikipedia, in a way that avoids too much prior knowledge, thus improving the evaluation index of many tasks, exceeding the performance of ELMo [17] and GPT [20], and being widely applied. But fine-tuning BERT directly on the end task with limited tuning data faces both domain and task awareness challenge. To reduce the burden of the end tasks, Xu H et al. [21] proposed to fine-tune BERT model on certain domain corpus to make the model have domain characteristics, achieved obvious results.

III. THE PROPOSED MODEL

The model we introduced is SEMLE model, which is a new end-to-end solution with semantic enhancement and multi-level label embedding in a hybrid deep neural network. The

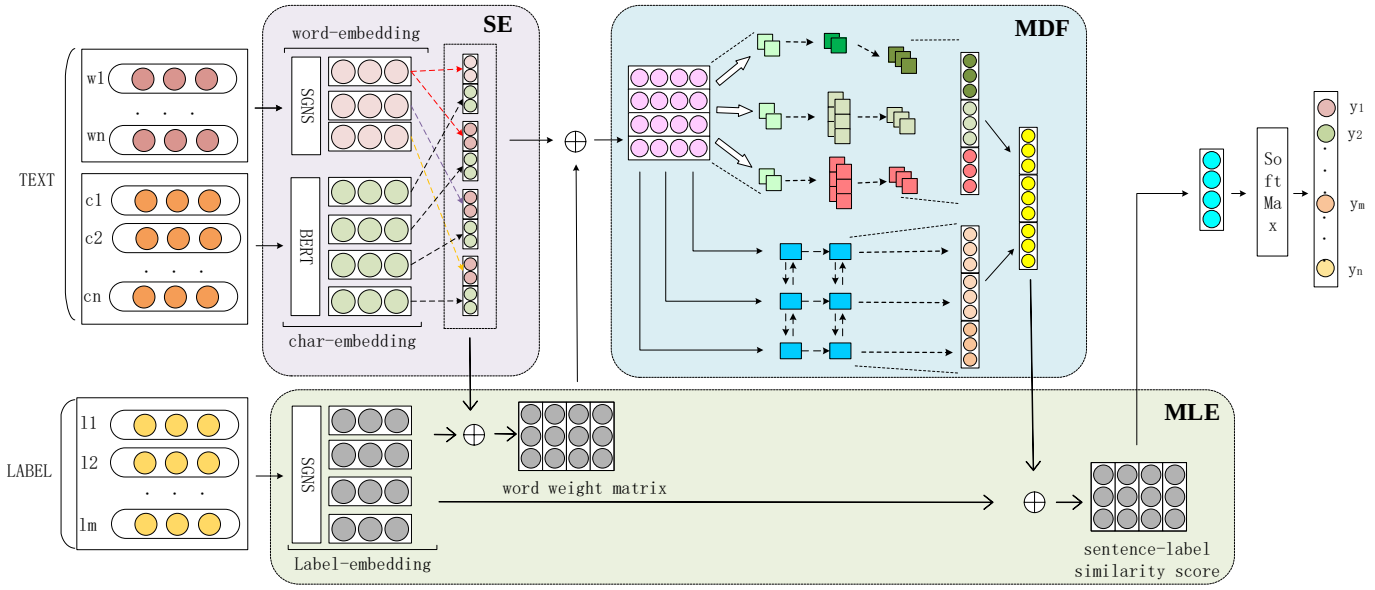


Fig. 1. Architecture of the SEMLE model. Where SE means semantic enhancement, MDF denotes the multi-dimensional feature fusion network and MLE denotes the multi-level label embedding.

whole framework of the model includes three components, such as semantic enhancement, multi-dimensional feature fusion network and multi-level label embedding, and illustrates in Figure 1. The detailed information will be designed as follow:

A. Semantic Enhancement

The goal of this module is to integrate the features of characters and words in the text together, so that each character can be embedded with the deep semantic information of the corresponding word. For a given sentence, the formal expressions of this sentence by the set of characters or words are shown as follow: $Sentence\_char = \{c_1, c_2, ..., c_n\}$ and $Sentence\_word = \{w_1, w_2, ..., w_m\}$, where n and m denote the number of characters and words in the sentence respectively, and $m \leq n$. Further, a fine-tune BERT is utilized as a mapping: ϕ: $R^n \rightarrow R^n \times R^d$ which represents a sentence mapping from n-dimension vector of characters into a $n \times d$ dimension embedding representation in character-level. Meanwhile, we donote a SGNS [22] model as a mapping φ: $R^m \rightarrow R^m \times R^{d'}$, which maps a m-dimension vector of words in one sentence into a $m \times d'$ dimensional embedding representation in word-level. The embedding representation in character-level and word-level of sentences are defined as follows:

$$v_{sentence\_char} = \phi(Sentence\_char) \tag{1}$$

$$v_{sentence\_word} = \varphi(Sentence\_word) \tag{2}$$

Then, each word representation can be embedded into the character representation of this word for semantic enhancement. The enhancement character-embedded representation is:

$$v_{enhance} = \{v_{e1}, v_{e2}, ..., v_{en}\} \tag{3}$$

where, v_{ei} denotes an enhancement vector representation of character embedding, which is composed of the character embedding and related word embedding. For example, there are words of "the media", the character vector and word vector are expressed as $\{v_{c1}, v_{c2}\}$ and $\{v_w\}$, respectively. After the processing of semantic enhancement, it can be represented as: $\{v_{e1}, v_{e2}\} = \{v_{c1} + v_w, v_{c2} + v_w\}$, where, "+" means an operator of joint.

B. Multi-dimensional Feature Fusion Network

Each sentence always is sequence of characters or words. Due to the operator of semantic enhancement, the sequential feature of every word reflecting deep semantic in context is weakened. Considering the dependency behavior among different characters in one sentence, a series of RNN models have been used to extract the sequence features [7], [12]. Accordingly, the bidirectional GRU is selected to enhance semantic representations for sentence vectors. The calculation strategy of a GRU at the step t are listed as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{4}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{5}$$

$$g_t = \tanh(W_h x_t + U_h(r_t \circ h_{t-1})) \tag{6}$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ g_t \tag{7}$$

Where,x_t denotes the input data at time t. h_t and h_{t-1} are the hidden units at time t and t-1, respectively. z_t is update gate. r_t is reset gate. σ,tanh,$\circ$ denote the sigmod activation function, the tanh activation function and element-wise multiplication. In bidirectional GRU, the forward and backward outputs of the hidden layer are aggregated together for uniform definition about the output of text sequence features:

$$v_{output\_gru} = [\overrightarrow{h_t} : \overleftarrow{h_t}] \tag{8}$$

In the task of short text classification, some characters or words in sentences have significant influence on the quality of classification, so how to measure the effect about local features in text will bring a new opportunity for optimizing the capability of the algorithm. Inspired by Inception Network [23] in computer vision, we used three sets of CNN with different size of convolution kernel for extraction n-gram features. The first CNN stack has two convolution kernels and both size are 1. The convolution kernel sizes of second CNN stack are 1 and 3 respectively. The third one is the same as the second one, because the length of most words in text is usually less than 5. all convolution modules will take one-dimensional convolution and a selectively add activation functions after convolution. Finally, local features are expressed as:

$$v_{cnn} = [C^1 : C^3 : C^3] \tag{9}$$

Where, C^i is the output of convolutional neural network with convolution kernel size i. In order to represent more text features and make the sentences have both sequence features and local features, we aggregate the representation results of Bi-GRU and multi-scale CNN together to form the final sentence representation:

$$v_{abstract\_sent} = [v_{cnn} : v_{output\_gru}] \tag{10}$$

C. Multi-level Label Embedding

In text classification, the classification labels have strong semantic information and play a critical role for directly guiding on the classification of contents [19]–[21]. In this paper, the labeled information is used for character filtering and auxiliary classification at the word-level and sentence-level respectively.

Word-level Label Embedding: Due to amount of useless words for text classification in one sentence, some noise data will be introduced when the word is embedding to generate a word vector. Once these word vectors are directly used in the downstream task, it will not only consume the huge resources of computation but also influence the performance of classification. While, the category labels can be used to screen information closely related to classification and formalized definition is represented as follow: $Label = \{l_1, l_2, ..., lnum\}$, where num denotes the number of labels. Therefore, the label embedding can be defined as:

$$v_{label} = \varphi(Label) \tag{11}$$

The compatibility of label-character pairs can be measured by the cosine similarity:

$$M_{weight\_char} = v_{enhance} \cdot v_{label}^T \tag{12}$$

Where, M_{weight_char} means a weight matrix with size of $n \times num$, $v_{enhance}$ means an enhancement vector representation, v_{label}^T means label embeddings, $\cdot$ denotes dot multiplication. When we add the label weights into each corresponding character embeddings, then the character weight can be calculated:

$$\alpha_i = \sum_{j=1}^{num} M_{weight\_char}[i][j] \tag{13}$$

$$\alpha = [\alpha_1, \alpha_2, ..., \alpha_n] \tag{14}$$

Where, α_i denotes the weight of the i-th character. α denotes the normalized weight matrix of text representation:

$$\alpha_i = \frac{\alpha_i}{\max(\alpha)} \tag{15}$$

The text embedding representation can be calculated by multiplication between the character embedding and the normalized weight matrix:

$$v_{update} = v_{enhance} \circ \alpha \tag{16}$$

Where, v_{update} means a character representation vector filtered by labels, $\circ$ denotes an element-wise multiplication. Meanwhile, We will take advantage of v_{update} as input to the Multi-dimensional Feature Fusion Network instead of $v_{enhance}$.

Sentence-level Label Embedding: In general, the sentence representation will be compressed in same dimension as the number of classification labels by a feedforward neural network. However, this kind of method is easy to introduce noise in the compression process without prior guidance. To solve the above problem, we propose a similar calculation method of label-sentence pairs to replace feedforward neural network:

$$v_{similarity} = v_{label} \cdot v_{abstract\_sent} \tag{17}$$

Where, $v_{similarity}$ is similarity score of corresponding label, v_{label} means label embeddings, $v_{abstract\_sent}$ means the output of the Multi-dimensional Feature Fusion Network, $\cdot$ means multiplication of corresponding vectors.

Finally, we employ Softmax activation function to obtain the classification label:

$$P(\tilde{y_i}|sentence) = Softmax(v_{similarity}) \tag{18}$$

$$\hat{y} = \arg\max P(\tilde{y_i}|sentence) \tag{19}$$

Where, $\hat{y}$ denotes a prediction label. $P(\tilde{y_i}|sentence)$ means the probability probability of the i-th label in the sentence. $Softmax$ is softmax activation function.

D. Model Training

Due to the classification of news headline is discrete, we employ a multi-label cross entropy as the loss function to calculate the whole loss:

$$Loss = -\sum_{i}^{N} \sum_{j}^{num} y_j^{(i)} \log \hat{y}_j^{(i)} + \lambda ||\theta|| \tag{20}$$

Where, N denotes the size of the training set, y denotes the markup label of each category, and θ denotes all the training parameters. Cross entropy is used to describe the difference between the predicted value and the real value of the model, and the stochastic gradient descent algorithm is used to optimize and adjust the model parameters.

IV. EXPERIMENTS

A. DataSet Introduction

To evaluate the performance of our model in news headline classification, we executed a series of experiments on Chinese news headline categorization dataset [1] derived from Task 2 of NLPCC 2017, which collected huge amount of news headlines with 18 categories from several Chinese news websites, such as Toutiao, Sina and so on. The label of all categories are shown in Table I. The number of characters in most titles is less than 40, with 21.05 the average. Headlines are even shorter, most of which are less than 20, with 12.07 the average. In this paper, the training set contains 156,000 pieces of data, and the development set and test set contain 36,000 headline of news, respectively. However, this corpus is an imbalance data set with different quantity of news in different category. Especially, there are only 4,000 pieces of training set in the four categories of essay, story, regimen and discovery, but 10,000 pieces of other data sets.

In addition, we also adopted the same evaluation methods defined by NLPCC 2017, including macro-averaged precision, recall and F1. Meanwhile, we add the accuracy as a new evaluation index.

TABLE I
NEWS DATASET LABEL

Game	Essay	Baby	Regimen	Fashion	History
Military	World	Society	Travel	Food	Discovery
Car	Entertainment	Finance	Story	Tech	Sports

B. Candidate Models for Comparison

To compare with our framework, we selected two types of benchmark methods: the baseline methods provided by NLPCC 2017 and some new classification method proposed by the researchers later.

- **LSTM** [1]: a bidirectional LSTM network with pre-training language models by GloVe for word embedding.
- **CNN** [1]: a deep CNN networks with convolution core size of different scales.
- **NBOW** [24]: a neural Bag-of-Words model containing the task specific word importance weights
- **FastText** [25]: a FastText network combined with pre-training language models to enhance semantic representation.
- **Fusion System** [26]: a text classification method based on GRU and multi-mode binary classification voting mechanism.
- **Expand System** [25]: a semantic enhancement method for extending keywords in domain-specific datasets.
- **STCKA** [27]: a method of enhancing the semantic representation of short text by retrieving knowledge from external knowledge sources..
- **CEA** [28]: a Category Expert Attention matrix to extract sentence features from different category perspectives.

C. Implementation Details

In this paper, we select a public pre-training deep representation language model, named BERT-Base, Chinese, which has 12 layers, 768 hidden size, 12 attention heads per layer and 110M parameters. The dimension of the SNGS embedding are 768 and the length of sentence is 64. In a multi-scale convolution module, the channel numbers on CNN with 1 × 1 kernel size are 128, 64 and 32, respectively; the channel numbers of the other group CNN with 3×3 kernel size are 192 and 96. The hidden size of the bidirectional LSTM is 768. During the training process, the Adam optimizer is selected and the learning and dropout rates are divided into two parts: 5e-5, 0.1 in BERT module, and 2e-4,0.5 in other modules. Besides, we set the batch size of the train, validate and test to 32, 8 and 8 respectively, and the epoch ranges from 0 to 10.

Additionally, all experiments in the paper are completed on the open source framework PyTorch on Linux CentOS 7.6.1810 system with NVIDIA TITAN Xp GPU (12G graphics memory).

D. Results And Analysis

In TABLE II, the results of the comparison experiments show that our model, SEMLE, obtains 84.88%, 84.7%, 84.74% and 84.75% in macro-averaged precision, recall, F1 and accuracy, respectively. Compared to the best method of benchmark, our method boosts 1.5% F1 and 1.6% accuracy and achieves the state-of-the-art performance.

Some neural networks models, such as GRU and CNN, have shown that they have the capability for automatically encoding the text in sequence and local features and improve the performance of text classification. But, these models with shallow network can not represent the deep semantic information, so it only widely can be used as basic module for extracting text features. Since NBOW model considered the statistical information of words, the performance of classification is greatly improved than the traditional models of LSTM and CNN. Furthermore, a fusion System, based on LSTM and CNN, is to merge two kinds of text feature together and to form a classification framework, which designs a multi-mode binary classification voting mechanism to improve classification performance. Due to this strategy is too simple, the voting mechanism cannot optimize the final performance greatly. Compared with above-mentioned models, the model of semantic enhancement network at the word level achieved better results than FastText and Expand System. The external knowledge introduced some noise, but corresponding module is not designed in the downstream task to reduce the classification error caused by it. In addition, both of STCKA and CEA networks use the attention mechanism to filter the noise by adjusting the weight and extend the text representation. In addition, the method of multi-angle classification is introduced in CEA model to extract classification features from multiple aspects, which model had achieved the best classification performance before we propose the SEMLT model.

Our method obtains the best performance in this field so far, which may be derived from three parts as follow: (1)

TABLE II
COMPARISON OF EXPERIMENTAL RESULTS

Model	Macro P	Macro R	Macro F	Accuracy
LSTM	77.5	76.8	77.1	76.8
CNN	79.0	78.4	78.7	78.4
NBOW	79.7	79.0	78.7	78.4
Fasttext	81.0	80.5	80.8	80.5
Fusion System	81.41	81.14	81.16	-
Expand System	83.20	83.10	83.10	83.10
STCKA	-	-	-	80.11
CEA	84.12	83.34	83.21	83.14
SEMLE	**84.88**	**84.75**	**84.74**	**84.75**

enhancement embedding for semantic representation of text. We utilize BERT model to obtain the character vector and SGNS model to obtain the word vector in same sentence, and then the word vector should be embedded into the character vectors to enhance semantic features of short text. (2) In the phase of semantic extraction, local features and sequence features of the text are extracted by multi-scales CNN model and Bi-GRU model, respectively. The enhancement semantic information of sentences can be expressed by integrating these features. (3) We propose a creative strategy of multi-level label embedding. The category labels, with strong semantic feature, can enhance text representation at word level. Then, the similarity score between sentence and label is used to optimize the classification strategy at the sentence level.

Meanwhile, due to the imbalance of data set, the performance of classification about each category is discussed in this section. The experimental results show that the classification quality of all categories to achieve the state-of-arts about all the measures in the dataset, especially, there are four kinds of categories, such as Essay, Story, Regimen and Discovery, who contain less data than others but the performers is not the worst. The F1 value on each category by our model's is shown in Figure 2. On the contrary, the two categories, i.e., World and Society, have amount of data, but the performance of classification is somewhat lower than others. Therefore, our model have a certain robustness to adapt the situation of data imbalance. While, the reason of the lower performance may be derived from the smaller data scope. To comprehensively learn all the semantic information of the category labels are impossible.

To verify the convergence of the model, we analyze the influence of F1 value waved with epochs and learning rate. The experimental results are shown in Figure 3. In the training phase, we design three groups of experiments, i.e., the larger and the smaller learning rate, and the hierarchical learning rate, to analyze the influence of learning rate on F1 value. The large learning rate makes the experimental results fluctuate greatly, because the pre-training model already has prior knowledge. Although the F1 value continues rise with the small learning rate, but the speed of rise is become slower and slower. Our model employs the hierarchical learning rate, whose capability of convergence is quickly and steady. Then, we

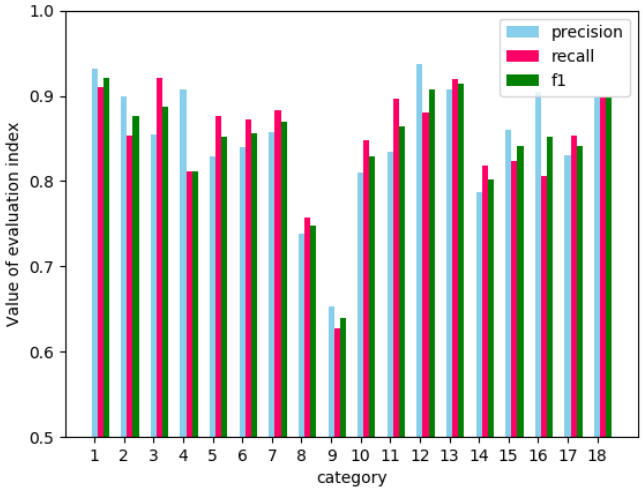


Fig. 2. Value of evaluation index of each category of SEMLT's. The figures from 1 to 18 represent Game, Essay, Baby, Regimen, Fashion, History, Military, World, Society, Travel, Food, Discovery, Car, Entertainment, Finance, Story, Tech and Sports.

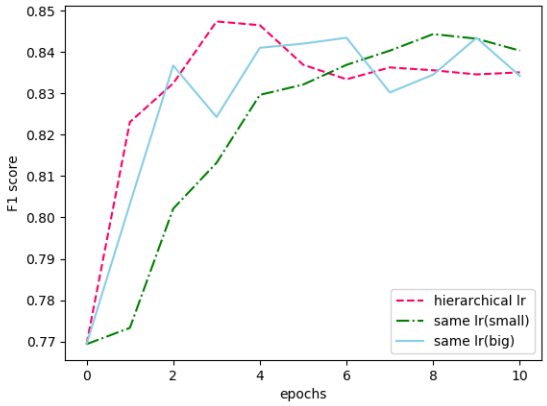


Fig. 3. The influence of epochs

design an experiment to analyze the classification performance of our model changed with epochs, the results show that the performance of model has a rapid growth trend before the third epochs and achieves peak value in the third epoch, then rate of rise is become steady fluctuate and decline. Maybe, there are two reasons can explain why this model has the capability of quick convergence. (1) In experiment, the fine-tuning strategy is executed in the pre-training model on the Chinese news data set, i.e, Sogou dataset, which can learn the deep characteristics of news text. (2) In the downstream task, the learning rate of our model is larger than the pre-training model, and the kind of hierarchical learning rate can accelerate the convergence rate of the model. Meanwhile, different dropout rates are adopted to prevent the model from overfitting and make the model

TABLE III
RESULTS OF ABLATION STUDY

Models	Game	Essay	Baby	Regimen	Fashion	History	Military	World	Society	Travel	Food	Discovery	Car	entertainment	Finance	Story	Tech	sports	F1	acc
BERT	0.8698	0.8557	0.8366	0.818	0.8377	0.8429	0.849	0.7349	0.5946	0.7937	0.8539	0.8963	0.8911	0.7588	0.8217	0.8339	0.8194	0.8792	0.8215	0.8214
BERT-Finetune	0.9007	0.8655	0.8738	0.7913	0.8367	0.839	0.8454	0.7402	0.606	0.8068	0.8181	0.8181	0.8983	0.7824	0.8103	0.8225	0.8083	0.8799	0.8241	0.8239
BERT+SE	0.8999	0.87	0.8719	0.8425	0.8316	0.8377	0.8555	0.749	0.6173	0.8164	0.8676	0.9035	0.9051	0.7659	0.8278	0.8392	0.8338	0.8756	0.8339	0.8335
BERT+SE+CG	0.9109	0.869	0.88	0.8523	0.8503	0.8504	0.8628	0.7433	0.614	0.8127	0.8546	0.9058	0.9053	0.7916	0.8311	0.8359	0.8358	0.889	0.8387	0.8385
BERT+SE+CG+WLE	0.9131	0.8722	0.879	0.8563	0.8509	0.8556	0.8566	0.746	0.6261	0.8179	0.8688	0.9099	0.9092	0.796	0.8314	0.8449	0.8253	0.8902	0.8416	0.8422
BERT+SE+CG+SLE	0.9132	0.865	0.8836	0.8564	0.8503	0.8507	0.8636	0.7413	0.6274	0.8139	0.8624	0.9105	0.9096	0.7869	0.8361	0.8321	0.8384	0.8886	0.8406	0.8404
SEMLE(ours model)	0.921	0.8763	0.8869	0.8569	0.8517	0.856	0.8701	0.7472	0.6396	0.8285	0.8638	0.9079	0.9135	0.8019	0.8411	0.8521	0.8418	0.8977	0.8474	0.8475

reach the optimal performance in a short time.

E. Ablation Study

In order to verify the effectiveness of our proposed model, including pre-training and the downstream task, we conducted ablation experiments for the four main processes. The detailed experimental results are shown in TABLE III, where SE denotes the semantic enhancement module, C denotes the CNN semantic extraction, G denotes the bidirectional GRU, CG means multi-dimensional feature fusion network, WLE denotes is the word-level label embedding and SLE is the sentence-level label embedding. The detailed processes of ablation experiments are explained as follow:

(1) Firstly, we choice Sogou dataset in Chinese news to fine-tune based on the pre-training model and make it to satisfy the news field. The pre-training model already contains a large amount of corpus information, but lacks the professional knowledge of domain, which may lead to the phenomenon that one word has multiple meanings in different context. Especially, the correct semantic meaning of each word is very important in the short text classification. The experiment results show that the strategy of fine-tuning based on the pre-training model can effectively improve the classification performance of short text.

(2) In the process of semantic enhancement, we compare three approaches: the first one is to use a fine-tuned BERT's last layer as text embedding, which can extract the high-level semantic information and ignore the low-level syntactic information. The second one is to fuse the output of all layers of fine-tune BERT to represent text embedding, which can integrate both of low-level syntactic information and high-level semantic information to enhance the representation of words, but it also will aggravate the computational complexity.

In both of above-mentioned methods, every character only focuses on the global information of the sentence, but ignores the semantics feature at the word level. BERT+SE integrates the word information into the character vector to obtain the semantic information of the text from multiple dimensions, and improves the classification effectiveness.

(3) For multi-dimensional feature fusion network, we add different modules to verify the performance of the network. Where, the pre-training model utilizes the self-attention mechanism to add the contextual semantic information into the text representation. Due to the addition of semantic enhancement module, the original global information is weakened. The bidirectional GRU module is added to enhance the sequential information representation of text, which can obviously improve the performance of classification. In addition, local information of the text is extracted by multi-scale CNN module, compared with the original pre-training model, the classification results are improved by introducing the local feature in the text. Finally, both of local and sequence features are fused to make the sentence vectors with more richness semantics. Experimental results show that the multi-dimensional feature fusion network can greatly improve the classification performance.

(4) Due to the traditional One-Hot representation ignoring the label semantics, we propose a multi-level label embedding method. Firstly, the text embedding representation is filtered by the word level label embedding. Because the classification task usually only focuses on some words in a sentence, while other words in the sentence maybe become the noise data to the classification task. The label embedding at sentence level can assist classification by calculating the similarity score between the sentence vector and label, which avoid the noise derived from compressing the sentence representation to low-dimensional space by the feedforward neural network. The experimental results show that single embedding can improve the performance of classification, and two-layer embedding will achieve the best results.

V. CONCLUSION

In order to improve the performance of news headline classification, we analyze the current semantic representation, semantic enhancement, label-based embedding models and summarize their shortcomings. Based on these work, we propose a Chinese news headline classification method based on semantic enhancement and multi-level label embedding. Specifically, semantic enhancement components put more semantic information with the words embedding into characters. Moreover, a joint model of the bidirectional GRU and multi-scale CNN is designed as the extractor of sentence features to expand sentence semantics from multi-dimensions. In addition, we creatively provide a multi-level label embedding strategy, which can filter text vectors at the word level and promote the short text classification at the sentence level. The performance of SEMLE model surpasses the best methods in previous and achieves state-of-the-art on the task of Chinese news

headline categorization in NLPCC 2017. The ablation analysis shows that different components illustrate different capability to provide the fundamental work for SEMLE's architecture.

ACKNOWLEDGMENT

This work has received funding from "the World-Class Universities(Disciplines) and the Characteristic Development Guidance Funds for the Central Universities" (PY3A022), Shenzhen Science and Technology Project (JCYJ20180306170836595), the National Natural Science Fund of China (No.F020807), Ministry of Education Fund Project Cloud Number Integration Science and Education Innovation" (No.2017B00030), Basic Science Research Operating Expenses of Central Universities (No.ZDYF2017006). We would like to thanks them for providing support.

REFERENCES

[1] X. Qiu, J. Gong, and X. Huang, "Overview of the nlpcc 2017 shared task: Chinese news headline categorization," in *National CCF Conference on Natural Language Processing and Chinese Computing*, pp. 948–953, Springer, 2017.

[2] J. Wang, Z. Wang, D. Zhang, and J. Yan, "Combining knowledge with deep convolutional neural networks for short text classification.," in *IJCAI*, pp. 2915–2921, 2017.

[3] J. Li, Y. Cai, Z. Cai, H. Leung, and K. Yang, "Wikipedia based short text classification method," in *International Conference on Database Systems for Advanced Applications*, pp. 275–286, Springer, 2017.

[4] P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, pp. 806–814, 2016.

[5] Q. Chen, L. Yao, and J. Yang, "Short text classification based on lda topic model," in *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 749–753, IEEE, 2016.

[6] A. Mahabal, J. Baldridge, B. Karagol Ayan, V. Perot, and D. Roth, "Text classification with few examples using controlled generalization," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (Minneapolis, Minnesota), pp. 3158–3167, Association for Computational Linguistics, June 2019.

[7] B. Wang, "Disconnected recurrent neural networks for text categorization," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2311–2320, 2018.

[8] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," *arXiv preprint arXiv:1606.01781*, 2016.

[9] R. Johnson and T. Zhang, "Deep pyramid convolutional neural networks for text categorization," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 562–570, 2017.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[11] G. Wang, C. Li, W. Wang, Y. Zhang, D. Shen, X. Zhang, R. Henao, and L. Carin, "Joint embedding of words and labels for text classification," *arXiv preprint arXiv:1805.04174*, 2018.

[12] Y.-B. Kim, D. Kim, A. Kumar, and R. Sarikaya, "Efficient large-scale neural domain classification with personalized attention," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Melbourne, Australia), pp. 2214–2224, Association for Computational Linguistics, July 2018.

[13] B. Demirel, R. G. Cinbis, and N. Ikizler-Cinbis, "Learning visually consistent label embeddings for zero-shot learning," *arXiv preprint arXiv:1905.06764*, 2019.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[16] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[17] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[18] C. Xu, C. Paris, S. Nepal, and R. Sparks, "Cross-target stance classification with self-attention networks," *arXiv preprint arXiv:1805.06593*, 2018.

[19] H. Zhang, L. Xiao, W. Chen, Y. Wang, and Y. Jin, "Multi-task label embedding for text classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, (Brussels, Belgium), pp. 4545–4553, Association for Computational Linguistics, Oct.-Nov. 2018.

[20] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *URL https://s3-us-west-2. amazonaws. com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf*, 2018.

[21] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Bert post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232*, 2019.

[22] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, "Analogical reasoning on chinese morphological and semantic relations," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 138–143, Association for Computational Linguistics, 2018.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[24] I. Sheikh, I. Illina, D. Fohr, and G. Linarès, "Learning word importance with the neural bag-of-words model," in *Proceedings of the 1st Workshop on Representation Learning for NLP*, (Berlin, Germany), pp. 222–229, Association for Computational Linguistics, Aug. 2016.

[25] Z. Yin, J. Tang, C. Ru, W. Luo, Z. Luo, and X. Ma, "A semantic representation enhancement method for chinese news headline classification," in *National CCF Conference on Natural Language Processing and Chinese Computing*, pp. 318–328, Springer, 2017.

[26] X. DONG, R. SONG, F. ZHU, and Q. ZHU, "Multi-model based news headline classification," *Journal of Chinese Information Processing*, vol. 32, no. 10, pp. 73–81, 2018.

[27] J. Chen, Y. Hu, J. Liu, Y. Xiao, and H. Jiang, "Deep short text classification with knowledge powered attention," *arXiv preprint arXiv:1902.08050*, 2019.

[28] S. Chen, M. Wang, J. Zhang, and L. He, "Classify sentence from multiple perspectives with category expert attention network," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

R-Cane: A Mobility Aid for Visually Impaired

Kanak Manjari*, Madhushi Verma† and Gaurav Singal‡
Department of Computer Science Engineering , Bennett University
Greater Noida, India
Email: *KM5723@bennett.edu.in, †madhushi.verma@bennett.edu.in, ‡gauravsingal789@gmail.com

***Abstract*—An Electronic Travel Aid (ETA) has become a necessity for visually impaired to provide them proper guidance and assistance in their daily routine. As the number of blind persons are gradually increasing, there is a dire need of an effective and low-cost solution for assisting them in their daily tasks. This paper presents a cane called R-Cane which is an ETA for the visually impaired and is capable of detecting obstacles in front direction using sonar sensor and alerts the user by informing whether the obstacle is within the range of one meter. In R-Cane, tensorflow object-detection API has been used for object recognition. It makes the user aware about the nature of objects by providing them voice-based output through bluetooth earphones. Raspberry Pi has been used for processing and Pi camera has been used to capture frames for object recognition. Further, we have implemented Single Shot Multibox Detector (SSD) based four models for object detection. The experimental analysis shows that out of the four models, average F1 score of all the classes is highest for SSD_Mobilenet_v1 _Ppn_Coco model model.**

***Index Terms*—Electronic Travel Aids, Sensor, Assistive Technology, Visually Impaired, Ultrasonic Sensor, Raspberry Pi**

I. Introduction

World is a very beautiful place and we are fortunate enough to be able to see amazing things around us. But a person suffering from vision loss has to face many difficulties in their day-to-day tasks. According to World Health Organization (WHO) [26], 1.3 billion of persons are estimated to be suffering from vision imapirment out of which 188.5 million persons have mild vision impairment, 217 million persons have moderate to severe vision impairment, and 36 million are blind [27]. The indoor and outdoor environment contains various obstacles of different shapes and sizes at various locations. Even sighted persons need to be careful sometimes to prevent themselves from collision with obstacles, and some of these obstacles become dangerous for visually impaired such as descending staircases, conical edges etc. that can cause injuries.

A white cane [25] is the oldest and universal solution which provide safe mobility for blind persons. This cane detects the objects in the environment around the person at ground level in the front direction. This cane cannot detect objects in different directions and also can not perform object recognition. The knowledge of presence and type of static and dynamic obstacles can offer safety and security to visually impaired persons. Some solutions in the form of cane already exists in market such as: Ultrasonic Cane [1,4], Electronic Cane [2], Smart Cane [3], and HALO [15] etc.

Assistive Technologies (AT) are used to assist visually blind which can be wearable or handheld. Wearable devices allow handsfree interaction with minimal use of hands while handheld devices require continuous hand interaction. Wearable devices are worn on different parts of body such as: vests and garments [10,11], waist belt [12], devices worn on feet [14], and glasses [13]. Although variety of assistive devices are currently available but white cane is the oldest and the conventional device used by visually impaired. For safe and quick navigation different developments have been done using different technologies such as Global Positioning System (GPS) [7], Radio Frequency Identification (RFID) [6], Ultrasonic [9], Laser [5] and Global System for Mobile Communication (GSM) [8].

Laser transmits invisible laser beams and produce different audio signals after detecting the obstacle. Ultrasonic is similar to laser technology and follows the same principle that is followed in laser. Based on the distance, different types of tones can be produced through it. GPS is most commonly used for navigation by blind as well as sighted people. It provides voice based output which is followed by users. RFID is also used for navigation, but it requires RFID tags for navigation [6].

The aim of our paper is to build a cost efficient handheld cane to assist blind and visually impaired and make them capable without the help of sighted persons. The system is raspberry pi based so that it can gain the benefit of high processing power to make it capable for real time processing. It has a pi-camera attached to raspberry which captures the images on which object recognition is done. Ultrasonic sensor is also attached to detect the object and know its range.

Our contribution includes building up of the cane for helping visually impaired using deep learning approach. This cane has the integration of both hardware and software components. Raspberry Pi is attached with the cane where object detection model is integrated that performs object detection as well as recognition. Tensorflow framework of deep learning have been used here for serving the purpose of detection. The combination of raspberry pi and Tensorflow object detection API integrated into cane is a different approach than the methods used in the earlier developed canes.

The remaining paper has been organized in following se-

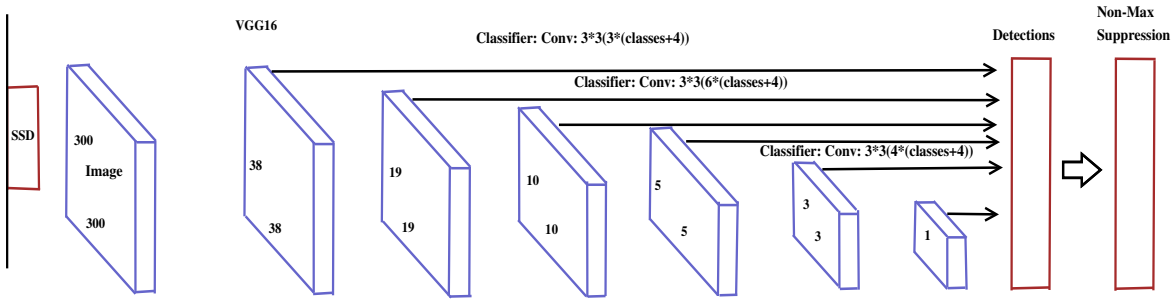


Fig. 1. Multi-Scaling in SSD Model

quence. Related work and methodology has been discussed in Section II and Section III respectively. In Section IV, experimental setup has been described. Finally, Conclusion has been presented in Section V followed by the resource information.

II. Related Work

A variety of travelling aids have been developed to increase the mobility of visually impaired. Some of these aids related to our work have been discussed in this section.

The author in [1] proposed a small and portable Ultrasonic Cane for visually blind with the aim to replace traditional long cane. Ultrasonic cane has few advantages over long cane like longer detection range and the capability to detect the overhanging objects. The authors in [2] have developed a low cost electronic cane for blind people for the purpose of obstacle detection and recognition using two ultrasonic sensors and one monocular camera. It works well in indoor environment but nothing has been stated about its robustness in urban areas.

Another device introduced in [3] is linked with a GSM-GPS module to locate the blind person and to establish a bi-directional communication path in a wireless fashion. It was developed with the aim to guide users who are visually impaired or partially sighted. Another device integrated with cane was developed by authors in [15] which provide haptic alerts during navigation to alert visually impaired of low-hanging obstacles. It is affordable but somewhat heavier than traditional cane. A stereo-image processing based system [23] was developed where stereo camera and stereo earphones was integrated on helmet. Image of the scene in front of user is captured through stereo camera and those images are processed to extract features for assisting blind persons in navigation. The information about the object present is provided to user in musical stereo sounds.

Drishti [24] is a navigation system for blind persons which uses a wearable computer and a vocal communication interface to guide users in travelling in indoor and outdoor environment. The working range of this system is limited with the range of wireless network and installation cost is also high. A RGB-camera based solution, Navigation Assistance for Visually Impaired (NAVI) [22] has been developed to assist visually blind persons through sound commands. To remove the use of multiple sensors, RGB-camera is used here to utilize visual and range information which eased the process of complex image processing tasks.

Bbeep [20] is another recent aid developed for visually impaired to help them in navigation in crowded environment. It is a suitcase based system that alerts user as well as nearby persons from collision using sonic feedback through pose estimation. Recently, a path guiding robot [21] was also developed to help visually impaired with the aim to replace guide dogs. It has the capability to move along multiple path as well as retrace them. And thus, make it easy for blind persons to navigate in indoor as well as outdoor environment.

Single Shot Multibox Detector (SSD) [18] is a model for object detection in real time. Faster Region-Based Convolution Neural Network (Faster R-CNN) [16] is also an object detection model which make use of Region Proposal Network (RPN) to create bounding boxes and classify those objects. It has a good accuracy but not suitable for real-time applications. You Only Look Once (YOLO) is another object detection model which is suitable for real-time applications but is not so accurate. It cannot detect small size objects such as a tennis ball. SSD model uses multi-scale features and eliminates the use of RPN which was used in Faster RCNN. This improvement of SSD over Faster RCNN helps it in achieving accuracy using lower resolution images and raises the speed. As shown in Fig. 1, SSD uses base network of VGG-16 which creates feature maps with decreasing sizes. These varying size feature maps are used for scale variance of objects. Detector and classifier are applied to each feature map. Multi-scale features makes it better than YOLO model in terms of accuracy as model gets trained to detect obejcts at different scale. SSDLite is an extension of SSD model where kernel size is modified and depthwise separable convolution is performed to make it lighter and faster.

MobileNetV2 [19] is an extension of MobileNetV1 where a module is introduced with inverted residual structure. Non-linearities in narrow layers present in MobileNetV1 has been removed. In the first layer, depthwise convolution is performed where a lightweight filtering is done. In the next layer, a pointwise convolution is

performed for detecting pattern and performing feature extraction. An activation function, ReLU is used because of its robustness while low-precision computation.

Despite of the presence of many aids to help visually impaired, there is still the need of a cost-effective and handy solution for them to make their day to day life easier. The solutions that have been proposed till now are either heavy or costly and therefore it does not suit the needs of the visually impaired.

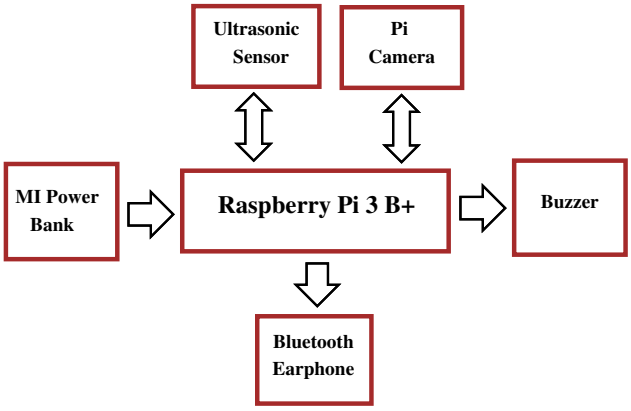


Fig. 2. Block Diagram of the Proposed System

III. Methodology

In this section, components used for development of this device has been discussed followed by the approach and algorithm that has been implemented in R-Cane.

A. Components of R-Cane

The system architecture mainly consists of six parts: raspberry pi 3 B+, ultrasonic sensor, pi camera, vibration motor, MI power bank and bluetooth earphone as shown in Fig. 2. Raspberry Pi is the main controller on which all other components have been attached. Ultrasonic sensor and pi camera provide the data captured by them back to raspberry pi. Vibration motor activates when any obstacle is sensed by ultrasonic sensor and voice-based output is provided through bluetooth earphones. MI power bank has been used to provide power to the raspberry pi.

Raspberry Pi needs a New Out of the Box Software (NOOBS) which contains Raspbian operating system. Tensorflow library has been used here to develop, train and test machine learning models. The TensorFlow object detection API is an open source framework that is built on top of TensorFlow which makes it easy to construct, train and deploy object detection models. Open Source Computer Vision Library (OpenCV) has also been used which is an open source computer vision and machine learning software that helps in providing a baseline and infrastructure to computer vision applications and accelerate the use of machine learning in commercial products.

B. Approach

An object detection model is used to detect objects and the location of objects. For example, a model might be trained with images that contain cats and dogs, along with a label that specifies the class of object they represent (e.g. a cat, or a dog), and data specifying where each object appears in the image. Object detection can perform classification and localization of multiple objects present in an image which is being used in many areas such as crowd management, traffic management, medical imaging, and computer vision.

Four SSD based models have been deployed in raspberry pi for object detection purpose. Although these models are SSD based, but are slightly different from each other in terms of their configuration as stated in Table I Batch size, number of steps, regularization and decay are few hyperparameters which can be fine tuned to obtain better accuracy. Batch size varies from 24 to 2048 for these models which defines the number of samples to be worked upon before updating the model's internal parameters. Larger batch size would require more memory space but for larger datasets using batches makes the process faster. Number of steps for these models varies from 10000 to 200000 which makes impact on the training time. L2 regularization is used to avoid the risk of overfitting by discouraging the learning of complex model whose value is 0.00004 for these models. Decay rate for these models vary from 0.97 to 0.99997 which controls how quickly or slowly a neural network learns a problem. SSDLite_mobilenet_v2_coco model is the lighter version of SSD with 27ms speed and 22 mAP achieved when trained on coco dataset. SSD_mobilenet_v1_ppn_coco model uses Pooling Pyramid Network (PPN) to make predictions through shared box detector which achieved 26 ms speed and 20 mAP when trained on coco model which do not makes it better than SSDlite. SSD_mobilenet_v1_0.75_depth_coco models uses 0.75 depth multiplier for better detection and achieved 26 ms speed and 18 mAP when trained on the same coco models. And, SSD_mobilenet_v1_coco is heavier than other three models but has good detection performance of 21 mAP.

In Fig. 5, the whole architectural process of convolution and pooling in SSD model is shown in a diagram. An RGB image of size 224*224 having three channel is convolved and ReLU activation function is applied to introduce non-linearity of model. Max pooling is performed to reduce the computational parameters and size is reduced to 112*112. After this process of convolution and pooling, the data is passed to a fully connected layer and then softmax activation function is applied for classification.

C. Algorithm

The algorithm for developing this system requires the presence of a software model for object detection. Initially, frames are captured by pi camera and resized to 224*224. Then, convolution, classification and detection of four

classes are performed using few SSD-based models. If object is present, the device provides voice-based output to the user about the type of object and confidence through bluetooth earphone.

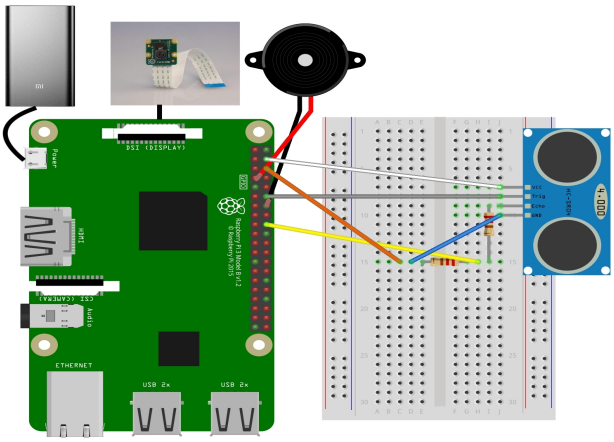


Fig. 3. Inter-Connection of Components

Algorithm 1 Object Detection & Recognition

Require: Weight & model file for SSD-based models
Ensure: Object name & their range

1: Frame is captured & resized to (224*224)
2: Convolution with ReLU, classification using softmax & detection is performed
3: Object detection performed for four classes
4: **if** (object==1) **then**
5: Ultrasonic sensor activated
6: Object name & range is provided through earphone
7: **else**
8: Go back to the beginning of program
9: **end if**

TABLE I
EVALUATION PARAMETERS AND ITS VALUE FOR ALL OBJECT-DETECTION MODELS

Object Detection Model	**Parameters**			
	Batch Size	***Num Steps***	***Regularizer (l2)***	***Decay***
SSD_mobilenet _v1_0.75_depth_coco (Model1)	2048	10000	0.00004	0.97
SSD_mobilenet _v1_coco (Model2)	24	200000	0.00004	0.99997
SSD_mobilenet _v1_ppn_coco (Model3)	512	50000	0.00004	0.97
SSDlite_mobilenet _v2_coco (Model4)	24	200000	0.00004	0.99997

IV. EXPERIMENTAL RESULTS

In this section, Experimental Setup and Performance Metrics has been discussed. The whole setup including the circuitry connection of various components has been presented in the first sub-section and the parameters used for the evaluation of performance of the system has been discussed in the second sub-section.

A. Experimental Setup

The entire system is organized with different cost-effective components to provide flexibility and comfort to visually impaired or partially sighted people as shown in Fig. 3.

Fig. 4. Model of Developed Stick (Front View, Top View, Side View)

Raspberry Pi is the main processing board used here which is a cheap, pocket-sized Personal Computer (PC) that fits into a computer screen /TV and utilizes a standard console and mouse. It is a competent little gadget that empowers individuals of any age to explore processing, and to figure out how to program in different languages like Python and Scratch. It can do all that we would anticipate that a computer should do, from perusing the web and playing top notch video, to making spreadsheets, word-handling, and making diversions.

The raspberry pi camera is a high-definition camera module which can capture image as well as video. It is supported by all the versions of raspberry pi and is mainly used in security applications and wildlife camera traps. Ultrasonic sensor can find the distance between user and obstacle to prevent the user from colliding and an alarm is provided to the user in the form of a vibration from vibration motor. The object recognition is done from the

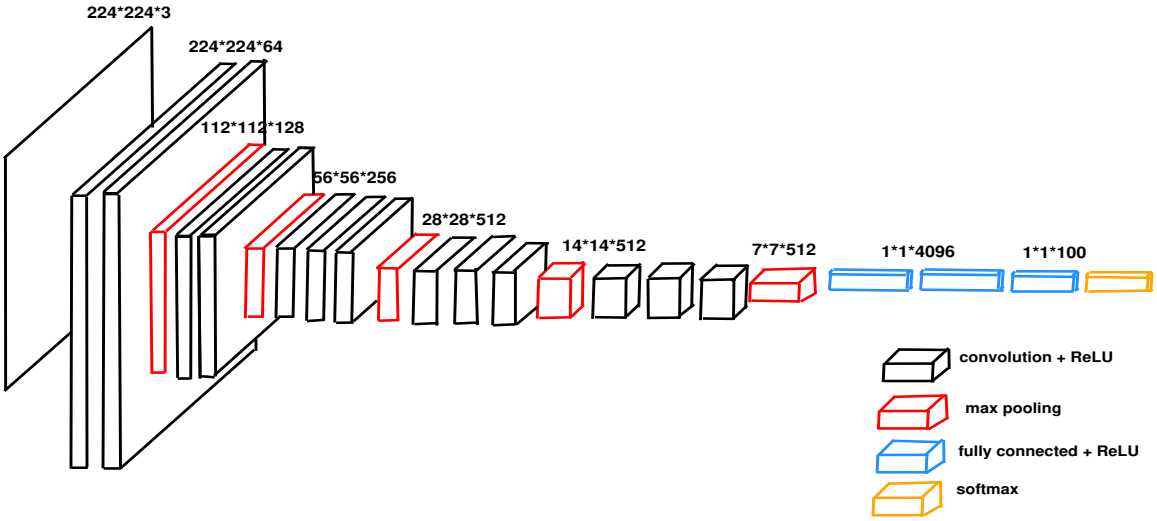


Fig. 5. Architecture of SSD Model

image captured through pi camera using deep learning model for object detection and this information about the obstacle is provided to user through bluetooth earphone. The system is powered by a power bank to make it more durable. R-Cane along with the integrated components have been presented in Fig. 4. Front view, top view and side view of the R-Cane being used by user has been shown here.

B. Performance Metrics

In this section, result obtained from few SSD based object detection model available in tensorflow is discussed on the basis of Precision, Recall, F1 score, Frames Per second (FPS) and Confidence of each detected object in a similar scenario. Precision is the measurement of accuracy of prediction. Recall measures how well we find all the positive results. FPS is the frequency of consecutive images called frames that appears on screen. Confidence is the probability of presence of an object in the anchor box.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Using equation (1) & (2),

$$F1Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{3}$$

Here, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

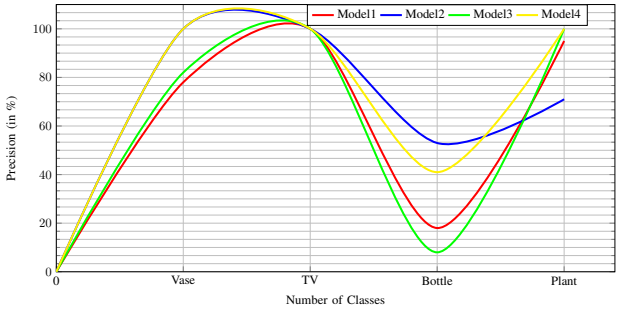


Fig. 6. Precision of Models for Four Classes

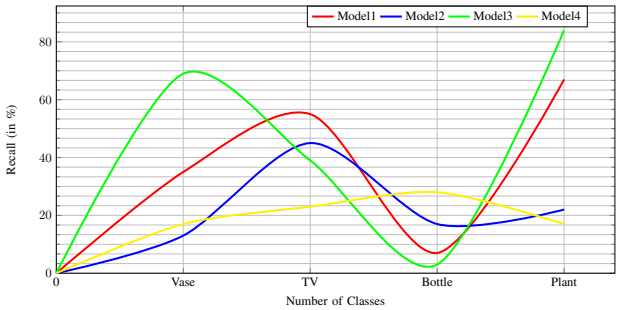


Fig. 7. Recall of Models for Four Classes

The scenario is indoor where TV, vase, bottles etc. are present. All the models including Model1, Model2, Model3 and Model3 are deployed in the system and for each model FPS and confidence of these objects are noted which is presented in the Table II. FPS of SSDlite_mobilenet_v2_coco model is higher than other three models while capturing the same indoor scenario using pi camera as SSDlite is lighter than SSD based models. The confidence of two objects "vase" and "TV" is higher when SSD_mobilenet_v1_coco model is used while for

other two objects "Bottle" and "Plant" confidence is higher when SSD_mobilenet_v1_ppn_coco model was used.

TABLE II
COMPARATIVE ANALYSIS OF FEW OBJECT-DETECTION MODEL DEPLOYED IN THIS DEVICE

Object Detection Model	**Frames Per Second**	**Confidence of Detected Objects**			
		Vase	***TV***	***Bottle***	***Plant***
SSD_mobilenet _v1_0.75_depth_coco (Model1)	0.23	49%	64%	40%	47%
SSD_mobilenet _v1_coco (Model2)	0.37	58%	74%	45%	42%
SSD_mobilenet _v1_ppn_coco (Model3)	0.16	54%	72%	50%	51%
SSDlite_mobilenet _v2_coco (Model4)	0.50	50%	42%	40%	61%

TABLE III
F1 SCORE OF OBJECT DETECTION MODELS FOR FOUR CLASSES

Object Detection Model	**F1 Score**				**Average**
	Vase	***TV***	***Bottle***	***Plant***	
SSD_mobilenet _v1_0.75_depth_coco (Model1)	0.48	0.70	0.09	0.78	0.51
SSD_mobilenet _v1_coco (Model2)	0.22	0.61	0.25	0.33	0.35
SSD_mobilenet _v1_ppn_coco (Model3)	0.74	0.56	0.04	0.90	0.56
SSDlite_mobilenet _v2_coco (Model4)	0.29	0.37	0.33	0.29	0.32

Using equation (3), F1 score of all the four models for all the four classes are shown in Table III. It is observed that Model3 achieves the highest F1 score for two classes Vase and Plant. For class TV and Bottle, Model1 and Model4 achieves the highest F1 score than other models. The average value of F1 score of all the classes for each model is also shown in last column of Table III where Model3 achieves the highest score. Using equation (1) & (2), precision and recall has been calculated. In Fig. 6 and Fig. 7, the results obtained for Precision and Recall of the four models has been presented where SSD_mobilenet_v1_0.75_depth_coco has been represented as Model1, SSD_mobilenet_v1_coco as Model2, SSD_mobilenet_v1_ppn_coco as Model3 and SSDlite_mobilenet _v2_coco as Model4. Using Model3, best precision is achieved for classes TV and Plant and best recall is achieved for classes Vase and Plant. Using Model2, best precision is achieved for classes Vase and TV and best recall achieved for classes TV and Plant. Using Model1, best precision and recall is achieved for classes TV and Plant. Using Model4, best precision is achieved for classes Vase, TV and Plant & best recall achieved for classes TV and Bottle.

V. CONCLUSION

In this paper, we have developed a raspberry pi based cane called R-Cane for guiding visually impaired. In order to provide ease and independence to users, a microcontroller is attached with the cane in which object detection model is deployed. It helps visually impaired to recognize the obstacles present in front of them and the presence of those object is detected through ultrasonic sensor. Evaluation of the system is done by attaching the system to the cane. It has been determined through experimental analysis that out of the four models, SSD_mobilenet_v1_ppn_coco model achieves the highest average value of F1 score for all the classes.

VI. RESOURCE

Working code used for integration of object detection model and ultrasonic sensor in R-Cane can be found on this link: https://github.com/kmanjari/R-cane.

REFERENCES

[1] Hoydal, T. O., and J. A. Zelano. "An alternative mobility aid for the blind: the'ultrasonic cane'." Proceedings of the 1991 IEEE Seventeenth Annual Northeast Bioengineering Conference. IEEE, 1991.

[2] Bouhamed, S. A., Eleuch, J. F., Kallel, I. K., & Masmoudi, D. S. (2012, July). New electronic cane for visually impaired people for obstacle detection and recognition. In 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012) (pp. 416-420). IEEE.

[3] Alshbatat, Nour, and Abdel Ilah. "Automated Mobility and Orientation System for Blind or Partially Sighted People." International Journal on Smart Sensing & Intelligent Systems6.2 (2013).

[4] Kumar, Krishna, et al. "Development of an ultrasonic cane as a navigation aid for the blind people." 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, 2014.

[5] Benjamin Jr, J. Malvern, and Nazir A. Ali. "An improved laser cane for the blind." Quantitative Imagery in the Biomedical Sciences II. Vol. 40. International Society for Optics and Photonics, 1974.

[6] Want, Roy. "An introduction to RFID technology." IEEE pervasive computing 1 (2006): 25-33.

[7] Misra, Pratap, and Per Enge. "Global Positioning System: signals, measurements and performance second edition." Global Positioning System: Signals, Measurements And Performance Second Editions, (2006).

[8] Mouly, Michel, Marie-Bernadette Pautet, and Thomas Foreword By-Haug. The GSM system for mobile communications. Telecom publishing, 1992.

[9] Rozenberg, L., ed. Physical principles of ultrasonic technology. Vol. 1. Springer Science & Business Media, 2013.

[10] Bahadir, Senem Kursun, Vladan Koncar, and Fatma Kalaoglu. "Wearable obstacle detection system fully integrated to textile structures for visually impaired people." Sensors and Actuators A: Physical 179 (2012): 297-311.

[11] Lee, Young Hoon, and Gérard Medioni. "RGB-D camera based navigation for the visually impaired." Proceedings of the RSS. 2011.

[12] Mahalle, Sushant. "Ultrasonic Spectacles & Waist-Belt for Visually Impaired & Blind Person." IOSR Journal of Engineering 4 (2014): 46-49.

[13] Xiang, K., Wang, K., Fei, L., & Yang, K. (2019, May). Store sign text recognition for wearable navigation assistance system. In Journal of Physics: Conference Series (Vol. 1229, No. 1, p. 012070). IOP Publishing.

[14] Patil, Kailas, Qaidjohar Jawadwala, and Felix Che Shu. "Design and construction of electronic aid for visually impaired people." IEEE Transactions on Human-Machine Systems 48.2 (2018): 172-182.
[15] Wang, Yunqing, and Katherine J. Kuchenbecker. "HALO: Haptic alerts for low-hanging obstacles in white cane navigation." 2012 IEEE Haptics Symposium (HAPTICS). IEEE, 2012.
[16] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
[17] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
[18] Jonathan Hui, 'SSD object detection: Single Shot MultiBox Detector for real-time processing' , 2018. [Online]. Available: https://medium.com/@jonathan_hui/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06.[Accessed: 14-Mar-2018].
[19] Sik-Ho Tsang, 'Review: MobileNetV2 — Light Weight Model (Image Classification)' , 2019. [Online]. Available: https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c. [Accessed: 19-May-2019]
[20] Kayukawa, S., Higuchi, K., Guerreiro, J., Morishima, S., Sato, Y., Kitani, K., & Asakawa, C. (2019, April). BBeep: A Sonic Collision Avoidance System for Blind Travellers and Nearby Pedestrians. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (p. 52). ACM.
[21] Megalingam, R. K., Vishnu, S., Sasikumar, V., & Sreekumar, S. (2019). Autonomous Path Guiding Robot for Visually Impaired People. In Cognitive Informatics and Soft Computing (pp. 257-266). Springer, Singapore.
[22] Aladren, A., López-Nicolás, G., Puig, L., & Guerrero, J. J. (2014). Navigation assistance for the visually impaired using RGB-D sensor with range expansion. IEEE Systems Journal, 10(3), 922-932.
[23] Balakrishnan, G. N. R. Y. S., Sainarayanan, G., Nagarajan, R., & Yaacob, S. (2006). A stereo image processing system for visually impaired. International Journal of Signal Processing, 2(3), 136-145.
[24] Ran, Lisa, Sumi Helal, and Steve Moore. "Drishti: an integrated indoor/outdoor blind navigation system and service." Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the. IEEE, 2004.
[25] Sjöström, Calle. "Virtual haptic search tools–The white cane in a haptic computer interface." Assistive technology: Added value to the quality of life, AAATE 1 (2001): 124-128.
[26] World Health Organization (2018, Oct 11). [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
[27] Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, et al.; Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health. 2017 Sep;5(9):e888–97.

RoadWay: Lane Detection for Autonomous Driving Vehicles via Deep Learning

Himanshu Singhal[1], Riti Kushwaha[2], Gaurav Singal[3]

CSE Department
[1] *IIIT, Vadodara*, Gujarat, 201551014@iiitvadodara.ac.in
[2] *MNIT Jaipur*, Rajasthan, riti.kushwaha07@gmail.com
[3] *Bennett University*, Greater Noida, UP, gauravsingal789@gmail.com

Abstract. Locomotion is basic to all human needs. Modern-day transport has come a long way but still far away from perfection and all-around safety. Lane Detection is a concept of demarcating lanes on the roads while the vehicle is moving. Lane detection algorithm is used in the project of autonomous self-driving vehicles. It has the capability of changing the vehicular movements on road to a great extent making them more organized and safe. This leap could provide for driver carelessness and avoid a lot of mishaps on the roads. Ride-hailing services such as Uber and Ola companies can use them to monitor drivers and rate them based on driving skills. We have designed and trained a deep Convolutional Network model from scratch for lane detection since a CNN based model is known to work best for image classification datasets. We have used multiple metrics values for hyper-parameters and took the ones which gave the better results. The training part is done on Supercomputer NVIDIA DGX V100. A deep learning approach has been proposed to successfully identify the lane in highways on video data.

1 Introduction

Lane Detection as the name suggests identifies and marks the lanes on the road so as to assist vehicular movements. Lane Detection even has the capability of guiding blind drivers to a certain extent by helping them navigate in particular lanes and applying brakes when the lane marked area before the car falls less than a predefined value based on the size of the vehicle. One of the major challenges faced is one of different road and weather conditions which we have taken up and tried to solve. There is no effect of illumination changes and road surfaces in the final predicted lane output.

Deep Learning is a supervised learning paradigm that takes in labeled datasets and develops the learning into a model. This model can then be used to predict the desired information for a completely new and unseen input i.e input which was not used in the training process. Deep Learning has taken off in the recent years due to increase computation power. It is much accurate as compared to machine learning and does not require to write complete algorithms to get the task done. It consists of different layers and their associated functions which are chosen depending upon the problem at hand.

Millions of lives are lost every year on roads due to unorganized traffic on roads. According to the data statistics by saving Life responsibility of drivers is the top contributor to road crash deaths, accounting for 80.3 percent deaths out of the total road crash fatalities in 2016. Out of the three vulnerabilities listed below, our model can efficiently solve the speeding and overtaking issues. Thus bringing down accidents on roads by 94.9 percent. By calculating the marked area on road and observing if it is safe to change lanes or not we can bring down the number of accidents down by a great extent. The following statistics have been taken from Save Life official survey data.

According to this data if we perform similar calculations, our model is bound to avoid more than 90 percent of such accidents which is truly revolutionary in its own sense. Coupled with other DAS (driver assistance systems) like sign board detection and pedestrian detection our model has the capability to transform the way cars move on road. We have developed, trained and tested our model from scratch for lane detection process. We have constructed multiple layers for convolution, de-convolution and pooling. Brief components of each layer in the model are listed in Implementation section.

The paper structure is followed by existing work and predefined rules with their pros and cons in section II. In section III, we have discussed the proposed approach, programming requirements and mathematical model for the approach. Later, we will be presenting the results and analysis of the approach and architecture of model in section IV. In the last, we have concluded our work with future possibilities.

2 Related Work

Lane detection is the major requirement in Autonomous self-driving car. Multiple algorithms have been proposed by researcher to provide a reliable solution. Time constraints is the critical part in this problem because a delay of seconds can cause a big accident. Authors are trying to solve this issue by using different image processing and Deep learning techniques.

2.1 Lane Detection with Image Processing

Through several years lane detection has been addressed through image recognition which is not a viable solution when the vehicle is on road. Image processing makes real-time systems slow thus completely moving it out of the picture. Deep Learning-based model is superior to it both in terms of speed and accuracy.

2.2 Lane Detection with Deep Learning

Lane Detection with deep learning has taken place in recent years due to increased computational power and large varieties of road data available. But most models lack on the part of accuracy and robustness. We have taken utmost care

Fig. 1. Image Processing Example

Cause Of Crash	Road Crash Fatalities	Percent Share
Speeding	73,896	61
Over-taking	9,462	7.8
Intake Of Alcohol	6,131	5.1

Table 1. Road Accident Statistics

to leverage our model for every condition is it road surfaces or weather changes that can occur on most roads.

Deep Learning is a much more advanced version of machine learning and in which we do not need to write algorithms to get our task done. This is why deep learning performs much better on image datasets where is not feasible to write accurate algorithms and achieve a high accuracy. We have thus used deep learning approach for our model. We have trained and tested our model from scratch thus enabling us to tweak changes according to our specific needs and problem domain. We have used sequential function in keras to lay layers one after the other. The summary function gives the complete training summary of the model along with epochs, trainable parameters, hyper-parameters along with the features of every layer.

Lane Detection is a hot research topic with quite a number of papers and journals in recent years. Though the only issue is accuracy with respect to different road and weather conditions. We have taken up the issue and tried to solve it through a different perspective and have achieved an accuracy of 96.34 percent. We have constructed our deep learning model from scratch using the sequential layers function in keras. We have laid out several convolution, de-convolution and pooling layers to extract the features in the dataset so as to predict the

features for the new image provided. The extracted feature pertaining to our problem is the demarcated lanes on the road.

Deep Learning is a feature extraction technique which extracts features from the provided dataset and the tries to predict results for unseen data. It has gained lot of attention from researchers around the world. Deep learning has been implemented by various researchers before and has various pros and cons to their implementation.

Most work that has been done is in form of opencv and image processing which is not promising on aspects such as accuracy and speed. It is even tedious for a researcher to go about it. Deep Learning provides a much better semantic segmentation for the images thus extracting features much more efficiently with high accuracy. Deep learning models tend to perform much better on image datasets and corresponding statistics. There has been work in deep learning adas but no perfect model has yet been made to overcome the mishaps on roads. Our model inculcates most incumbencies and brings out a really excellent model for deployment.

On-Board lane detection system [1] emphasizes to build a monocular vision system to locate lanes on the road in real time. A monocular camera is mounted on the vehicle to get road image. This road images are then fed to cannys edge detection algorithm to obtain an edge map from the corresponding image. After this, the candidates of the road lines are normalized using a matching process. After this it proposes to reinforce possible road lines and degrades the impossible lines. Linking condition is used to increase the confidence of lane lines. K means clustering provides for the localization of obtained road lines.
Pros:

1. Works well for most conditions on the roadway.
2. Computation cost is really less and model can perform without lags in real time.

Cons:

1. Image processing is not great solution when it comes to safety of lives on road
2. We cannot take advantage of positive or negative reinforcement from the data generated by already occurred situations
3. Issues like wrong parked vehicles, shadows of trees, bad quality lines, unusual pavements, dissimilar slopes and sharp curves can really be of inhibition to this model.

Rapidly adapting machine vision [2] uses machine vision techniques coupled with a RALPH (Rapidly adapting lateral position handler) vision system developed by Carnegie Melon University and Assist-Ware Technologies Inc. Ralph segregated the obtained image in three major steps i.e. sampling image, determining road curvature and finally accessing the lateral offset of vehicle with respect to the lane center. Pros:

1. Provides good results in standard conditions but noisy images carrying environmental issues like fog, rain etc this model suffers a lot.

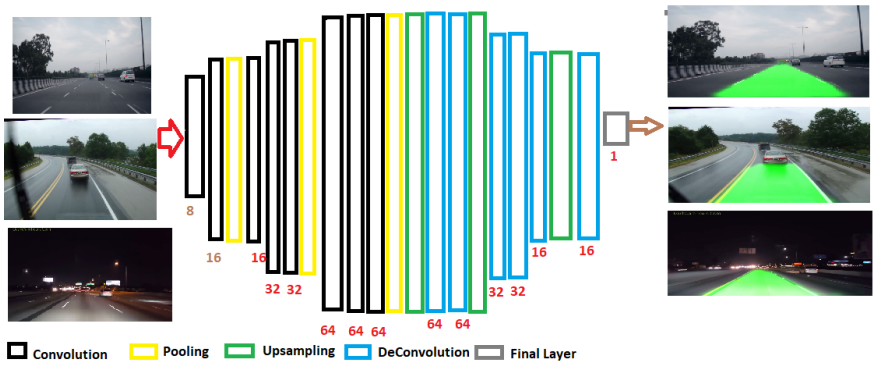


Fig. 2. Representation of CNN Model layers and Input/Outputs

2. For sharp curves the mod el is highly incompetent.
3. Locating specific features via hand programmed algorithms which is good to a great extent since there is less chance of loss of information.

Cons:

1. Hand programmed features make it very difficult to scale and span all road conditions in all types of atmospheric conditions.
2. Hand Programming algorithms make it completely difficult at back end with high computation power model running in background. It is not advisable to waste resources during deployment. Rather it is much better to spend a lot more resources during the training time.
3. We cannot take advantage of positive or negative reinforcement from the data generated by already occurred situations.
4. Hough transform was used which can be transformed for better accuracy.

Efficient lane detection based on artificial neural network [3] is quite revolutionary in its context. Rather than focussing on general image processing and analysis techniques, it took a major leap and proposed a based on Ellipsoidal Neural Network with Dendritic Processing (ENNDPs) to find a solution for the lane detection problem. The performance of the model thus created was validated by mounting a camera on the car which then navigated through the urban highways of Mexico City.
Pros:

1. Techniques are really accurate and display state of art in image processing.

Cons:

1. Large training and testing dataset for accuracy good enough for driving car on road.
2. Specialized hardware used for testing and training data. Hardware capabilities consist of high clock speed and ability to parallelize matrix calculations.
3. Such a hardware possess its own time and cost, and it is not viable for daily computing which makes it really non-scalable.

2.3 Predefined Theory

Deep Learning is a concept that has taken off in the recent years due to increased computational power and widespread data generation in amounts larger than the consumption level. Deep Learning is a supervised learning paradigm. Labeled datasets are provided as inputs and the model is trained based on a metric optimization. These metrics are often known as optimizer and loss function. Our model involves Adam optimizer and mean square error loss function. Initially, we had used SGD (stochastic gradient descent) optimizer but it had some downsides so we took up Adam optimizer for more robust training. Adam optimizer is an extension to SGD, developed in a more wider domain for effective results.

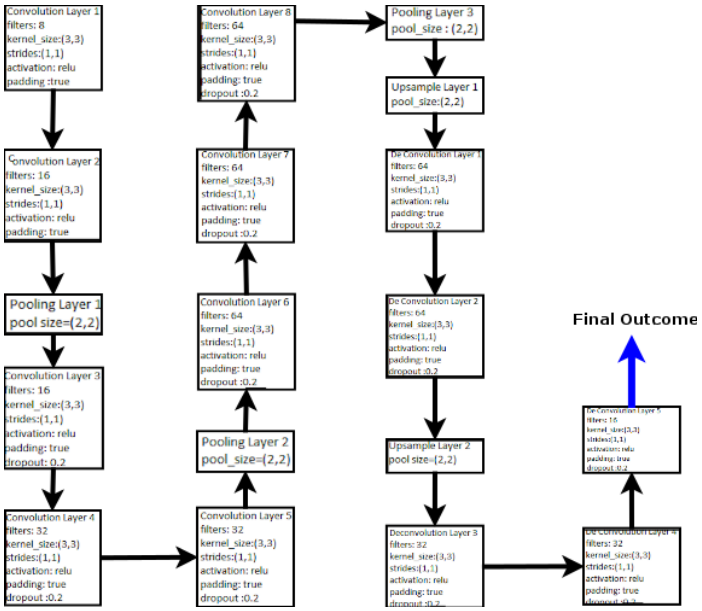


Fig. 3. Implemented CNN Model Visualization

3 Proposed approach

In this paper, we are providing a reliable and efficient solution for Lane detection in Indian road networks for self-driving autonomous vehicles.

3.1 Methodology

We have designed and trained a deep convolutional network (DCN) from scratch for lane detection since a CNN based model is known to work best for image

data. We have used different metrics values for hyper-parameters and took the ones which gave the best result. The training is done on NVIDIA-DGX V100 supercomputer because trainign a model requires lot of computation. A deep learning approach has been shown in Figure 2.

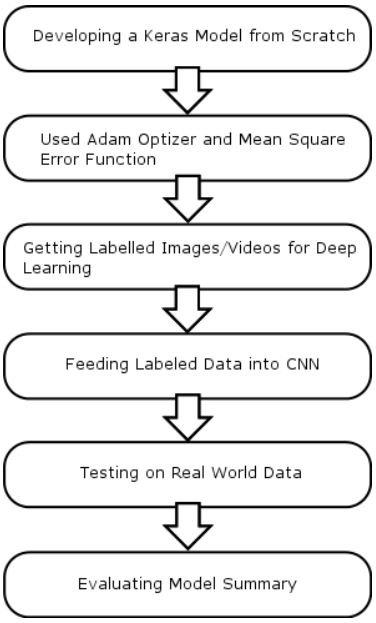


Fig. 4. Deep Learning Flowchart for the Lane Detection

We have taken several layers in our code from the convolution, deconvolution, pooling and up-sampling function provided in the sequential module of keras. We have laid all these layers one after the other in our model. We start with the convolutional layer where we have 8 filters. Filters determine the dimensionality of the output space. Thus our first layer has 8 dimensional output. We have a kernel size of (3, 3) which gives the length of 1D convolution window provided by the layer. Stride provides the stride length of layer which is (1, 1) for the first layer. Padding is done on the input layer so that the output has the same length as the input. Rectified Linear Unit (RLU) is used as an activation function for this layer. The similar approach is used for other convolutional layers which can be checked in the table listed below.

Lets dive down a step by step process of what we have done and achieved shown in Figure 3. We have taken labeled datasets and extracted the lane channel from them in a different color. This forms the labels during training process. So we have two files with us now. One is simple the road images obtained from the dataset and other is the labels file obtained from the dataset containing a different colored channel demarcating the road.

Now we design keras sequential model based on deep learning and set up different layers of convolution, pooling, up-sampling and de-convolution. These layers provide different input values which are fed into the layers based on the prediction model objective. We have shown the flow chart of our process in Figure 4.

We have also defined various hyper-parameters 2 that depict the accuracy of the model during training and testing phase. These are also listed in the below table. At the training time all the packages listed above should be updated.

One of the greatest upsides of our model is that it can work on video inputs and outputs rather than discrete images. This provides it robustness and ability to be deployed at real time.

Programming Requirements: We have used **tensorflow** and **keras** deep learning library in python. We have used **Anaconda** environment.

3.2 Mathematical Model

We have used an FCN based model since it is known to work best for image data. An FCN lacks any fully connected layers. The dataset contains the label to label pixels of road images and was used during training and testing. We have followed a lane segmentation approach using color models to classify the road from the rest of the background. Dataset was fed as input to the model to predict the road lanes. Dataset images were scaled down a bit to decrease the training time of the model. Data augmentations like image rotations and horizontal flips were performed to increase the amount of data.

We have taken real world dataset for training and have split testing and training data in 80/20. This combination often works well for most deep learning models.

Percentage of Diagram avoided

$$= 94.9/100 * 96.34/100 = 91.43$$

Table 2. Hyperparameter Values

Hyper Parameters	**Values**
Batch Size	130
Pool Size	(2,2)
Epochs	50
Loss Function	Mean Squared Error
Optimizer	Adam
Model	Keras Sequential
Batch Normalisation	(80,60,3)
Image Data Generator	ChannelShiftRange=0.3

4 Results and Analysis

We have taken a deep learning route in which we have laid out several keras sequential layers and tried to build a model using the same. Training and testing is done on the model and features are extracted from the road images. We have laid out several convolution, de-convolution and pooling layers and developed a FCN based architecture which is known to work best on image data. We have achieved an overall accuracy of 96.34 percent for different scenarios.

Fig. 5. Result for a Straight Road, downhill and uphill in Daylight

Our model is a part of ADAS (automatic driver assistance systems) that can very easily help drivers to be safe and keep other safe too. Our model can be built with other models like pedestrian detection, signboard analysis and traffic recognition to built a robust navigation systems. Ride hailing services can deploy our model in its core state or coupled with others to ensure safety for drivers as well as passengers.

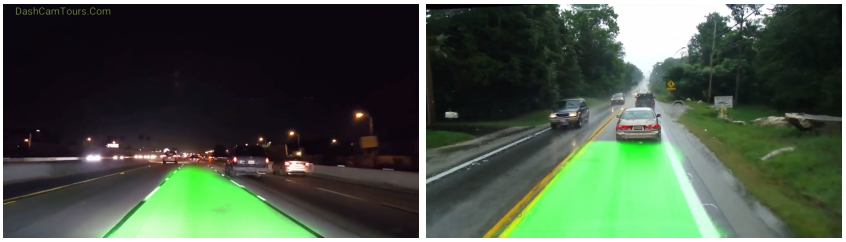


Fig. 6. Lane Detection Output for a Road in Night and rainy Scenario

Images in Figures 5 are taken from the output video obtained on feeding an input video of Gurgaon-Delhi highway in India. It is a normal daylight video with atmosphere neither to bright or too dark. Clockwise left to right, first image represents lane detection on straight road, second one with a downward slope and the third with an upward slope. As seen in the images, the result is pretty accurate with continuous lane markings on the current lane only till a vehicle on the same lane is encountered.

Figure 6 are taken from the video output of the model when fed with a night and rain input videos (right to left). The input videos are of the express

Table 3. Losses on different epochs

Epochs	Loss
10	0.076
30	0.045
50	0.021

highways in the US and represent how our model functions in different weather and atmospheric conditions. Length of lane markings depends on how far the next vehicle is in the current lane.

In table 3, we have shown the accuracy on different epochs. Going by calculations, our model alone is capable of saving a million lives every year. Moreover, we must not forget that it is a deep learning model and the more it is trained and followed the more; it has the chances of better training and getting reliable system from it. We can also have a negative reinforcement learning algorithm and thus making our model more robust for other users and scenarios.

This model training is required lots of computation for building an efficient system in terms of time and cost. For training purpose, we have used NVIDIA machine that save the resources by 100 times. We have also worked for optimization of this trained model for installing it on low power and resource constraint devices by applying different compression techniques.

5 Conclusion And Future Work

One of the greatest advances we would like to bring on-board is the lane detection in real time through mobile phone apps and it would really help real-time car drivers. When the demarcated area before the vehicle falls below a particular value depending on the size of the car either a lane change option or automatic breaking is triggered. This would help avoid collisions to a great extent. To avoid excessive lane changing safety messages are displayed for constant lane changes if there is sufficient area in front of the vehicle thus preventing rash driving. During lane changing, automatic lane changing headlights are turned on for other vehicles to notice and be safe. We would also like to improve upon our model in terms of RNN as it is known to work best in case of sequence inputs. Lanes on roads are presented in a form of sequence data and RNN will suites better for this approach. We have built a CNN model that identifies the lane in day, night, and rainy conditions. In parallel, we are also working on pedestrian detection on road while driving and add both the model can use for autonomous self-driving car.

Acknowledgment

leadingindia.ai and it's director Dr. Deepak Garg for continuous support throughout our project. Bennett University for granting full access of resources specially Supercomputer Nvidia DGX-1 V100 GPU for training and testing our model. Mr. Aditya Sharma, Program Manager, Microsoft, US for guiding us throughout the project.

References

1. Xiaodong Miao, Shunming Li, Huan Shen. "On-Board Lane Detection System For Intelligent Vehicle Based On Monocular Vision",International Journal On Smart Sensing And Intelligent Systems, 2011, Vol. 5, No. 4.
2. D. Pomerleau And T. Jochem, "Rapidly Adapting Machine Vision For Automated Vehicle Steering," In Ieee Expert, Vol. 11, No. 2, Pp. 19-27, Apr 1996.
3. Fernando Arce, Erik Zamora, Gerardo Hernndez, Humberto Sossa. Efficient Lane Detection Based On Artificial Neural Networks , Isprs Annals Of The Photogrammetry, Remote Sensing And Spatial Information Sciences, Volume Iv-4/W3, 2017 2nd International Conference On Smart Data And Smart Cities, 46 October 2017, Puebla, Mexico.
4. J. Li, X. Mei, D. Prokhorov And D. Tao, "Deep Neural Network For Structural Prediction And Lane Detection In Traffic Scene," In Ieee Transactions On Neural Networks And Learning Systems, Vol. 28, No. 3, Pp. 690-703, March 2017.
5. A. Krizhevsky, I. Sutskever, And G. E. Hinton, Imagenet Classification With Deep Convolutional Neural Networks, In Proc. Adv. Nips, 2012, Pp. 19.
6. B. Alexe, T. Deselaers, And V. Ferrari, Measuring The Objectness Of Image Windows,IEEE Trans. Pattern Anal. Mach. Intell. , Vol. 34, No. 11, Pp. 21892202, Nov. 2012.
7. M.-M. Cheng, Z. Zhang, W.-Y. L In, Andp.Torr, Bing:Binarized Normed Gradients For Objectness Estimation At 300 Fps, In Proc. Ieee Conf. Cvpr , Jun. 2014, Pp. 32863293.
8. G. H. Bakir, T. Hofmann, B. Schlkopf, A. J. Smola, B. Taskar, And S. V. N. Vishwanathan, Predicting Structured Data (Neural Information Processing). Cambridge, Ma, Usa: Mit Press, 2007.
9. Z. Kim, Robust Lane Detection And Tracking In Challenging Scenarios, Ieee Trans. Intell. Transp. Syst., Vol. 9, No. 1, Pp. 1626, Mar. 2008.
10. Kun Qian, Xudong Ma, Xian Zhong Dai, Et Al. Spatial-Temporal Collaborative Sequential Monte Carlo For Mobile Robot Localization In Distributed Intelligent Environments. International Journal On Smart Sensing And Intelligent Systems, 2012, Vol. 5, No. 2, Pp. 295-314.
11. Cretu, A.-M.; Payeur, P. Biologically-Inspired Visual Attention Features For A Vehicle Classification Task. International Journal On Smart Sensing And Intelligent Systems, 2011,Vol. 4, No. 3, Pp. 402-423.

12. Shen Huan,Li Shunming, Miao Xiaodong, Et Al. Intelligent Vehicles Oriented Lane Detection Approach Under Bad Road Scene. IEEE The Ninth International Conference On Computer And Information Technology. Xiamen, China, 2009, Pp.177-182.
13. S. S. Huang, C. J. Chen, P. Y. Hsiao, And L. C. Fu, On-Board Vision System For Lane, Xiaodong Miao, Shunming Li, And Huan Shen, On-Board Lane Detection System For Intelligent Vehicle Based On Monocular Vision 972.
14. Recognition And Front-Vehicle Detection To Enhance Drivers Awareness, IEEE International Conference On Robotics And Automation, 2004, 2456-2461.
15. Z. Zhang. A Flexible New Technique For Camera Calibration. Transactions On Pattern Analysis And Machine Intelligence, 2000, Vol. 19, No. 11, Pp.1330-1334.
16. Yingying Huang, Ross Mcmurran. Development Of An Automated Testing System For Vehicle Infotainment System. Advanced Manufacturing Technology. 2010. Vol. 51, No. 14, Pp.233-246.

A combined method for detecting seven segment digit detection on medical devices

Noppakun Boonsim
Faculty of Applied Science and Engineering
Khonkaen University, Nongkhai campus
Nongkhai, Thailand
boonsim@kku.ac.th

Saranya Kanjaruek
Faculty of Applied Science and Engineering
Khonkaen University, Nongkhai campus
Nongkhai, Thailand
kanjaruek@kku.ac.th

***Abstract*— Biometric data is created by a diversity of medical devices. Manual recording of biometric data from medical devices can be a time-consuming task. Seven segment digit is normally presented on medical devices, for example, blood pressure monitors, glucose meters and digital weight scales, etc. Computer image processing is utilized to automatically analyze seven-segment digit from medical devices for collecting large data sets of biometric data. The objective of this work is to detect a seven-segment digit screen from medical devices. The purposed method begins with seven segment screen detection using a deep learning technique. Afterward, a variety of image processing techniques and parameters are applied to locate the seven-segment digit positions. Experimental results were reported with an accuracy (F-measure) of 94 % utilizing 200 seven-segment digit images.**

Keywords-component; Seven-segment digits; medical device;

I. INTRODUCTION

According to the Thailand Global Health Strategic Framework from 2016 to 2020, one of the expected outcomes is to help people maintain good health. Promotion of healthcare services across Thailand is the strategic objective that helps to provide healthcare as a human right [1]. Collecting bio-metric data from medical device is the beginning scheme to provide healthcare services to society. Normally, health data (such as blood pressure and weight etc.) are collected manually from medical devices (such as glucometer, medical thermometers and digital weight scale) during the medical examination process.

Most of medical devices use seven segment digit (SSD) to represent the measuring data. Seven segment digit is made of seven LEDs (segment) and set as a rectangle. Each segment presents part of a numerical digit including decimal and hex. Therefore, the recognition process for seven segment digit is increasingly important for automatic recording of health data that can be used to predict or spot an early sign of diseases.

Seven segment digit recognition (SSDR) process based on the state of the art, character recognition process, basically consists of three steps, SSD detection, character segmentation and character recognition. SSD detection locates SSD in a medical image. Then, character segmentation separates many SSD regions into character. Each character is classified to a pre-defined class in recognition process. This research presents a method to detect SSD positions which is considered to be an important process because the performance of SSDR depends on the accuracy of the SSD detection.

II. RELATED WORKS

There are a number of previous researches dedicating to recognizing SSD. First, Ghugardare et. al [2] presented combining image processing techniques in order to recognize SSD for measuring instruments in 2006. The techniques, for example image banalization, filtering and projection and template matching were used to recognize SSD. The results were reported with an accuracy of 92 % which tested on 100 digital multi-meter images. However, the work was only tested on clear background images. The performance might be changed if implemented on images with complex background or light changing.

Sathiya et al. [3] proposed a SSD recognition technique on traffic light countdown time. This work began with traffic light detection using color technique and then Sathiya's technique can be extended to locate the countdown time. The research reported 88% of SSD recognition accuracy.

Bonačić et al. [4] presented the neural networks technique to recognize SSD character. This research experimented on variety of neural network topologies to find the best one. The highest tested result was a committee technique of neural network with 99% SSD recognition accuracy.

In 2016, Kulkarni et al. [5] studied a multi-modal technique such as image banalization, tilt correction, background elimination and image filtering to detect LCD display and used pixel density (distribution of black pixels) to recognize the SSD. The study reported recognition accuracy at 79 % which was tested on 800 LCD images including a wild range of variation in illumination and angular tilt conditions.

Lui [6] presented Histogram Orientation Gradient to extract dominant features of SSD within medical devices. SVM was used to classify two classes of detected texts into SSD or non-SSD. Moreover, the research applied Tesseract-OCR engine [7] to recognize SSD. The experimental results were reported that the SSD detection accuracy (F-measure) was 70%, tested on 266 images.

Shenoy and Aalami [8] presented smartphone-based application to automatically recognize SSD used in Apple's HealthKit. The algorithm began with user choosing type of health data to be recorded such as blood pressure, weight and glucose. Next, User defined the position of SSD display by using a prepared bounding rectangle. Finally, the SSD in the defined bounding rectangle was sent to the SSD recognition process. The recognition accuracy was reported at 98.2% and tested on 108 images.

III. METHODOLOGY

This research presents a combining method which tries to cover variation of SSD as much as possible. The proposed method, combining method, compounds with two techniques: deep learning and image processing. First, deep learning technique is used to detect SSD screen. The deep learning is continuously applied in variety of machine learning to increase the efficiency of the training processes. The use of deep learning technique to detect and identify is a plausible solution.

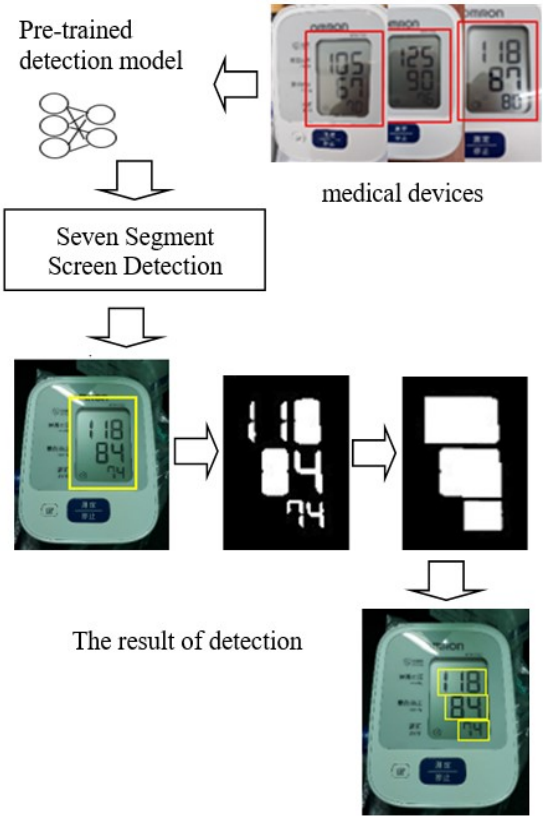


Fig. 1. Seven segment digit detection algorithm

Furthermore, this technique provides high accuracy on object detection and recognition [9]. Therefore, the proposed method utilizes the deep learning technique to detect SSD screen. The detected SSD screen is implemented by variety of image processing techniques to verify the true SSD regions as shown in Fig. 1.

A. *Seven segment screen detection*

SSDs from medical devices are located within SSD screens. The seven segment screen detection can shrink down the image and can reduce the percentage of false alarms (detected from other texts outside the SSD screen). This work applies faster Regional Convolution Neural Networks (fasterRCNN) [10], deep learning technique to detect SSD screen. The fasterRCNN detects SSD screen consists of two processes: training and testing for detecting SSD screen. For the training process, algorithm fits the network by backpropagation convolution neural network and training images (pre-defined position of SSD screen). The proposed research applies VGG19 network in Simonyan and Zisserman work [11] which has 19 layers deep and can classify 1,000 objects such as boats, airplanes and cups etc., in Deng work [12]. Fig. 2 shows the example training images and defined SSD screen positions.

Fig. 2. Example of training images with defined SSD screen position

B. *Seven segment digit detection*

Seven segment digit detection diagram as shown in Fig 3 verifies the true SSD regions from the obtained SSD screen. First step is a detected SSD screen step. After that the screen is converted to gray scale in a gray image conversion and then an adaptive local contrast enhancement is applied. SSD characters are detected in a region detection step by using MSER technique to detect regions size between 50 and 500 as illustrations in Fig. 4(a) and Fig. 4(b) displays binary (mask) image of detected regions.

While, Canny edge detection technique is used to extract SSD edges on the same image, Fig. 4(c) in an edge detection step. The edge detection step, an edge image is applied by morphological operation, dilation, to extend edge size as specified in Fig. 4(d). The result images from the last two steps (the region detection step and the edge detection step) are intersected in order to find potential candidate regions in an image intersection step as displayed in Fig. 4(e). Non-characters are discarded by considering geographic properties, the region aspect ratio value is more than 3, the eccentricity value is more than 0.995, the area is less than 100 pixels and the solidity value is less than 0.3 from Gonzalez et al. [13] in the non-character removal step. Fig. 4(f) shows the remaining regions and can be mapped to locate the SSD in the original image, as displayed in Fig. 3(g). Then, remaining characters are merged into region or text in a character merging step. Finally, region high (more than 27) and aspect ratio (between 1 and 4) are used to verify the true SSD regions in SSD verification as shown in Fig. 4(h).

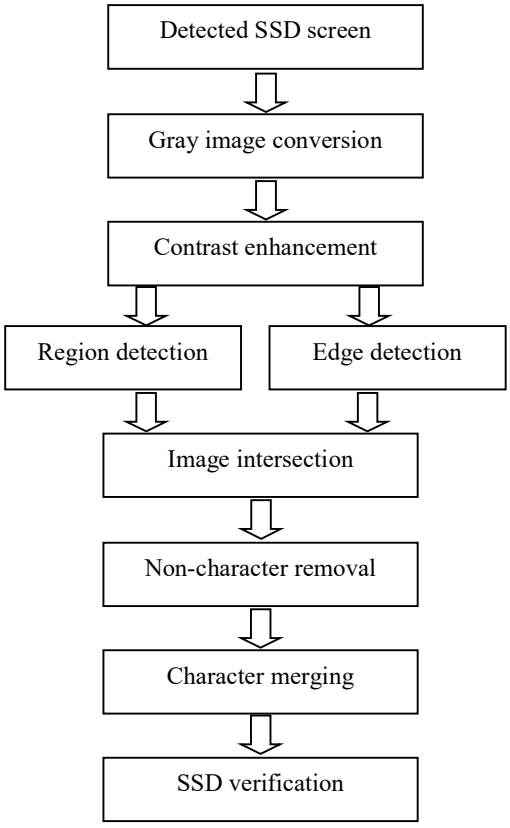


Fig. 3. Seven segment digit detection diagram

IV. EXPERIMENTAL RESULTS

A. Dataset and tool

The dataset has 400 images which were collected from internet 200 and captured from blood pressure meter 200 images which were separated into training and testing images. The experiment used MATLAB version 2018b (trial version) to implement the algorithm. The average computation time per image was less than 1 second. The performance of the research is evaluated by precision recall and F-measure value.

B. Results

Table I shows the comparison of SSD detection from other works and experimental results of the proposed method. The research results show SSD detection accuracy precision recall and F-measure 91%, 97% and 94% respectively. According to table I, the combined method shows higher performance than the existing works.

TABLE I. THE COMPARISON OF SSD DETECTION WORKS

Works	#	Technique	P	R	F
Ghugardare [2]	110	Image processing	N/A	N/A	0.96
Sathiya [3]	N/A	Color based	0.97	0.95	0.96
Lui [6]	266	HOG+SVM	0.76	0.55	0.70
Combined technique	200	FasterRCNN + image processing	0.91	0.97	0.94

C. Discussion

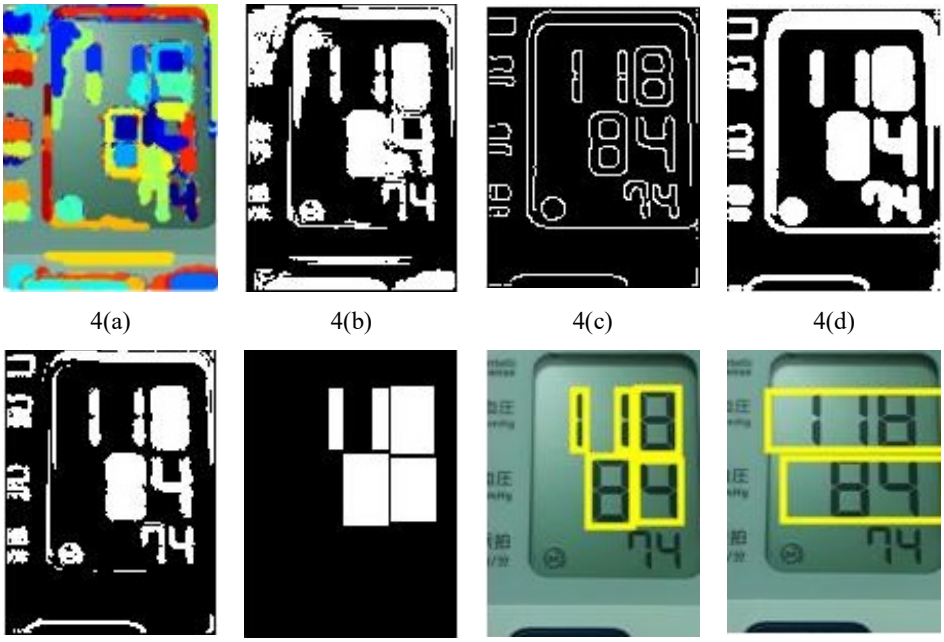
The research detection accuracy was reported at 94% which were tested on 200 images that is acceptable compared to typical values from existing results. With the proposed technique, deep learning technique, detects SSD screen before detects SSD text.

Although, the combined technique has shown high performance but the experimental dataset was small (200 images for training and 200 images for testing). Moreover, there are some problems of this method for example the proposed method cannot properly detect SSD screen as shown in Fig. 5(a-b) and sometime this method detects other similar text on screen, Fig. 5(c).

V. CONCLUSION

This research applies deep learning technique and combines with image processing techniques that attain 94% accuracy on SSD detection. Furthermore, deep learning technique can detect rotated SSD screen and partial occlusion SSD screen. However, this technique has some limitations such as the missed SSD screen detection and the detect larger area than SSD screen that provide more false alarms.

For future work, the performance might be improved if more training images are employed. Moreover, the research applied VGG19 network which has 19 layers if used with other deep networks the performance might be changed. When the detection is completed, recognition process will be implemented in order to automatically collect bio-metric data. The increasing of SSD detection directly affects the correct recognition rate. Moreover, the experiment should be based on real-life situations to justify the performance.



4(a) 4(b) 4(c) 4(d)

4(e) 4(f) 4(g) 4(h)

Fig. 4. Example images of SSD detection steps

5 (a) 5 (b) 5 (c)

Fig. 5. Examples of error detection

REFERENCES

[1] Ministry of public health, “Thailand Global Heath Strategic Framework 2016-2020,” Internet: www.bihmoph.net/download/czetH0Z6g2Wt.pdf, Dec. 2016, [Dec. 1, 2018].

[2] R. P. Ghugardare, S. P. Narote, P. Mukherji and P. M. Kulkarn, "Optical character recognition system for seven segment display images of measuring instruments," in *Proceeding of the IEEE Region 10 Conference (TENCON 2009)*, 2009.

[3] S. Sathiya, M. Balasubramanian, and D. V. Priya, “Real time recognition of traffic light and their signal count-down timings,” in *Proceeding of IEEE International Conference on Infomation Communication Embedded System*, 2014, pp. 1–6.

[4] I. Bonačić, T. Herman, T. Krznar, E. Mangić, G. Molnar and M. Čupić, “Optical Character Recognition of Seven-segment Display Digits Using Neural Networks,” in *Proceeding of the 32st International Convention on Information and Communication Technology, Electronics and Microelectronics*, Vol 3, 2015.

[5] P. H. Kulkarni and P. D. Kute, "Optical numeral recognition algorithm for seven segment display," in *Proceeding of the IEEE Conference on Advances in Signal Processing (CASP),* 2016, pp. 397-401.

[6] C. Liu, “Digits Recognition on Medical Device,” M.S Thesis, The University of Western Ontario, Canada, 2016.

[7] R. Smith, "An overview of the Tesseract OCR engine," in *Proceeding of the IEEE 9th International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 2, 2007.

[8] V. N. Shenoy and O. O. Aalami, “Utilizing Smartphone-Based Machine Learning in Medical Monitor Data Collection: Seven Segment Digit Recognition,” in *Proceeding of the American Medical Informatics Assosication Annual Symposium (AMIA)*, Vol. 2017, pp. 1564-1570.

[9] A. Krizhevsky, S. Ilya and H. E. Geoffrey, "Imagenet classification with deep convolutional neural networks," in *Proceeding of Advances in neural information processing systems*, 2012, pp. 1097-1105.

[10] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), pp. 1137-1149, 2017.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceeding of International Conference Learning Representation*s, 2015

[12] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[13] A. Gonzalez, L. M. Bergasa, J. J. Yebes and S. Bronte, "Text location in complex images," ," in *Proceeding of the IEEE 21st International Conference on Patten Recognition*, 2012.

Synchronization Control for Microgrid Seamless Reconnection

Thakul Uten
School of ICT
SIIT, Thammasat University
Pathum Thani 12120, Thailand
thakul.ute@gmail.com

Chalie Charoenlarpnopparut
School of ICT
SIIT, Thammasat University
Pathum Thani 12120, Thailand
chalie@siit.tu.ac.th

Prapun Suksompong
School of ICT
SIIT, Thammasat University
Pathum Thani 12120, Thailand
prapun@siit.tu.ac.th

***Abstract*—In the three phase distribution network, the microgrid is connected the utility grid at the point of common coupling (PCC) and can operate either in a grid-connected or standalone modes. However, at the moment of transition mode from standalone to grid-connected mode if the microgrid is reconnected to the utility grid without synchronization, the out-of-phase reclosing might be occur and lead to serious consequences e.g. large inrush current, overvoltage, system oscillations, and damage to system equipment itself. Thus, in order to achieve microgrid seamless reconnection, it is necessary to synchronize frequency, phase and amplitude voltage of the microgrid to the utility grid before reconnection. In this paper, a synchronization control has been proposed for microgrid seamless reconnection. The additional Distributed Generator (DG) is installed at PCC and assigned to be Dispatch Unit (DU). DU is controlled by Droop Control for generating the electric power to adjust frequency, phase and amplitude voltage of the microgrid at PCC during transfer to grid-connected mode. The proposed method is performed by eliminating the difference of frequency, phase and amplitude voltage during transition modes. The simulation has been performed in Matlab/Simulink.**

Keywords—Droop Control, Distributed Generator (DG), Dispatch Unit (DU), Smooth Transition

I. INTRODUCTION

To improve power quality and reliability of power supply, microgrids are used in many area of the electrical grid by integrating control systems, communication infrastructure, Distributed Generations (DGs) and also Distributed Energy Storages (DESs) with the local load in a distribution network. A microgrid is connected to the utility grid through circuit breaker or switch at the point of common coupling (PCC) and can operate either in a grid-connected or standalone mode with DGs or DESs supplying power to local loads.

Normally, a microgrid is operated on the grid-connected mode. When the fault occur in the grid, the circuit breaker or switch at PCC will operate in order to isolate a microgrid from the utility grid by the cause of reliability and stability of the utilities grid. It mean, a microgrid will be switched to the standalone mode. At this mode, a microgrid might be loss reliability and stability due to power imbalance and losing synchronization with the power system. After the fault was cleared, a microgrid will be transferred back to the grid-connected mode again. However, at the moment of reconnection if a microgrid is reconnected to the utility grid without synchronization, the out-of-phase reclosing might be occur and could be lead to serious consequences e.g. large inrush current, overvoltage, system oscillations, and damage to system equipment itself.

Thus, it is a great importance to develop a control strategy to improve the synchronizing process that can avoid the out-of-phase reclosing problem and increase an electrical stability as well. To achieve this goals, it is necessary to synchronize the microgrid voltage, frequency, and phase to the utility grid before transition modes.

Generally, the synchronization control strategy have two significant part, which are the phase detection technique and the synchronization control algorithm. Phase-Locked Loop (PLL) has been the state-of-the-art in phase detecting the phase angle of an input signal [1]. In case of three-phase system, the Synchronous Reference Frame PLL (SRF-PLL) is the one of PLL techniques that be used [2-8]. The SRF-PLL technique is extremely simple and provides a highly fast and accurate results in terms of synchronization when there is no distorted input signal or any unbalanced loads [9]. Nevertheless, [2] under distorted and/or unbalanced supply voltages, the extracted phase angle will have error. Then, Fran González-Espín et al.[10] proposed the Adaptive SRF-PLL technique which has a high rejection of the disturbances introduced by the voltage imbalance and by the voltage harmonic distortion, regardless of the grid frequency variation. Then, Di Shi et al. [11] proposed the newest phase detection by using Phasor Measurement Unit (PMU). PMU has the capability of precisely tagging the timestamp of voltage and current measurements. However, when compare the SRF-PLL with PMU, the cost of PMUs are quite extremely high which might be not suitable for using in the microgrid level. From literature, in order to use for three-phase system under balance voltage without harmonic distortion condition as well as can extract the phase angle. The SRF-PLL [2] is the most interesting technique that can operate under condition with simple and provides a highly fast and accurate results in terms of synchronization.

For the synchronization control algorithm which simple and DGs can cooperate without communication system is the control algorithm based on a droop control concept. A droop control is a kind of cooperative control that allows parallel connection of DGs sharing active and reactive powers. With the objective of paralleling DG units, the voltage and frequency reference of the grid will be generated. The droop control is responsible for adjusting the phase and the amplitude of the voltage reference of according to the active and reactive powers (P and Q) [12]. The conventional P–f and Q–V droop control technique is the most commonly used methodology for the DGs to share active and reactive powers [13].

From literature, a synchronization control based on droop control can be separated by synchronization algorithm into 3 types:

First type, synchronized the voltage by share compensated active and reactive power from all DGs in the microgrid [14-18]. Second type, synchronized the voltage by the synchronizing DG unit. The synchronizing DG unit is the one of DGs in the microgrid that was selected for the synchronization of the microgrid. To handle this responsibility, The synchronizing DG unit should be sufficiently large, dispatchable and close to the PCC [19]. And third type, synchronized the voltage by using the extra DG unit. The extra DG unit is a new DG that will be added to the microgrid and has been given extra responsibility to manage the voltage amplitude, frequency and phasor during transition modes only [8][20].

From many techniques, a synchronization control based on droop control with synchronized the voltage by the extra DG unit is interesting algorithm. This algorithm is suitable for the microgrid that does not have enough power to adjust the voltage and frequency to synchronize the system during transition modes. This algorithm allow to add the extra DG for taking specific duty to synchronize the amplitude, frequency, and phasor of voltage during transition modes only. The advantages are the DGs and DESs that connected the microgrid do not operate or feedout power over the limited of capacity during transition modes that have to compensate the active and reactive power for synchronizing the microgrid with the utility grid.

II. MICROGRID AND CONTROL STRATEGY

A. *Microgrid Configuration*

The microgrid model consist of at least two DGs to supply power to the grid, local loads with balance load. One of DGs is assigned as a Dispatch Unit (DU) which is responsible to compensate the active and reactive power for synchronizing during transition modes. The DGs and DU are connected to microgrid and microgrid is connected to the utility grid at PCC. Besides, DU shall be connected to the microgrid as nearest the synchronized position as possible, shown in Fig 1.

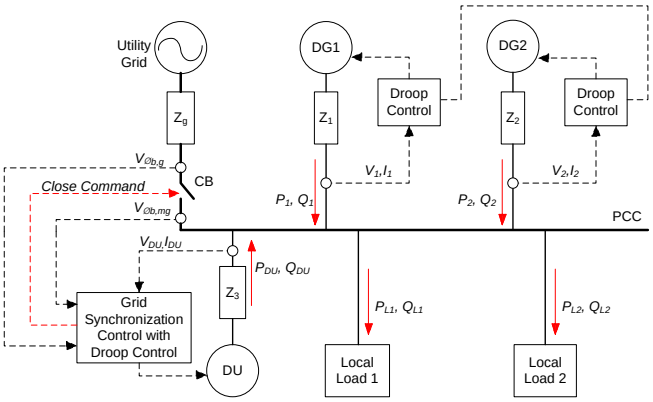


Fig. 1 Microgrid configuration with DU

B. *Phase Detection and Control Strategy*

The voltage, frequency and Phase angle of the microgrid and the utility are detected by the SRF-PLL [2] in V,I Transformation Process, as shown in Fig. 2.

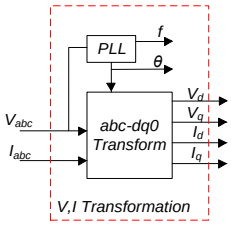


Fig. 2 SRF-PLL Phase detection

The control algorithm of the DGs is based on a droop control. The conventional P–f and Q–v droop control is used to sharing active and reactive powers between DGs in microgrid. The voltage and frequency of microgrid are regulated by droop control, as shown in Fig. 3.

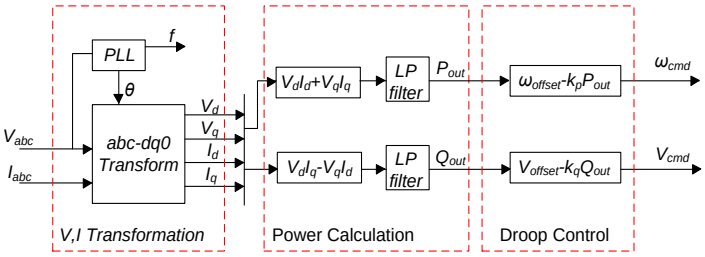


Fig. 3 Block diagram of droop control in DGs.

The droop control relations for the DGs in microgrid are given by

$$\omega_n = \omega_{offset} - k_{pn}P_n \quad (1)$$

$$V_n = V_{offset} - k_{qn}Q_n \quad (2)$$

Where ω_n and ω_{offset} are the output and offset frequency, respectively. V_n and V_{offset} are the output and offset voltages, respectively. P_n and Q_n are the real and reactive powers of each DGs in the microgrid, respectively. And k_{pn} and k_{qn} are the droop parameters. In this relations, ω_{offset} and V_{offset} are adjusted higher than the nominal value of grid frequency and voltage ($\omega_{nominal}$ and $V_{nominal}$, respectively). Droop coefficients can be determined as

$$k_{pn} = d\omega/P_{max,n} \quad (3)$$

$$k_{qn} = dV/Q_{max,n} \quad (4)$$

Where $d\omega$ and dV are the maximum allowable frequency and voltage deviation, respectively. And $P_{max,n}$ and $Q_{max,n}$ are the maximum output of active and reactive power of the n^{th} DG in the microgrid, respectively.

The droop control characteristic of DGs in standalone mode, as shown in Fig. 4.

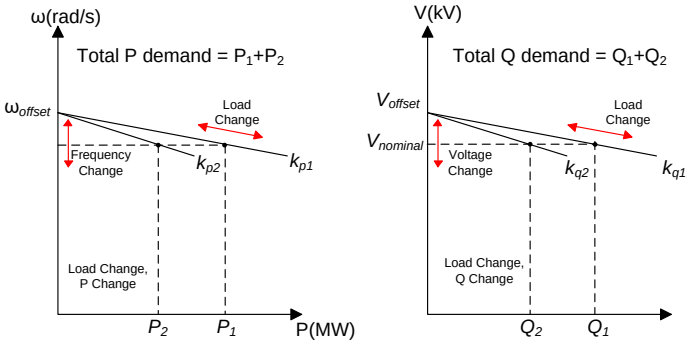


Fig. 4 Droop control characteristic in standalone mode

The control algorithm of the DU during transition mode from standalone to grid-connected mode is combine between the droop control and the synchronization control, as shown in Fig. 5. The voltage, frequency and phase angle in the microgrid are regulated for synchronization by adjustment of $\omega_{offset,du}$ during transition. The droop control characteristic of DU during transition mode, as shown in Fig. 6.

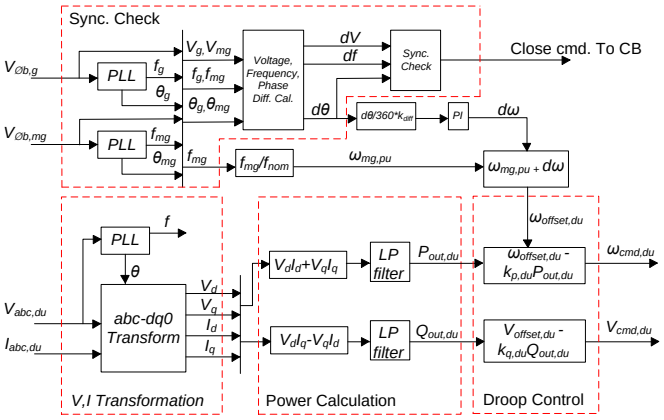


Fig. 5 Block diagram of droop control in DU

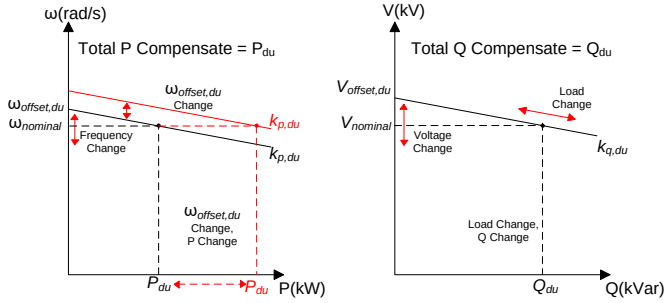


Fig. 6 Droop control characteristic during transition mode

The droop control relations for the DU in microgrid are given by

$$\omega_{du} = \omega_{offset,du} - k_{p,du}P_{out,du} \quad (5)$$

where $\omega_{offset,du} = \omega_{mg,pu} + d\omega$

$$V_{du} = V_{offset,du} - k_{q,du}Q_{out,du} \quad (6)$$

Where $\omega_{mg,pu}$ and $d\omega$ are the frequency of microgrid in per-unit at nearest synchronized position and the difference of frequency between the utility grid and microgrid, respectively. ω_{du} and $\omega_{offset,du}$ are the output and offset frequency of DU, respectively. V_{du} and $V_{offset,du}$ are the output and offset voltages of DU, respectively. $P_{out,du}$ and $Q_{out,du}$ are the real and reactive powers of DU, respectively. And $k_{p,du}$ and $k_{q,du}$ are the droop parameters of DU.

In standalone and grid-connected mode, DU will operate in idle mode.

$$\omega_{offset,du} = \omega_{mg,pu} \quad where\ d\omega = 0 \quad (7)$$

And during transition mode from standalone to grid-connected mode, DU will adjust its $\omega_{offset,du}$ by adding $d\omega$.

$$d\omega = ((k_p s + k_i)/s).(|\theta_g - \theta_{mg}|/(360*k_{diff})) \quad (8)$$

Where k_p and k_i are the PI controller coefficient. θ_g and θ_{mg} are the utility grid and microgrid voltage phase angle, respectively. And k_{diff} is the adaptive coefficient of the difference of voltage phase angle.

From adjustment of droop control characteristic, DU will feed the real and reactive powers into microgrid to adjust voltage, frequency, phase angle until two grid system are synchronized.

The operation of DU during transition from standalone to grid-connected mode, as shown in Fig 7.

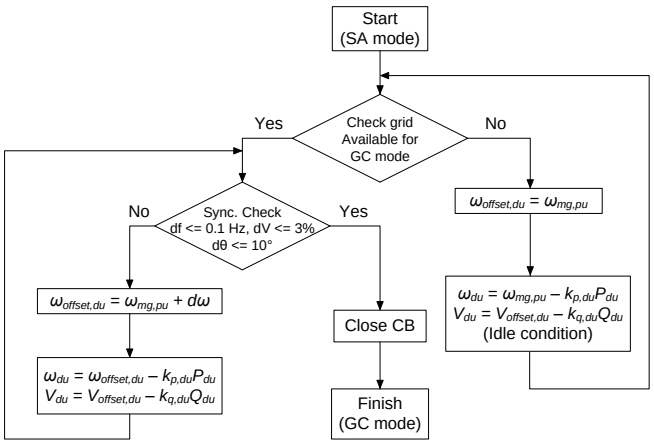


Fig. 7 The operation of the DU during transition from standalone to grid-connected mode

C. Synchronized Check

The synchronized check as shown in Fig. 8 with the synchronization parameter limits are following the limits specified in IEEE 1547-2003 standard, as shown in table I.

TABLE I. THE LIMITS SPECIFIED IN IEEE 1547-2003 STANDARD

Aggregate rating of DR units (kVA)	*Frequency difference (df, Hz)*	*Voltage difference (dV, %)*	*Phase angle difference (dθ, degree)*
0-500	0.3	10	20
>500-1500	0.2	5	15
>1500-10000	0.1	3	10

In this paper, Phase B is used for phase reference in synchronization process.

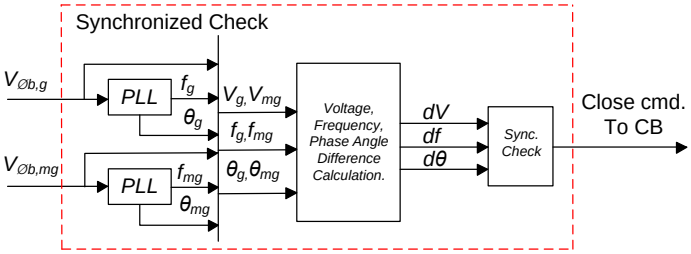


Fig. 8 Synchronized check

Where $V_{Øb,g}$ and $V_{Øb,mg}$ are the utility and microgrid voltage (phase B), respectively. V_g and V_{mg} are the utility and microgrid amplitude voltage, respectively. f_g and f_{mg} are the utility and microgrid frequency, respectively. θ_g and θ_{mg} are the utility and microgrid voltage phase angle, respectively. dV, df and $d\theta$ are the difference of amplitude voltage, frequency and voltage phase angle, respectively. During synchronized check if all parameters (dV, df and $d\theta$) are in the limits value, it will send the signal to close circuit breaker. Then, the microgrid will reconnect to utility grid with seamless reconnection.

The proposed strategy is implemented and simulated in Matlab/Simulink. Simulate under balance voltage without harmonic distortion.

III. Experimental Result and Discussion

In the simulation model, the grid code of the system is base on the Provincial Electricity Authority (PEA) distribution system. This model consists of two DGs, two local loads and one DU as shown in Fig.1. DGs and DU are the generator which controlled by droop control that mentioned in Chapter II. DU is located as nearest the synchronized position as possible. The microgrid is connected to the utility grid at PCC through circuit breaker. The system parameters are given in Table II.

TABLE II. Parameter of the Simulation Model

Parameter	***Value***	***Parameter***	***Value***
V_{base}	22 kV	dV	3 %
f_{base}	50 Hz	df	0.1 Hz
VA_{base}	50 MVA	$d\theta$	1 degree
P_{DG1}	5.40 MW	k_{p1}	0.1852
Q_{DG1}	2.62 MVar	k_{q1}	1.9118
P_{DG2}	3.60 MW	k_{p2}	0.2778
Q_{DG2}	1.74 MVar	k_{q2}	2.8677
P_{load1}	4.47 MW	ω_{offset}	1.0140 pu
Q_{load1}	1.56 MVar	V_{offset}	1.8197 pu
P_{load2}	4.28 MW	k_p	4.5
Q_{load2}	1.41 MVar	k_i	1
In the simulation model, the difference of voltage phase angle $d\theta$ is reduced from 10º to 1º for seamless reconnection as much as possible. All parameters are converted and used in Per-Unit (pu).			

Nevertheless, two important factors that affect to the synchronized process directly must be concern. The first factor is a dθ during synchronization that can be any value (0 – 360 degree). It mean that if $d\theta$ is high, it might take more time to synchronized.

The second factor is the capacity of DU which affect directly to synchronization time and peak power as well. In this paper, the $d\theta$ values are defined as 5, 10, 15, 20, 25, 30, 35, 40 and 45 degree, respectively. Since the capacity of DU are not more than 2.5% of aggregate capacity of DGs in microgrid, the capacity of DU is defined as 50, 100, 150, 200 and 250 kVA, respectively.

A. Droop Control Parameter

The relation between k_{diff} and synchronization time (T_{sync}) as well as peak power (P_{max}) of DU during transition mode. From the graph as shown in Fig. 9 and Fig. 10, show the relation between k_{diff} and T_{sync}, k_{diff} and P_{max} where $d\theta$ = 30º, respectively.

From graph in Fig. 9, T_{sync} will gradually increase when k_{diff} is increased. On the other hand, in Fig. 10, P_{max} trend to decrease while k_{diff} is increasing.

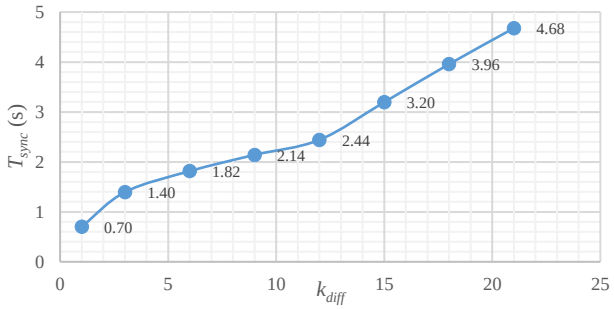


Fig. 9 k_{diff} – T_{sync} relationship

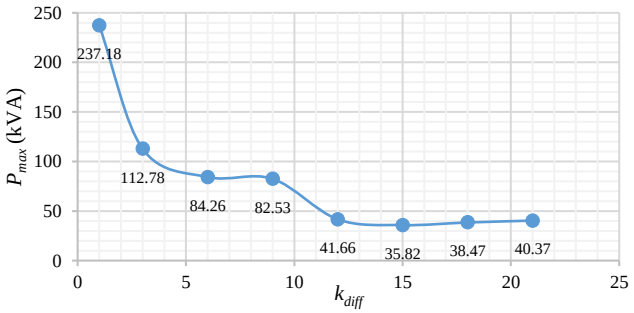


Fig. 10 k_{diff} – P_{max} relationship

From this relation, using low k_{diff} value will take short time to synchronized system but it need k_{diff} value from 12 to 21 are interesting because DU at the same capacity but synchronization time can change via adjust k_{diff}. So, graph as shown in Fig. 11 show the trend of k_{diff} and P_{max} at any $d\theta$ where DU is same capacity.

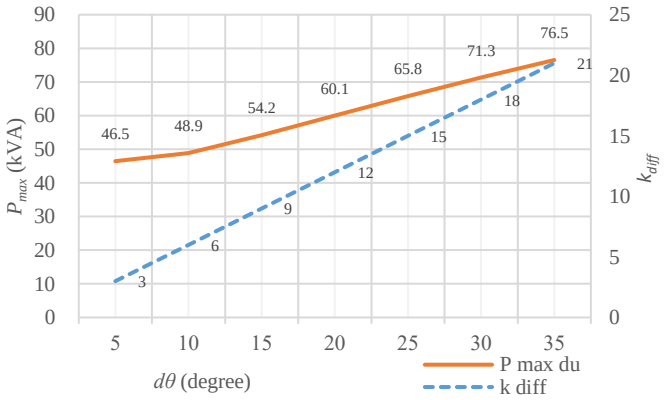


Fig. 11 $d\theta$- k_{diff} , P_{max} relationship

Thus, k_{diff} should be concerned and defined with suitable value. In this paper, the value of k_{diff} and $d\theta$ in simulation model are given in Table III.

TABLE III. Parameter of Droop Control

Parameter	***Value***		
$d\theta$ *(degree)*	5	10	15
k_{diff}	3	6	9
$d\theta$ *(degree)*	20	25	30
k_{diff}	12	15	18
$d\theta$ *(degree)*	35	40	45
k_{diff}	21	24	27
Where DU capacity are 50, 100, 150, 200 and 250 kVA			

B. Simulation Results

The system parameter are following as Table II and III. The Fig. 12 show the *dV*, *df* and *dθ* between microgrid and the utility grid before synchronization process where *dθ* is 30 degree and DU capacity is 200 kVA. At the initial state, the microgrid is operated in Stand-alone mode with DU that is in idle mode. The synchronization process is started at *t*=1.00 s and completed at *t*=4.38 s. It is mean that time consumed by two grid synchronization process is around 3.38 s.

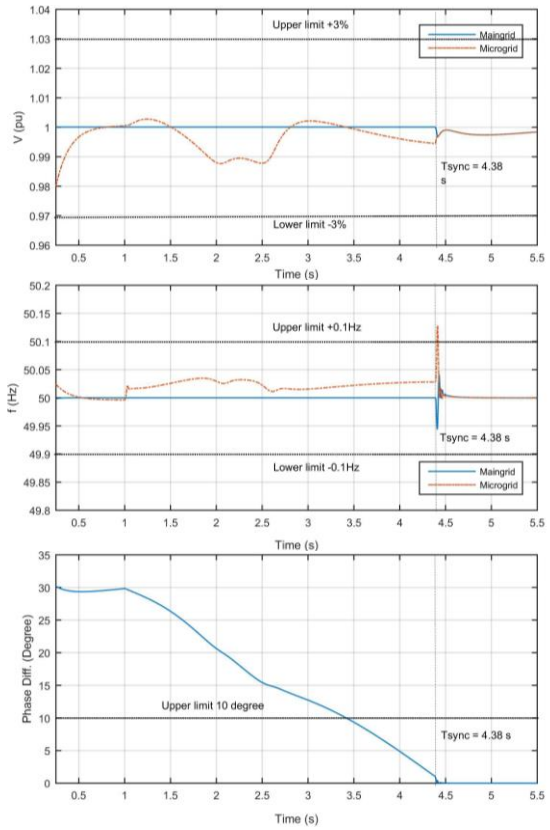


Fig. 12 the *dV*, *df* and *dθ* between two grids

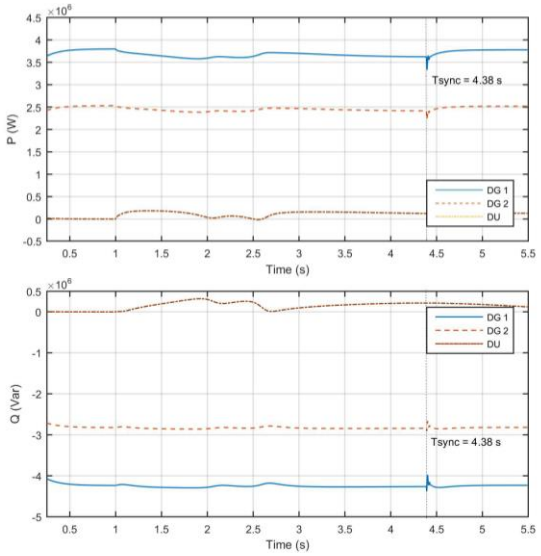


Fig. 13 Power injected by DU during Synchronization

During Synchronization process, the DGs and DU are controlled by Droop Control. The *dV*, *df* and *dθ* were eliminated by DU which will inject the power to adjust the voltage magnitude and frequency of microgrid. When the *dV*, *df* and *dθ* are in the criteria, the synchronized check will send a command to close a circuit breaker in order to reconnect the microgrid and the utility grid. From the graph in Fig.13, there are some power oscillation because of the slightly variation of *dV*, *df* and *dθ* that be generated during close circuit breaker. After microgrid were in grid-connected mode, DU will change to idle mode again.

From the control strategy, when compare the result between microgrid reconnection with and without synchronized. While circuit breaker was closing to reconnect microgrid to utility grid, Power oscillation can be reduced if *dθ* is decreased. In Fig. 14 show power during microgrid reconnection with *dθ* are 30, 20, 10 and 1 degree, respectively. From the graph in Fig. 14, Peak powers are around 3.65/2.72/1.83/1.08 times of nominal power where *dθ* are 30, 20, 10 and 1 degree, respectively. The comparison between a case of *dθ* = 10 degree (IEEE standard) and a case of *dθ* = 1 degree, a power oscillation still high. Thus, for seamless reconnection as much as possible, in this paper, the *dθ* criteria is reduced from 10 degree to 1 degree.

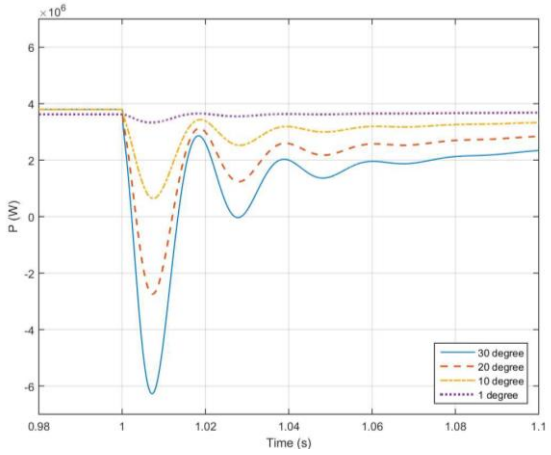


Fig. 14 Power oscillation during reconnection

In another case, the *dθ* and DU capacity comparison is shown in Fig.15. From graph in Fig.15, when T_{sync} is compare with identical *dθ* in various DU capacity. T_{sync} tend to decrease dramatically when the DU capacity increase. On the other hand, T_{sync} is depend on *dθ* when compare with identical DU capacity.

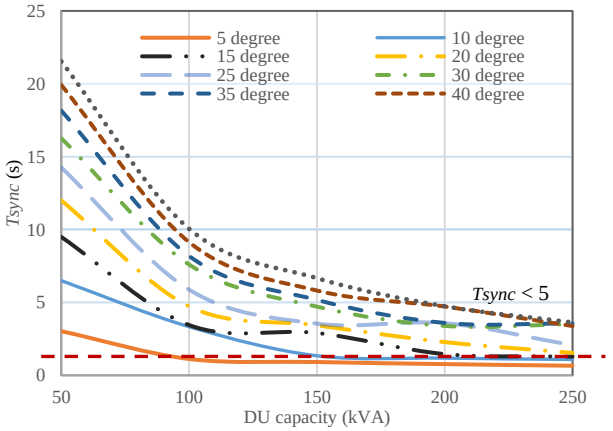


Fig. 15 DU capacity compare with *dθ and* T_{sync}

Therefore, from the simulation result the DU capacity is not more than 2.5% of aggregate capacity of DGs in microgrid and the T_{sync} less than 5 s are suitable and acceptable in this paper.

IV. CONCLUSION

In this paper, the synchronization control for microgrid seamless reconnection is presented. For this control strategy can be used in the three phase distribution network with microgrid in order to synchronize between microgrid and utility grid by adding DU that its capacity not more than 2.5% of aggregate capacity of DGs in microgrid with T_{sync} less than 5 s. The purpose control strategy could be used to reduce power oscillation and also adapted with zero-crossing technique in order to decrease the voltage oscillation and more smoothly transfer to grid-connected mode as well.

ACKNOWLEDGMENT

The researcher would like to thank Provincial Electricity Authority (PEA) and Sirindhorn International Institute of Technology (SIIT) that given an good opportunity and also supported the scholarship for study further in SIIT.

REFERENCES

[1] Felice Liccardo, Pompeo Marino, and Giuliano Raimondo, "Robust and Fast Three-Phase PLL Tracking System," IEEE Transactions on Industrial Electronics., vol. 58, no. 1, pp. 221-231, Mar. 2010.

[2] Se-Kyo Chung, "A Phase Tracking System for Three Phase Utility Interface Inverters," IEEE Transactions on Power Electronics., vol. 15, no. 3, pp. 431-438, Aug. 2002.

[3] Yilmaz Sozer, and David A. Torrey, "Modeling and Control of Utility Interactive Inverters," IEEE Transactions on Power Electronics., vol. 24, no. 11, pp. 2475-2483, Aug. 2009.

[4] Irvin J. Balaguer, Qin Lei, Shuitao Yang, Uthane Supatti, and Fang Zheng Peng, "Control for Grid-Connected and Intentional Islanding Operations of Distributed Power Generation," IEEE Transactions on Industrial Electronics., vol. 58, no. 1, pp. 147-157, May. 2010.

[5] V. Kaura, and V. Blasko, "Operation of a phase locked loop system under distorted utility conditions," IEEE Transactions on Industry Applications., vol. 33, no. 1, pp. 58-63, Feb. 1997.

[6] Salvatore D'Arco, and Jon Are Suul, "A Synchronization Controller for Grid Reconnection of Islanded Virtual Synchronous Machines," Power Electronics for Distributed Generation Systems (PEDG), 2015 IEEE 6th International Symposium on., Aug. 2015.

[7] Tine L.Vandoorn, Bart Meersman, Jeroen D.M. De Kooning, and Lieven Vandevelde, "Transition From Islanded to Grid-Connected Mode of Microgrids With Voltage-Based Droop Control," IEEE Transactions on Power Systems., vol. 28, no. 3, pp. 2545-2553, Mar. 2013.

[8] Md. Nayeem Arafat, Ali Elrayyah, and Yilmaz Sozer, "An Effective Smooth Transition Control Strategy Using Droop-Based Synchronization for Parallel Inverters," IEEE Transactions on Industry Applications., vol. 51, no. 3, pp. 2443-2454, Nov. 2014.

[9] Sadaf Sadeghian Sorkhabi, and Alireza Bakhshai, "Microgrid Control Strategies and Synchronization Techniques during Transition between Grid-Connected and Stand-alone Mode of Operation," Telecommunications Energy Conference (INTELEC), 2015 IEEE International., Sep. 2016.

[10] Fran Gonzalez-Espin, Emilio Figueres, and Gabriel Garcera, "An Adaptive Synchronous-Reference-Frame Phase-Locked Loop for Power Quality Improvement in a Polluted Utility Grid," IEEE Transactions on Industrial Electronics., vol. 59, no. 6, pp. 2718-2731, Oct. 2011.

[11] Di Shi, Yusheng Luo, and Ratnesh K. Sharma, "Active Synchronization Control for Microgrid Reconnection after Islanding," Innovative Smart Grid Technologies Conference Europe(ISGT-Europe), 2014 IEEE PES., Feb. 2015.

[12] Juan C. Vasquez, Josep M. Guerrero, Mehdi Savaghebi, Joaquin Eloy-Garcia, and Remus Teodorescu, "Modeling, Analysis, and Design of Stationary-Reference-Frame Droop-Controlled Parallel Three-Phase Voltage Source Inverters," IEEE Transactions on Industrial Electronics., vol. 60, no. 4, pp. 1271-1280, Apr. 2013.

[13] Somesh Bhattacharya, and Sukumar Mishra, "Efficient power sharing approach for photovoltaic generation based microgrids," IET Renewable Power Generation., vol. 10, no. 7, pp.973-987, Jul. 2016.

[14] Fen Tang, Josep M. Guerrero, Juan C. Vasquez, Dan Wu, and Lexuan Meng, "Distributed Active Synchronization Strategy for Microgrid Seamless Reconnection to the Grid Under Unbalance and Harmonic Distortion," IEEE Transactions on Smart Grid., vol. 6, no. 6, pp. 2757-2769, Mar. 2015.

[15] Masoud Karimi-Ghartemani, Prasanna Piya, Mohammad Ebrahimi, and S. Ali Khajehoddin, "A universal controller for grid-connected and autonomous operation of three-phase DC/AC converters," Energy Conversion Congress and Exposition (ECCE), 2015 IEEE., Oct. 2015.

[16] Dhananjay Gautam, and Hema Rani P, "Microgrid System Advanced Control In Islanded and Grid Connected Mode," Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on., Jan. 2015.

[17] Pornchai Chaweewat, jai Govind Singh, Weerakorn Ongsakul, A.K. Srivastava, "Synchronization Control and Droop Control of Microgrid Operation," Green Energy for Sustainable Development (ICUE), 2014 International Conference and Utility Exhibition on., Jun. 2014.

[18] Yunwei Li, D.M. Vilathgamuwa, and Poh Chiang Loh, "Design, analysis, and real-time testing of a controller for multibus microgrid system," IEEE Transactions on Power Electronics., vol. 19, no. 5, pp. 1195-1204, Sep. 2004.

[19] Tine L.Vandoorn, Bart Meersman, Jeroen D.M. De Kooning, and Lieven Vandevelde, "Transition From Islanded to Grid-Connected Mode of Microgrids With Voltage-Based Droop Control," IEEE Transactions on Power Systems., vol. 28, no. 3, pp. 2545-2553, Mar. 2013.

[20] Zhongwei Chen, Wei Zhang, Jiuqing Cai, Tao Cai, Zhiqiang Xu, and Nana Yan, "A Synchronization Control Method for Micro-Grid with Droop control," Energy Conversion Congress and Exposition (ECCE), 2015 IEEE., Oct. 2015.

Dialogue Breakdown Detection for Understanding Comics with Deep Learning

1st Ryo Iwasaki
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
iwasaki@ss.cs.osakafu-u.ac.jp

2nd Naoki Mori
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
mori@cs.osakafu-u.ac.jp

***Abstract*—For research on computer understanding of comics, due to the copyright, we suffer from the number of data. Thus, we require a model which understands comic contents with less data. Unsupervised models can be trained on data without annotations. We compared two unsupervised approaches—bidirectional encoder representations from transformers (BERT) and Skip-Thought—on Japanese comics because BERT achieved state-of-the-art results on the many datasets and Skip-Thought was a models learning the continuity of text. After all, BERT outperformed Skip-Thought. However, the models did not solves problems with unique expressions which datasets for dialogue breakdown detection did not contain but comics datasets. We concluded that we needed more comics data or ordinary conversation data similar to talks in comics for computer understanding of comics.**

***Index Terms*—natural language processing, japanese comics, unsupervised learning, dialogue breakdown detection**

I. Introduction

Bidirectional encoder representations from transformers (BERT) achieved state-of-the-art results on many datasets [1]. BERT is pre-trained using unsupervised learning and is subsequently fine-tuned. The approach is similar to VGG-16 model [2] in computer vision. However, models are seldom fine-tuned in natural language processing (NLP).

In NLP, models have typically used pre-trained word vectors: word embeddings [3], [4]. Additionally, models that learn generic features at the sentence level have been developed. Except for word embedding layers, these NLP models have been trained from scratch.

Generic vectors are relatively easy to obtain for words but are very difficult to obtain for sentences. To obtain sentence vectors, two approaches exist: training on specific tasks and unsupervised learning. Sentence features produced by a model trained on specific tasks are good for these tasks, but we doubt whether the features are generic or not. On the other hand, we think that vectors produced by unsupervised learning are more generic, but in many cases the vectors have a low quality. Additionally, we consider that the vectors produced by the encoder-decoder models trained with unsupervised learning are features where the inputs are compressed, while the original inputs have more information.

This work was supported by JSPS KAKENHI Grant, Grant-in-Aid for Scientific Research(B), 19H04184.

Despite the compression, it may be worth to obtain generic sentence features. Although machine learning requires a lot of training data to achieve good performance, it can be difficult to prepare some types of data, for example, comics in Japanese. In this case, unsupervised approaches can be considered.

For research on computer understanding of comics, less datasets exist because of copyright issues. To avoid them, exisiting datasets should be used, but they have problems: a small number of data and the lack of information to analyze in NLP. It is difficult to increase the number of comic data, and thus we use also data not produced by comics.

In this paper, we compare two unsupervised approaches—BERT and Skip-Thought [5]—on Japanese text. We solve the problem of dialog breakdown detection in computer understanding of comics. In Section 2, we describe related work. In Section 3, we introduce Skip-Thought and BERT. In Section 4, we show the experimental setup and our results. Finally, we show the conclusion.

II. Related Work

In this section, we describe previous research on computer understanding of comics, with related NLP models.

A. Understanding comics with deel learning

Firstly, from the aspect of comic as data for computing research such as tasks in image processing; Rigaud et al. [6], [7] focused on balloons and their association to comic characters and compared the performance of optical character recognition to extract speech text in comics. Matsui et al. [8] proposed a manga-specific image retrieval system. This system retrieved comics that had pictures resembling sketches that people had drawn. Fujino et al. [9] focused on four-scene comics. They solved an order recognition task for understanding structures in comics. All of them analyzed only pictures in comics or annotations closely related to the pictures.

Secondly, we focused on the aspect of creating multilingual dataset by machine translation for comics. It is hard to translate comics automatically. Mantra [10] is one of the important technologies which automatically translate Japanese comic images to English comic images based on the pair-images in existing comic databases. It works well for several character words and narration. However, it is still difficult to translate completely

while keeping the meanings of the reference text because Japanese comics have many kinds of onomatopoeias and name suffixes. Unique expression makes understading comics hard in spite of remarkable progress of neural translation. English sentences tends to be longer than Japanese sentences if they are directly translated. Moreover, that is not only in neural translation, but also probably in dialogue breakdown detection.

Thirdly, several studies focus on the aspect of creating multi-modal dataset which consists of natural languages and images. Because of the development of fields in deep learning using big data, we can obtain several types of datasets for photos [11] and clip arts [12]. For research in computer's understanding comics, Manga109 [8], [13] , which was used in this paper, has 109 comics in Japanese. eBDtheque, a dataset Guerin et al. [14] created, has comics in various languages such as: Japanese, English, and French. They contain pictures, texts, annotated places of characters and balloons. However, they were not enough for research because for example texts in the dataset are not aligned in correct order.

On the other hand, Four-scene comic story dataset [15] has various annotations for tasks related to comic contents to analyze them. However, it has less number of data than Manga109.

Thus, in this paper, we used Manga109 and Four-scene comic story dataset, and made a dataset related to comics manually.

B. Models in NLP

Distributed word representation has a long history since 1986 [16]. Recently, it has been an essential technique in NLP.

Word2vec [3] is a well-known model that calculates word vectors. For example, the vectors obtained from word2vec learned the male/female relationship: operation with the vectors "King - Man + Woman" results in a vector very close to "Queen." Methods to train word vectors have been widely applied since word2vec appeared. Word vectors are not only fixed but can be used as initial values of embedding layers; this approach outperforms the embedding layers with random values [17].

In addition to word2vec, models for word representation include Glove [4], fastText [18], and ELMo [19]. Glove is a model that uses the statistics of word occurrences in a corpus. fastText is a model with the phrase representations, position-dependent weighting, and subword information. ELMo [19] is a model to obtain contextualized word vectors. The representations differ from traditional word embeddings: in ELMo, each token is assigned a representation that is a function of the entire input sentence. ELMo is a bidirectional language model, and BERT is based on it.

Moreover, many methods to obtain sentence vectors were developed. For example, paragraph vectors or doc2vec [20] is an unsupervised method that learns document (or paragraph) vector representations. The model is based on word2vec and is simple. Although word2vec learns generic word vectors, the vectors from the model are not so good, because the model does not learn the order of words or context. The model is only trained along the context windows in corpora.

Skip-Thought [5] is an unsupervised model with long short-term memory. According to our hypothesis, it should perform well in solving the problem of dialogue breakdown detection, because it learns the continuity of text.

Additionally, we assume that BERT is suited for dialogue breakdown detection, because BERT achieved state-of-the-art results on many datasets, and it also learns the continuity of text in the task of next-sentence prediction (NSP).

With encoder-decoder models, we can interpret outputs from the encoders as sentence vectors regardless of the training task. Although these models produce high-quality vectors, they are tuned only to their respective task. Thus, we exclude these models from our experiment.

III. Models

A. Skip-Thought

Skip-Thought is an approach to unsupervised learning of a generic, distributed sentence encoder [5].

The model is trained on the continuity of text from books. Specifically, using the input sentence received in the encoder, the decoder reconstructs the previous sentence and the next one. With Skip-Thought, sentences that share semantic and syntactic properties are mapped to similar vectors.

Skip-Thought has a simple vocabulary expansion method to encode words that were not seen during the training. Let $\mathcal{V}_{\mathrm{w2v}}$ denote the word2vec embedding space and let $\mathcal{V}_{\mathrm{rnn}}$ denote the recurrent neural network embedding space. Then, under the assumption that $\mathcal{V}_{\mathrm{w2v}}$ is larger than $\mathcal{V}_{\mathrm{rnn}}$, we can define a mapping function f for these models: $\mathcal{V}_{\mathrm{w2v}} \rightarrow \mathcal{V}_{\mathrm{rnn}}$ parameterized by a matrix $\mathbf{W}$ such that $\boldsymbol{v'} = \boldsymbol{W}\boldsymbol{v}$.

In this experiment, we trained our model on 2.8 GB of Wikipedia dumps in Japanese[1]. However, we used the data only partially: we considered only series of sentences because continuity was important. Moreover, we used sentences with ten words or more.

B. BERT

BERT is a language representation model [1]. It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As the name suggests, BERT includes transformers [21][2].

BERT requires two training steps: pre-training and fine-tuning. In the pre-training step, the model is trained on masked language model (MLM), the unsupervised learning, and NSP, the classification task. On MLM, the sentences with masked words are used as data, although standard conditional language models can only be trained left-to-right or right-to-left. When predicting the next word by using bidirectional information,

[1]https://dumps.wikimedia.org/jawiki/

[2]They are based on attention mechanisms, not recurrence and convolutions. Recurrent networks typically factor computation along the symbol positions, and step-by-step computation is required. That precludes parallelization and requires more training time.

TABLE I
THE CONDITION OF MODEL 1-A

Kernel (SVM)	rbf
C (SVM)	21.4533
Gamma (SVM)	0.001070

TABLE II
THE CONDITION OF MODEL 1-B

Min sample split (RF)	13
Max leaf nodes (RF)	25
Criterion (RF)	entropy

the model predicts the word by indirectly seeing itself. On NSP, the model classifies whether in a pair of input sentences one sentence directly follows the other sentence or not. In the fine-tuning step, we only need to retrain the pre-trained BERT model with an additional layer. This step is easy and, compared to pre-training, relatively inexpensive.

In our experiment, we use PyTorch transformers[3] and the pre-trained model[4]. We train the model on Japanese text with words separated by Juman++ [22] and preprocessed by byte-pair encoding [23].

IV. EXPERIMENTS

A. Experimental Setup

We evaluated five models on dialogue breakdown detection:

1) Use Skip-Thought to get sentence vectors and classify the vectors as the input into:
   a) support vector machine (SVM)
   b) random forest (RF)
   c) multi layer perceptron (MLP)

   They were all models in Scikit-Learn. The hyper-parameters of the models are tuned by Optuna [24].
2) Fine-tune the pre-trained BERT model.

TABLE I, II, III, IV, V shows hyper-parameters of Model 1-a, 1-b, 1-c, 2 and the pre-trained Skip-Thought Model, respectively. The hyper-parameters of the models in Scikit-Learn that are not be written in the table have the default values. When pre-training the Skip-Thought model, we did not use sentences whose length was less than ten words. Note that the pre-trained Skip-Thought model and the pre-trained BERT model have the same vocabulary, but some words were not registered because they did not appear during the training.

We used datasets from the Dialogue Breakdown Detection Challenge (DBDC) 1, 2, 3 [25], and a comics test dataset that included Four-scene comic story dataset and Manga109. We split DBDC datasets in a training dataset and a test dataset. For testing, we used both the test part of the DBDC dataset and the comics test dataset. Importantly, the comics test dataset was not used for training.

[3]https://github.com/huggingface/pytorch-transformers
[4]http://nlp.ist.i.kyoto-u.ac.jp/index.php

TABLE III
THE CONDITION OF MODEL 1-C

Hidden layer sizes (MLP)	329
Solver (MLP)	adam
Activation (MLP)	tanh
Alpha (MLP)	0.033687
Learning rate init (MLP)	8.4266×10^{-4}

TABLE IV
THE CONDITION OF MODEL 2

Hidden size	768
Max length (of inputs)	128
Vocabulary	32006
Transformer blocks	12
Self attention heads	12
Hidden activation function	Gaussian error linear units [26]
Dropout probability	0.1

DBDC 1, 2, 3 included conversations with a computer system, and models detected breakdowns of user's words after the system's words. Thus, inputs contained two lines: the user's words and the system's words. The conversations were annotated by annotators: they assigned labels depending on whether the conversation was broken down or not, and we calculated average scores of annotations. In the experiment, the scores of the label "Not a breakdown," "Possible breakdown," and "Breakdown" are 0, 1, and 2, respectively. The label of the talk is 0 ("Not a breakdown") when the average is < 1, and 1 (breakdown) when the average is ≥ 1.

As mentioned above, we created the comics test dataset that included Four-scene comic story dataset and Manga109. In Manga109, all comics used in the experiments were four-scene comics. They were "OL lunch" and "Koukousei no hitotachi." The models predicted whether the words were next to the character's words in the comics or not. Positives in the dataset were pairs of continuous lines in Four-scene comic story dataset and "OL lunch." Negatives were pairs of a line in the four-scene comic story dataset or "OL lunch" and a line in "Koukousei no hitotachi." We picked up randomly, but they were broken down.

TABLEVI shows statistics for the datasets. The data were used for training the SVM, RF, and MLP models and fine-tuning of the BERT model. Note that inputs of the SVM, RF, and MLP model are made by concatenating the vectors of the two lines.

B. Experimental Results

TABLE VII shows the results of the experiment. In our experiments, Bert outperformed the other models. Our initial hypothesis was that Skip-Thought would performed well, because it would learn the continuity of sentences. However, it did not perform as expected because it was tuned to data for unsupervised learning which were 10 words or more. We think that shorter sentences might make noises. Moreover, we think that fine-tuning the pre-trained Skip-Thought encoder like BERT makes the performance of representations well.

TABLE V
THE CONDITION OF THE PRE-TRAINED SKIP-THOUGHT MODEL

Hidden size	768
Max length (of inputs into Skip-Thought)	64
Vocabulary	29933
the number of data	6850507

TABLE VI
THE NUMBER OF DATA FOR DIALOGUE BREAKDOWN DETECTION. VALIDATION DATA IS 10% OF TRAINING DATA PICKED UP RANDOMLY.

Dataset		unbreakdown	breakdown
DBDC dataset	Training data	2132	2887
	Test data	201	299
comics test dataset	Test data	50	50

Thus we will try to re-learn Skip-Thought and fine-tune it to the task of dialogue breakdown detection.

BERT performed well, but the score was 0.7289 on DBDC and 0.6162 on the comics test dataset. BERT was not made for dialogue breakdown detection, and inputs into it contained only one or two sentences because of the system design. Dialogue breakdown detection has specific challenges. One of those is, for example, requiring acknowledgment of previous talks. Because in this paper our models answered questions relying on one sentence, our models did not solve the problems requiring previous talks.

How about the comics test dataset? Firstly, comics have sentences with subtle differences from wikipedia data and novels. Characters often give responses like 'I see' or 'right' to the first speakers, and even the first speakers just mutter to themselves (and then, the second characters gave responses to the mutter). Here, we consider that if characters give simple responses, this is "Not a breakdown," but the model requires more data to understand that.

Secondly, comics like shorter sentences because characters must be in balloons and a character must be written in the proper size for reading. Fig. 1 and 2 show rates of the length of the first speaks and the responses in the training dataset, the DBDC test dataset and the comics test dataset, respectively. As shown in Fig. 1, the rates of the length of the first speaks in all dataset were similar. On the other hand, as shown in Fig. 2, the rates of short sentences in comics test dataset were high, and actually models could not solve the data.

After all, data in dialogue breakdown detection datasets could not be substituted for comics data, and we need to prepare more comics data or ordinary conversation data similar to talks in comics for future experiments.

V. CONCLUSION

We solved the problem of dialogue breakdown detection with two unsupervised models, Skip-Thought and BERT, and analyzed their performance on the comics data. In this experiment, models were trained on no comic data. According to our results, BERT outperformed Skip-Thought. However, the result show that, without comics data, models did not understand the comics. For this reason, we will prepare more comics data or ordinary conversation data similar to talks in comics and will experiment with the models fine-tuned on these data.

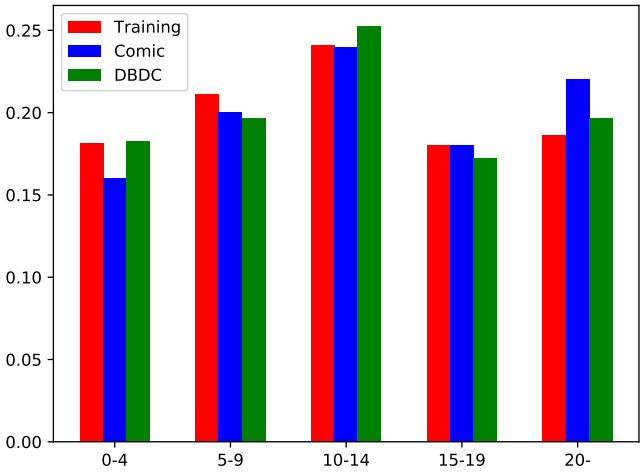


Fig. 1. Rates of the length of the first speaks. Training, Comic and DBDC show the training dataset, comics test dataset and DBDC test dataset, respectively. The all datasets have similar rates of the length.

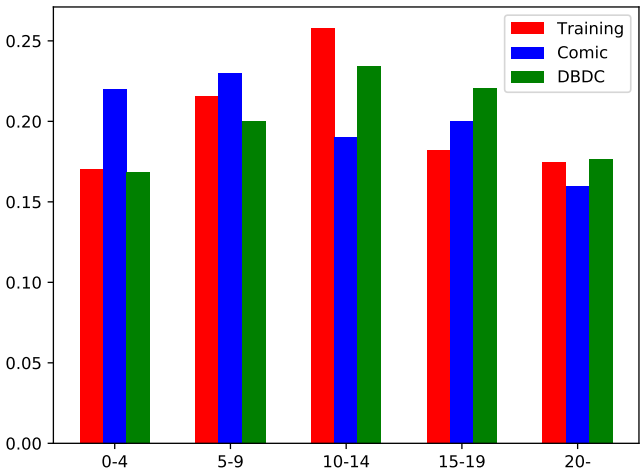


Fig. 2. Rates of the length of the responses. Training, Comic and DBDC show respectively the training dataset, comics test dataset and DBDC test dataset, respectively. Responses in the comic dataset are shorter than in the others.

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition.

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space.

TABLE VII
THE RESULTS OF THE EXPERIMENT

Model Name	Train				
	Accuracy	Precision		Recall	
		unbreakdown	breakdown	unbreakdown	breakdown
Model 1-a	0.7237	0.7361	0.7178	0.474202	0.905193
Model 1-b	0.6895	0.6845	0.6917	0.489681	0.903799
Model 1-c	0.6687	0.7630	0.6479	0.420731	0.859532
Model 2	0.946888	0.972144	0.930129	0.902275	0.980514

Model Name	Test (DBDC)				
	Accuracy	Precision		Recall	
		unbreakdown	breakdown	unbreakdown	breakdown
Model 1-a	0.643286	0.580419	0.668539	0.313432	0.855704
Model 1-b	0.625250	0.549295	0.655462	0.298507	0.832114
Model 1-c	0.635270	0.614457	0.639423	0.278606	0.842281
Model 2	0.724900	0.705128	0.733918	0.547261	0.854118

Model Name	Test (the comic dataset)				
	Accuracy	Precision		Recall	
		unbreakdown	breakdown	unbreakdown	breakdown
Model 1-a	0.43	0.393939	0.447761	0.16	0.74
Model 1-b	0.55	0.575757	0.537313	0.22	0.86
Model 1-c	0.49	0.461538	0.494252	0.14	0.78
Model 2	0.616161	0.703703	0.583333	0.387755	0.84

[4] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2014.

[5] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Skip-thought vectors.

[6] Christophe Rigaud, Nam Le Thanh, Jean-Christophe Burie, Jean-Marc Ogier, Motoi Iwata, Eiki Imazu, and Koichi Kise. Speech balloon and speaker association for comics and manga understanding. In *ICDAR*, pages 351–355. IEEE Computer Society, 2015.

[7] Christophe Rigaud, Srikanta Pal, Jean-Christophe Burie, and Jean-Marc Ogier. Toward speech text recognition for comic books. In *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding*, MANPU '16, pages 8:1–8:6, New York, NY, USA, 2016. ACM.

[8] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017.

[9] Saya Fujino, Naoki Mori, and Keinosuke Matsumoto. Recognizing the order of four-scene comics by evolutionary deep learning. In *Distributed Computing and Artificial Intelligence, 15th International Conference, DCAI 2018, Toledo, Spain, 20-22 June 2018.*, pages 136–144, 2018.

[10] Mantra : Manga traslator. https://mntr.jp/.

[11] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollr, and C. Lawrence Zitnick. Microsoft coco captions: Data collection and evaluation server. *CoRR*, abs/1504.00325, 2015.

[12] Stair captions: Constructing a large-scale japanese image caption dataset. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 417–421, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[13] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017.

[14] Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier, and Arnaud Revel. ebdtheque: a representative database of comics. In *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1145–1149, 2013.

[15] Miki Ueno. Four-scene comic story dataset for softwares on creative process. In *New Trends in Intelligent Software Methodologies, Tools and Techniques - Proceedings of the 17th International Conference SoMeT_18, Granada, Spain, 26-28 September 2018*, pages 48–56, 2018.

[16] G. E. Hinton, J. L. McClelland, and D. E. Rumelhart. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Distributed Representations, pages 77–109. MIT Press, Cambridge, MA, USA, 1986.

[17] Yoon Kim. Convolutional neural networks for sentence classification.

[18] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. Advances in pre-training distributed word representations.

[19] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations.

[20] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need.

[22] Hajime Morita, Daisuke Kawahara, and Sadao Kurohashi. Morphological analysis for unsegmented languages using recurrent neural network language model. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2015.

[23] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units.

[24] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework.

[25] R. Higashinaka, K. Funakoshi, M. Inaba, Y. Tsunomori, T. Takahashi, and N. Kaji. Overview of dialogue breakdown detection challenge 3. In *Proceedings of Dialog System Technology Challenge 6 (DSTC6) Workshop*, 2017.

[26] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus).

Deep Neural Network Pretrained by a Support Vector Machine

Hironori Yamamoto
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
h.yamamoto@ss.cs.osakafu-u.ac.jp

Naoki Mori
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
mori@cs.osakafu-u.ac.jp

***Abstract*—Recently, deep neural networks (DNNs) have shown strong performance in many applications. Some models achieve state of the art in a wide range of fields, such as natural language processing and image processing. Support vector machines (SVMs) have also been a popular approach thanks to their performance. Their criterion selects effective variables in a dataset, and kernel methods help the models extract useful features for prediction. In this paper, we propose a unique method to apply parameters of a pretrained SVM to a DNN. Our proposed method initializes the first layer of a DNN to behave as the pretrained SVM or to extract powerful features from input variables. As a result, the DNN is successfully tuned during the training, because it inherits the advantages of SVMs. To show performance of our proposed method in the experiments, we apply it to classification of a toy dataset and to a sentiment analysis of movie reviews. The results show that the pretrained parameters have a significant effect on the optimization of DNNs.**

***Index Terms*—support vector machine, deep neural network, kernel method**

I. Introduction

Recently, deep neural networks (DNNs) have achieved remarkable performance in many fields, such as natural language processing (NLP) and image processing. To represent the input data, DNNs generate useful features in their hidden units or parameters; thus, some pretrained models can be successfully used for other tasks. Support vector machines (SVMs) [1] [2] [3] are also widely used nowadays. In SVMs, the objective function is designed to find effective variables, and kernel tricks help to model non-linear and linear data. Thanks to these features, SVMs handle input variables effectively.

In this paper, we propose a unique method to transfer parameters of a pretrained SVM classifier into the first layer of a DNN. The resulting model has the advantages not only of a DNN but also of the pretrained SVM classifier. In our experiments, we classify a toy dataset and perform a sentiment analysis with our proposed method.

II. Related Works

Many types of DNNs are pretrained and then reused for other tasks. In image processing, convolutional autoencoders (CAEs) [4] are popular. CAEs are autoencoders that consist of convolutional neural networks [5]. Input features are effectively compressed by training the models to restore inputs from low-dimensional hidden units. As a result, compressed features can be applied to various tasks. Such models are called generative neural networks, because they can also generate new data. Classification models, such as VGG-16 [6] and a residual network (ResNet) [7], are widely used as pretrained models. This is because transferring parameters of models trained on a huge dataset with many general labels improves the results on specific tasks (such as classification of images in a narrow domain). Research on NLP tasks also has the same tendency. Word2Vec [8] and Glove [9] map words into feature vectors that reflect the meaning. They are trained by predicting words from other words appearing in the same contexts. Additionally, some models can handle sentences. Sequence-to-sequence (seq2seq) models generate feature vectors of sentences [10]. They consist of autoencoder's architecture and use recurrent neural networks (RNNs) or the long short-term memory [11] as an encoder and a decoder to handle time dependency. The encoder part provides compressed representations from RNNs' state. Next, the decoder part uses a language model to restore the inputs from the composed features. More recent models [12], [13] prelearn functions of entire sentences, whereas traditional models, as described above, tried to assign good representations to words or sentences. They achieved significant improvements in the representation learning tasks of NLP.

III. Support Vector Machine

Support vector machines are supervised learning models that perform classification and regression. In this paper, we focus on classification tasks. Thus, this section describes SVM classifiers (hereinafter, shortly referred to as SVMs). SVMs solve binary classification tasks by finding a hyperplane to maximize the margin between two classes. The hyperplane is defined by some data points sampled from the input, which are called support vectors. There are two types of definitions of margin: hard margin and soft margin. A hard margin is used for linearly separable data; otherwise, a soft margin is used. In this paper, we use soft margin models.

At first, a dataset D is given as

$$D = \{(\boldsymbol{x}_i, y_i) \,|\boldsymbol{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^{n}, \qquad (1)$$

This work was supported by JSPS KAKENHI Grant, Grant-in-Aid for Scientific Research(B), 19H04184.

where p is a dimension of input variables, and n is the number of data points. A predicted label $\hat{y}$ by an SVM is obtained by

$$\hat{y} = \mathrm{sgn}\left(\sum_{i=1}^{n_\mathrm{s}} y_{s(i)} a_{s(i)} K\left(\boldsymbol{x}_{s(i)}, \boldsymbol{x}\right) + b\right), \tag{2}$$

where n_s is the number of support vectors and $s(i)$ denotes a suffix of the i-th support vector, both $a_{s(i)}$ and b are training parameters, K is a kernel function, and sig is a signum function. Equation (2) indicates that SVMs employ kernel trick to avoid explicit feature mapping, and the training of the model selects a subset of $\{\boldsymbol{x}_i\}_{i=1}^n$ as support vectors to establish an efficient feature mapping.

IV. DNN pretrained by SVM

In our proposed method, we transfer the powerful feature mapping of an SVM to a DNN. This section describes the details of the migration and shows a simple example.

A. Migration of SVM to DNN

At first, an SVM is pretrained for a dataset D, i.e., the support vectors are selected, and parameters a_s and b are optimized. In (2), parameter b can be added to the summation as

$$\hat{y} = \mathrm{sgn}\left(\sum_{i=1}^{n_\mathrm{s}} \left(y_{s(i)} a_{s(i)} K\left(\boldsymbol{x}_{s(i)}, \boldsymbol{x}\right) + \frac{b}{n_\mathrm{s}}\right)\right). \tag{3}$$

Next, (3) is rewritten as

$$\hat{y} = \mathrm{sgn}\left(\sum_{i=1}^{n_\mathrm{s}} h_{s(i)}\right), \tag{4}$$

$$h_{s(i)} = y_{s(i)} a_{s(i)} K\left(\boldsymbol{x}_{s(i)}, \boldsymbol{x}\right) + \frac{b}{n_\mathrm{s}}, \tag{5}$$

where $h_{s(i)}$ can be regarded as a hidden unit. In other words, SVMs produce a hidden layer $\boldsymbol{h}$ as

$$\boldsymbol{h} = \begin{pmatrix} h_{s(1)} \\ h_{s(2)} \\ \vdots \\ h_{s(n_\mathrm{s})} \end{pmatrix}. \tag{6}$$

Our method passes a hidden layer $\boldsymbol{h}$ to conventional neural network layers instead of summing up the elements of $\boldsymbol{h}$. The proposed method has the following expected benefits.

- The DNN can take advantage of the effective kernel functions of SVMs.
- The number of units of the first layer is predefined, because they are optimized in the SVM's training process.

B. Tuning Types

There are three tuning ways of the DNN: fine-tuning, transfer learning, and full tuning. Fine-tuning inherits the pretrained parameters of the SVM and optimizes them in the training step of the DNN. Transfer learning also inherits the parameters and fixes them in training the DNN. Full tuning only reproduces the structure of the SVM: in this case, the DNN has the same structure as ones of fine-tuning and transfer learning, but all the parameters are initialized randomly.

TABLE I
CUSTOM PARAMETERS OF DATA GENERATOR.

name	value
n_samples	50,000
n_features	2
n_redundant	0
class_sep	0.01
random_state	42

C. Example

This part shows a simple example: a three-layered deep neural network pretrained by an SVM. The first layer is pretrained as mentioned above. Thus, first hidden layer $\boldsymbol{h}_1$ is given as

$$\boldsymbol{h}_1 = \begin{pmatrix} h_{s(1)} \\ h_{s(2)} \\ \vdots \\ h_{s(n_\mathrm{s})} \end{pmatrix}. \tag{7}$$

Then, traditional fully connected layers are employed as

$$\boldsymbol{h}_2 = f\left(\boldsymbol{W}_1 \boldsymbol{h}_1 + \boldsymbol{b}_1\right), \tag{8}$$

$$\hat{y} = \sigma\left(\boldsymbol{W}_2 \boldsymbol{h}_2 + \boldsymbol{b}_2\right), \tag{9}$$

where $\boldsymbol{W}_i$, $\boldsymbol{b}_i (i \in \{1, 2\})$ are the weight and bias of a fully connected layer, f is an activation function, and σ is a logistic function. Note that SVMs predict labels as $\{1, -1\}$, but this model produces the probability of a certain class. As a result, this model employs binary cross-entropy loss, while SVMs use a hinge loss.

V. Experiments 1: Toy Dataset

This experiment shows the features of our proposed method using a simple model described in Subsection IV-C and a two-dimensional dummy dataset. Note that, in this experiment and next, we do not consider difference of computational time between our proposed model and other models as comparison, although our model requires more cost than the compared models. We leave reducing computational time for feature work.

A. Dataset

A dummy dataset from SciKit-Learn (version: 0.20.4) [14] is used in this experiment. The dataset is generated by a method named "make_classification," which generates a random classification problem. TABLE I shows the custom parameters of "make_classification" for this experiments. The other parameters use their default variables. Fig. 1 shows samples of the generated dataset. The generated data are sampled from two normal distributions assigned to each label.

We use 80% of the dataset to train the model, and 20% is for checking the model's performance. The former is called the train set, and the latter is called the test set.

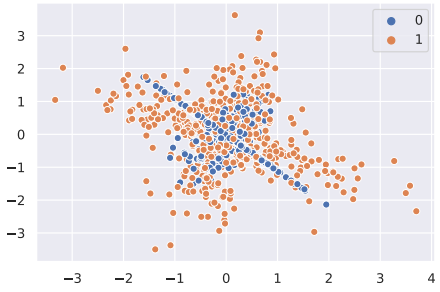


Fig. 1. Samples of the generated dataset

TABLE II
PARAMETERS OF THE SVM.

parameter name	value
not-meta-optimized SVM	
C	1.000
γ	1.000
kernel	rbf
meta-optimized SVM	
C	92.24
γ	8.856
kernel	rbf

TABLE III
PERFORMANCE OF THE SVMS.

metric	0	1	micro avg
not-meta-optimized SVM			
f1-score	0.7744	0.6870	0.7378
precision	0.6779	0.8560	0.7378
recall	0.9029	0.5738	0.7378
meta-optimized SVM			
f1-score	0.7940	0.7553	0.7763
precision	0.7339	0.8367	0.7763
recall	0.8648	0.6884	0.7763

TABLE IV
PARAMETERS OF THE DNNS.

parameter name	value
DNN pretrained by not-meta-optimized SVM	
size of $\boldsymbol{h}_1$	18,250
size of $\boldsymbol{h}_2$	145
activation f	$\tanh$
α (Adam)	8.362×10^{-5}
tune type	fine tuning
DNN pretrained by meta-optimized SVM	
size of $\boldsymbol{h}_1$	18,392
size of $\boldsymbol{h}_2$	94
activation f	$\tanh$
α (Adam)	2.303×10^{-5}
tune type	fine tuning

B. SVM Setup

Before training a DNN, we train two SVMs: one with not-meta-optimized hyperparameters and one with meta-optimized hyperparameters, to analyze effects of the parameters on our proposed method. The parameters of the latter SVM are optimized by Optuna software [15], which optimizes them based on a Bayesian optimization algorithm. TABLE II shows the parameters of the two SVMs.

TABLE III shows the performance of the SVMs for the test set. The table points out that Optuna succeeds in the meta-optimization about hyper parameters.

C. Results and Analysis

In this subsection, we analyze the results of transferring the parameters of the pretrained SVM to a simple DNN described in Subsection IV-C. We use 20% of the train set as a validation set to select the final model. We test the model at the epoch when it obtains the best accuracy on the validation set.

First, two SVMs with the same parameters in Subsection V-B are trained on the train set without the validation set. Then, we train two DNNs based on each SVM. TABLE IV shows the parameters of the DNNs. The parameters of DNNs are optimized by Optuna, except for the number of units of $\boldsymbol{h}_1$, because it equals to the number of the support vectors of the pretrained SVM.

TABLE V shows the performance of the DNNs pretrained by the SVMs. According to our results, the proposed method improves performance of the not-meta-optimized SVM. Thus, our model handles feature mapping of the SVM as a network layer. Additionally, the results show that the proposed model performs as successfully as the meta-optimized SVM. Fig. 2 shows the prediction of the pretrained SVMs and proposed models. As shown in Fig. 2, the DNN based on the meta-optimized SVM have a tendency to overfit the train set. That is, an well-optimized SVM does not always become a good pretrained model for our proposed method.

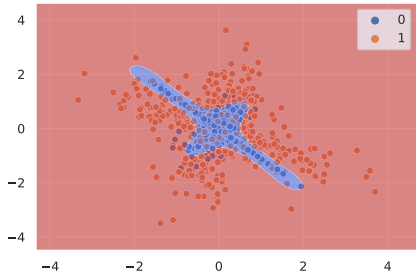
Next, we compare the effects of the three tuning types from Subsection IV-B, using the parameters based on meta-optimized SVM except for the tune type. Fig. 3 shows metrics over training epochs for each tune type. There is a clear difference between the models that use the SVM parameters and not. That is, the model's structure is difficult to train, because it has a large number of units in $\boldsymbol{h}_1$ and kernel functions. However, pretrained parameters make the optimizations of the DNN easier, and the resulting model performs well.

VI. EXPERIMENT 2: SENTIMENTAL ANALYSIS

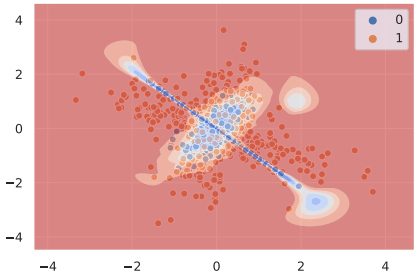
In this experiment, we apply our proposed method to a sentiment analysis dataset [16] as a real-world application, comparing the simple model (Subsection IV-C) with an SVM and a multilayer perceptron (MLP).

A. Preprocessing

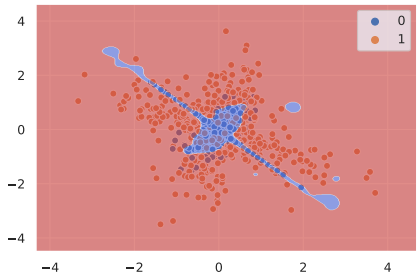
The dataset contains 50,000 reviews, which are separated in the same manner as in Section V. Validation set is only separated for the MLP and the proposed method; in other



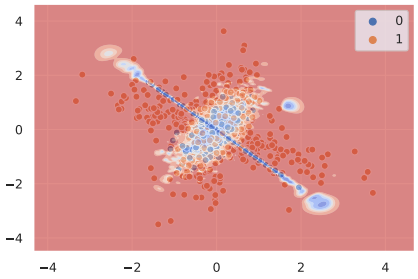
(a) Pretrained SVM with not optimized parameters.



(b) DNN pretrained by the SVM with not optimized parameters.



(c) Pretrained SVM with not optimized parameters.



(d) DNN pretrained by the SVM with not optimized parameters.

Fig. 2. Prediction of the pretrained SVMs and proposed models.

TABLE V
PERFORMANCE OF THE DNNS

metric	0	1	micro avg
DNN pretrained by not-meta-optimized SVM			
f1-score	0.7918	0.7693	0.7811
precision	0.7528	0.8161	0.7811
recall	0.8351	0.7275	0.7811
DNN pretrained by meta-optimized SVM			
f1-score	0.7914	0.7646	0.7788
precision	0.7467	0.8199	0.7788
recall	0.8417	0.7163	0.7788

words, the SVM is trained on the complete training dataset. We convert words in the reviews into word vectors produced by a word2vec pretrained on a part of the Google News dataset[1]. Next, input dataset is constructed by calculating an average of word vectors for each review. Target values are represented by 0 (negative) and 1 (positive).

B. Setup of Models

In this subsection, we describe parameters of the models that are optimized by Optuna. TABLE VI shows the parameters of the SVM, the MLP and the proposed model, which are optimized optuna. The parameters of the SVM, as a comparison, are inherited to the pretrained SVM in our proposed model.

[1]Obtained at https://code.google.com/archive/p/word2vec/

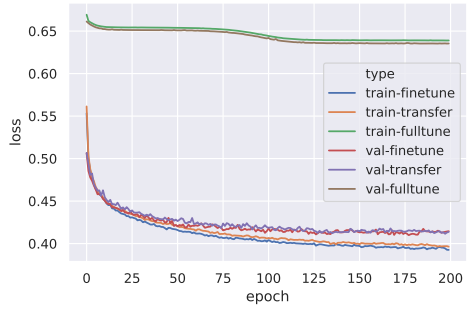
[2]$\mathcal{N}(\mu, \sigma)$ denotes a normal distribution with the mean μ and variance σ^2.

TABLE VI
PARAMETERS OF THE MODELS.

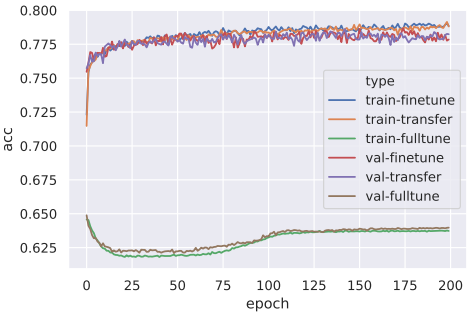
parameter name	value
SVM	
C	3.697
γ	9.927
kernel	rbf
MLP	
size of first units	797
first activation	tanh
size of second units	516
second activation	tanh
α (Adam)	5.993×10^{-4}
proposed model	
size of $\boldsymbol{h}_1$	13,510
size of $\boldsymbol{h}_2$	254
activation f	tanh
α (Adam)	1.508×10^{-6}
tune type	transfer learning
weight initializer	$\mathcal{N}(1, 0.05)^2$

C. Results and Analysis

TABLE VII show performance of the SVM, the MLP and the proposed model. The SVM outperforms other models; thus, our proposed model has not been optimized successfully in the training. In other aspects, it outperforms the MLP; that is, the proposed method handles features that the MLP does not.



(a) Losses.



(b) Accuracies.

Fig. 3. Metrics over training epochs for each tuning type. In the figures, "train-" refers to results for train data and "valid-" for validation data.

TABLE VII
PERFORMANCE OF THE MODELS

metric	negative	positive	micro avg
SVM			
f1-score	0.73108	0.72786	0.7295
precision	0.72678	0.73225	0.7295
recall	0.73544	0.72352	0.7295
MLP			
f1-score	0.68335	0.73515	0.7116
precision	0.75742	0.67957	0.7116
recall	0.62248	0.80064	0.7116
proposed model			
f1-score	0.72472	0.71445	0.7197
precision	0.71192	0.72804	0.7197
recall	0.73800	0.70136	0.7197

VII. Conclusion

In this paper, we proposed a method to transfer parameters of a pretrained SVM to a DNN. According to the experimental results, the method helps the DNN handle the feature mapping of the pretrained SVM, and the pretrained parameters have a significant effect on the optimization of the proposed example. We have some feature works. The following shows our important future work. First, we will conduct research on optimization method for out proposed model to reduce computational cost. Second, we will explore regularization methods and extend our proposed method by combining some pretrained SVMs with different kernels or parameters to obtain a more robust prediction. Finally, we are interested in finding more effective structures of the DNNs, because inheriting all the support vectors makes the first layer huge.

References

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[2] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[3] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm.

[4] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked Convolutional Auto-encoders for Hierarchical Feature Extraction," in *Proceedings of the 21th International Conference on Artificial Neural Networks - Volume Part I*, ser. ICANN'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 52–59.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1556, 2014.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[8] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 3111–3119.

[9] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global Vectors for Word Representation." in *EMNLP*, vol. 14, 2014, pp. 1532–1543.

[10] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3104–3112.

[11] J. Schmidhuber and S. Hochreiter, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep Contextualized Word Representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, Jun. 2018, pp. 2227–2237.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. ACM, 2019, pp. 2623–2631.

[16] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Jun. 2011, pp. 142–150.

The novel index of the similarity between hand-drawn sketches for machine learning

1st Ryosuke Fujii
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
fujii@ss.cs.osakafu-u.ac.jp

2nd Naoki Mori
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
mori@cs.osakafu-u.ac.jp

3rd Makoto Okada
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
okada@cs.osakafu-u.ac.jp

***Abstract*—We have dealt with sketches hand-drawn by human hands. They are creations including human emotions and sensibility. To handle sketches on a computer, it is necessary to have an index that quantitatively evaluates the stroke-order. We have already proposed a sketch similarity index for pen movement by cosine similarity. However, we also found that human cognition strongly depends on the shape of a sketch when they understand what is drawn. For this reason, we propose a new similarity index, which is the old index added to a shape similarity index. As a result, the similarity index composed of cosine similarity and Structural Similarity (SSIM) has a strong positive correlation with the human evaluation of similarity. Then, the correlation coefficient is 0.5667. Therefore, we define an objective evaluation model for the stroke-order similarity that incorporates not only the pen movement of a sketch but also the shape of one, and we verify its effectiveness by our experiments.**

***Index Terms*—sketch-rnn, similarity, stroke-order, Kansei**

I. Introduction

Artificial intelligence (AI) has recently attracted significant attention owing to the development of deep learning. While AI exhibits performance that surpasses human in the field of simple pattern recognition, it is still difficult to apply AI to fields related to human emotions and sensitivities, especially to creative works. Among the creations, sketches are highly important because they are expressions that can be broadly empathic, regardless of age, nationality, or culture.

To analyze sensibilities by sketches, we proposed Creative Animating Sketchbook with sketch-rnn (CASOOK-SR). CASOOK-SR has 2 systems. First, Kansei on still images is analyzed by Creative Animating Sketchbook (CASOOK) [1]. Second, features on hand-drawn lines are given by sketch-rnn [2] [3]. The information about hand-drawn lines contains as many feelings as still images because the stroke-order decision process is subconscious. So, we introduced sketch-rnn into CASOOK to extract sensibility on hand-drawn sketches. Also, in order to apply CASOOK-SR to real problems, we proposed the Visual Search Automotive Production System (VSAPS) [4]. The system generates search pictures automatically.

Furthermore, sketches are considered to be time-series data composed of a combination of partial parts like natural language. For these reasons, sketches have been regarded as one of the important AI issues in the field of sensibilities [5]. Nevertheless, sketches are different from natural language in that there are no distinct partial components such as words. Therefore, we need an index to measure the similarity between sketches in order to extract and evaluate sketch parts, strokes, with similar shapes.

However, comparing the similarity between stroke-orders was one of the difficult tasks. In order to solve this problem, we used the similarity index composed of the time-series average of cosine similarity [6]. While the comparison index succeeded in evaluating pen movement, it turned out that human cognition strongly depends on the hand-drawn shapes. Therefore, we define a similar degree of not only pen movement but also hand-drawn shapes as a new objective evaluation model for the similarity between stroke-orders and confirm its effectiveness by our experiments.

II. Related studies

A. Sketch-rnn

Sketch-rnn is a model that uses deep learning to complement and infer a user's sketches by training them. The input is the stroke-order data including the time series, and the latent vector is the intermediate output. Also, the final output is the stroke-order vector making it possible to generate images. Unlike a convolutional neural network (CNN) [7], sketch-rnn can reproduce the process of sketch drawing. In sketch-rnn, strokes are to a sketch in hand-drawn works what words are to a sentence in natural language.

Sketch-rnn is a sequence-to-sequence variational autoencoder (VAE) [8]. Fig. 1 shows the model of sketch-rnn. The input and output of sketch-rnn are stroke-orders data composed of a five-dimensional vector $(\Delta x, \Delta y p_1, p_2, p_3)$, where Δx, Δy represent the moving distance of the pen from the preceding point. In addition, p_1, p_2, and p_3 represent the state in which the pen is in contact with the paper, away from the paper, and the drawing is finished.

1) The Data of Stroke-Order: The stroke-order data converted from hand-drawn sketches are stored in a matrix. The line direction of the stroke-order data is kept along with the time series, the information about the state of the pen from the beginning to the end of the drawing of the hand-drawn sketch. The column direction of the stroke-order data holds the state of the pen at a certain time t, and the relative values of the x coordinates, the relative values of the y coordinates,

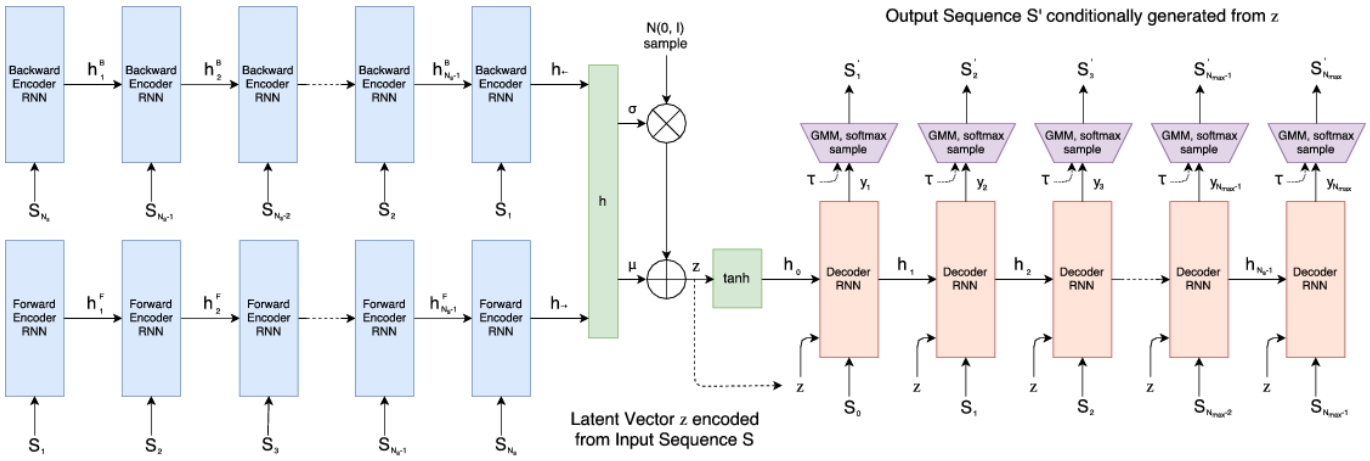


Fig. 1. The Model of Sketch-rnn

and the logical values of the line drawing start and drawing end respectively. However, the origin of x and y coordinates means the start of drawing sketches.

B. CASSOK

CASOOK provides a sketch drawn by users, such as children, using motion. CASOOK analyzes sensibility information from a given sketch and generates motion suitable for the sensibility.

C. CASOOK-SR

CASOOK-SR is our proposed system, which provides an interface for processing the sensibility features in sketches by analyzing the stroke-order and pixel information. The original, CASOOK system, can obtain the features from only the pixel information. To process the stroke-order information, CASOOK-SR adopts sketch-rnn which is a strong model in terms of stroke-order.

D. VSAPS

We propose the use of VSAPS, which is a core module of CASOOK-SR, for application to a real problem, namely, a visual search picture puzzle, which is a game requiring attention that involves an active scan of a visual environment for a particular object placed among other objects. In this study, the user has to recognize both a still image and the stroke-order. To realize the game, VSAPS uses 2 features. First, by introducing sketch-rnn, VSAPS can show users a visual search puzzle with not only the final image but also the stroke-order motions, and it is this stroke-order motion that makes VSAPS an original concept. Second, VSAPS can create visual searches based on sketches drawn by users. VSAPS provides questions and chooses objects with the users as soon as their sketches. Consequently, as the key to an interactive system, no 2 questions are the same. These 2 points enable VSAPS to extract sketch features immediately. Fig. **??** shows an example of a play screen in VSAPS. On the left side of the figure, VSAPS shows a sketch drawn by a user. On the right side of the figure, VSAPS shows some choices for a visual

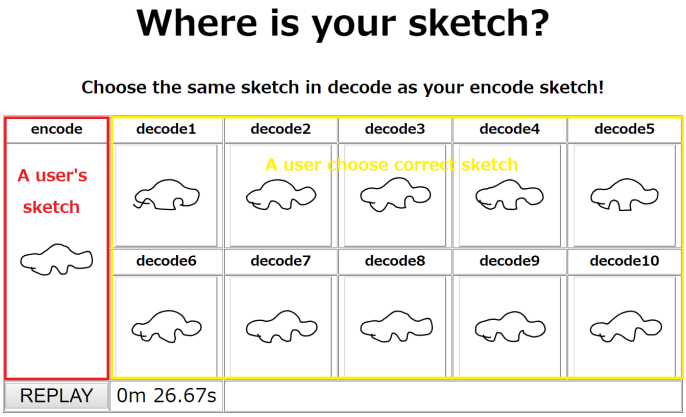


Fig. 2. Screen of VSAPS

search picture puzzle. Players need to choose the choices, are composed of motion or images. The purpose of the game is to choose the correct answer to the choices given.

III. Propose

Equation (1) shows the definitional formula of the similar index about hand-drawn sketches. Our purpose is to replace the human evaluation score with the evaluation score given (1).

$$S = (1 - w)S_{\mathrm{v}} + wS_{\mathrm{p}} \tag{1}$$

Equation (1) has 2 functions, S_{v} and S_{p}.S_{v} evaluates the similarity with pen movement in a sketch, and S_{p} evaluates the similarity with the shape of a sketch. However, $0 \leq S_{\mathrm{v}} \leq 1$, $0 \leq S_{\mathrm{p}} \leq 1$, and $0 \leq w \leq 1$ are satisfied in (1). Furthermore, S_{v} and S_{p} indicate that the 2 sketches are more similar as they approach 0, and that the 2 sketches differ as they approach 1. w is a hyper-parameter related to the similarity of the pen movement and the similarity weight of the image shape. The input data of both S_{v} and S_{p} is the same 2 sketches. However, the input format of S_{v} is stroke-order data, and the input

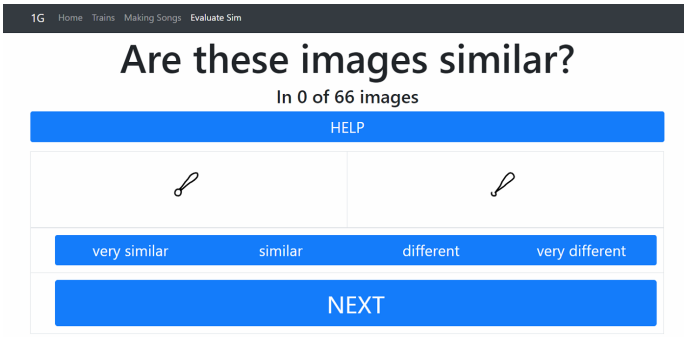


Fig. 3. The Screen in Our Questionnaire

format of S_{p} is a raster image. Therefore, (1) has the following 2 points. The smaller the value of (1), the lower the similarity between 2 different hand-drawn sketches. On the other hand, the larger the value of (1), the higher the similarity between 2 different hand-drawn sketches.

IV. Experiment

A. Purpose

The purpose of this experiment is to create a quantitative evaluation index of similarity in the drawing order. In previous studies, we evaluated pen movement by using cosine similarity. However, it was found that the similarity evaluated by humans strongly depends on the shape [3]. For this reason, in this experiment, not only the movement of the pen but also the similarity evaluation index that incorporates the evaluation of the shape will be verified.

If this can be achieved, we will be able to search for sketches without learning. Accordingly, it is expected that the computer's understanding of human creation can be further deepened. We strongly believe that methods that can be decomposed into meaningful parts in image analysis will be very unique and innovative.

B. Method

The effectiveness of our proposed evaluation index for the similarity of hand-drawn sketches defined by (1) is confirmed by a questionnaire experiment. 12 sketches are prepared, 3 for each of the 4 image class. In the experiment, 9 people are evaluated using ${}_{12}\mathrm{C}_2$ patterns, that is, combinations of 66 patterns. We confirm the validity of the accuracy of (1) by the Pearson correlation coefficient [9]. It is used between the evaluation of (1) and human evaluation. To realize the validation, 9 collaborators answer a questionnaire about the similarity between 2 sketches.

1) Format: In the questionnaire experiment, 9 collaborators are evaluated in four stages of "very similar", "similar", "different" and "very different" about the similarity of stroke-order for 2 different sketches. When the questionnaire results are analyzed, "very similar" is converted 4, "similar" is converted 3, "different" is converted 2, and "very different" is converted 1. Fig. 3 shows actually experimental screen.

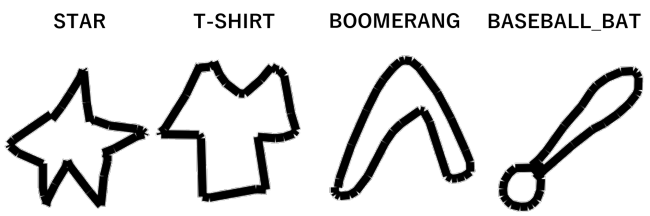


Fig. 4. The Chosen Samples

2) Dataset: We used 4 image class from QuickDraw! [10]; star, t-shirt, baseball_bat and boomerang. They are stroke-order data sets published by Google. These all have the data structure of IV-B2. In addition, these stroke-orders are restored to images, and stroke-order data and raster images are paired. Fig. 4 shows an example of a sketch used in the experiment.

Next, the reason for selecting each image class is described. The star and t-shirts were chosen because their shapes were very different, but the pen movements were very similar. boomerang was chosen because of its rich pattern of sketch orientations. baseball_bat was chosen because it is simple but has few patterns to draw. By using these features, the validity of the proposed (1) is confirmed.

C. Validation

This session describes how to verify the validity of (1). There are 4 processes in total. First, we prepare some functions of S_{v} and S_{p} in (1). Next, we will annotate sketches that are similar to humans by conducting a questionnaire experiment. Then, the hyper-parameters are tuned so that the correlation coefficient between the evaluation of similarity by humans and the evaluation of similarity by (1) is maximized. Finally, using the obtained hyper-parameter and similarity functions, we compare the human evaluation with (1) for the similarity evaluation of the unknown sketches and validate (1). This is checked to see if the obtained hyper-parameters have generalization performance or are dependent on annotate data.

1) Functions: We defined three candidates for S_{v} in (1). The first candidate is shown in (2). Equation (2) was proposed in our research. It was successful for the similarity between 2 sketches in terms of only pen movement. Although (2) only supported the movement of the pen, it did not contribute greatly to the shape evaluated by humans. So, we create functions such as from (3) to (6).

Equation (3) and (4) focus on the difference in pen movement at time t. Equation (4) gives the average value of cosine similarity to give weight at the beginning of the drawing. Equation (5) gives the average cosine similarity of the vector connecting the drawing start position and the time t position. Equation (6) is prepared to check whether there is an element similar to the subset in the drawing order.

$$S_{1\mathrm{v}} = \frac{1}{2}\{1 + \frac{1}{T}\sum_{t=1}^{T}\frac{(\sum_{i=1}^{t}\boldsymbol{U}_i)\cdot(\sum_{i=1}^{t}\boldsymbol{V}_i)}{(||\sum_{i=1}^{t}\boldsymbol{U}_i||)(||\sum_{i=1}^{t}\boldsymbol{V}_i||)}\} \quad (2)$$

$$S_{2\mathrm{v}} = \frac{1}{2}\{1 + \frac{1}{T}\sum_{t=1}^{T}\frac{\boldsymbol{U}_t \cdot \boldsymbol{V}_t}{||\boldsymbol{U}_t||||\boldsymbol{V}_t||}\} \quad (3)$$

$$S_{3\mathrm{v}} = \frac{1}{2}\{1 + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\sum_{i=1}^{t}\frac{\boldsymbol{U}_i \cdot \boldsymbol{V}_i}{||\boldsymbol{U}_i||||\boldsymbol{V}_i||}\} \quad (4)$$

$$S_{4\mathrm{v}} = \frac{1}{2}\{1 + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\sum_{s=1}^{t}\frac{(\sum_{i=1}^{s}\boldsymbol{U}_i)\cdot(\sum_{i=1}^{s}\boldsymbol{V}_i)}{(||\sum_{i=1}^{s}\boldsymbol{U}_i||)(||\sum_{i=1}^{s}\boldsymbol{V}_i||)}\} \quad (5)$$

$$S_{5\mathrm{v}} = \frac{1}{2}\{1 + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\sum_{s=1}^{t}\frac{1}{t-s+1}\sum_{i=s}^{t}\frac{\boldsymbol{U}_i \cdot \boldsymbol{V}_i}{||\boldsymbol{U}_i||||\boldsymbol{V}_i||}\} \quad (6)$$

U and V in (2) to (6) are matrices containing the stroke-order of a sketch in the form of IV-B2. T has an arbitrary time and satisfies $0 < T \leq \min(l_U, l_V)$. Here, l_U indicates the time length of the sketch stroke-order matrix U, and l_V indicates the time length of the sketch stroke-order matrix V. All of (2) to (6) satisfy $0 \leq S_{\mathrm{v}} \leq 1$. The closer S_{v} is to 0, the more different it is in pen movement. On the contrary, the closer S_{v} is to 1, the more similar it is in pen movement.

We defined 2 candidates for S_{p} in (1). The first candidate is shown in (7). The second candidate is shown in (8).

$$S_{1\mathrm{p}} = 1 - \frac{1}{L^2}\frac{1}{n^2}\sum_{i=0}^{n}\sum_{j=0}^{n}(A(i,j) - B(i,j))^2 \quad (7)$$

$$S_{2\mathrm{p}} = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{({\mu_A}^2{\mu_B}^2 + c_1)({\sigma_A}^2{\sigma_B}^2 + c_2)} \quad (8)$$

In (7) and (8), A and B are made in 2 steps. First, U and V are converted to gray-scale raster images. Second, A and B are obtained on the first step products through a Gaussian filter [11]. All images have the same vertical and horizontal size of n. L indicates the dynamic range of the pixel value, and $L = 256$ is used in the experiment.

In (7), Mean Square Error (MSE) [12] is given as the similarity of the shape. However, since the minimum value is 0 and the maximum value is L^2, normalization is added so that the minimum value is 0 and the maximum value is 1. In addition, adjustments have been made so that the similarity is highest when the maximum value is 1.

In (8), Structural Similarity (SSIM) [13] is given as the similarity of the shape.

In (8), μ_A, μ_B is the average of the pixel values of images A and B. Also, $sigma_A$ and $sigma_B$ represent the variance of the pixel values of images A and B. Additionally, σ_{AB} represents the correlation of the pixel values of images A and B. In (8), c_1 and c_2 are $(k_1 L)^2$ and $(k_2 L)^2$. k_1 and k_2 use the recommended values of 0.01 and 0.03.

TABLE I
THE RANGE OF EACH HYPER-PARAMETER

Hyper-parameter	Kind	Min	Max
w	Uniform	0.0	1.0
σ	Uniform	0.0	10.0

TABLE II
THE BEST HYPER-PRAMETERS

Functions	w	σ	Best Correlation
$S_{1\mathrm{v}}$, $S_{1\mathrm{p}}$	0.8643	0.4540	0.7138
$S_{2\mathrm{v}}$, $S_{1\mathrm{p}}$	0.8380	0.7202	0.7044
$S_{3\mathrm{v}}$, $S_{1\mathrm{p}}$	0.8384	0.7142	0.7336
$S_{4\mathrm{v}}$, $S_{1\mathrm{p}}$	0.8588	0.5140	0.7431
$S_{5\mathrm{v}}$, $S_{1\mathrm{p}}$	0.7823	0.7076	0.6976
$S_{1\mathrm{v}}$, $S_{2\mathrm{p}}$	0.6807	1.0734	0.7168
$S_{2\mathrm{v}}$, $S_{2\mathrm{p}}$	0.6468	0.9477	0.6971
$S_{3\mathrm{v}}$, $S_{2\mathrm{p}}$	0.6258	1.2552	0.7431
$S_{4\mathrm{v}}$, $S_{2\mathrm{p}}$	0.6449	1.3082	0.7566
$S_{5\mathrm{v}}$, $S_{2\mathrm{p}}$	0.5684	1.1804	0.6915

2) Hyper Parameter: In (1), w is set as a hyper-parameter. In addition, it was found that S_{p} is 1.0 due to the most white in the grayscale raster image just converted from the stroke-order data in the preliminary experiment. For this reason, it is added as a preprocessing to apply a Gaussian filter to a gray-scale raster image to S_{p}. In this case, σ, which is one of the parameters of the Gaussian distribution [14], is used as the hyper-parameter in (1).

Hyper-parameters w and σ are optimized by optuna [15]. The correlation coefficient between human similarity evaluation and similarity evaluation by (1) is maximized. The values are determined as a best-tuned hyper-parameters. TABLE I shows the range of parameters used in optimization. The optimal values are obtained by updating 1,000 times each.

V. RESULT AND DISCUSS

A. Best Hyper Parameter

TABLE II shows hyper-parameters that maximize the correlation coefficient between similarity by humans and similarity using (1). TABLE II gives that (1) combined (5) and (8) and assigned $w = 0.6649$ and $\sigma = 1.3082$ is the strongest correlation with human evaluations. At the same time, the correlation between the similarity evaluations by humans and by (1) is 0.7566, which is a very strong positive correlation.

First, let us consider σ. Fig. 5 shows the images before and after Gaussian filter processing under the condition of the correlation coefficient 0.7566. The input image in S_{p} is slightly blurred by the Gaussian filter as Fig. 5. This relaxation is controlled by σ, and it is found that 1.00 to 1.30 is effective in combination with (8). Also, it is found that 0.45 to 0.71 is effective when used in combination with (7).

Next, let us consider w. TABLE III shows the mean and the standard division about the similarity of the pen movement and the shape under the condition that the correlation coefficient is 0.7566. It is a fact that similarity evaluation by human depends on the shape of a sketch because the specific gravity of pen movement and shape between 2 sketches in best 1 are a ratio

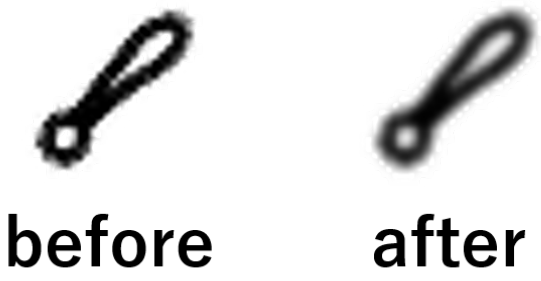


Fig. 5. The Comparison of Blurred and Blurring image

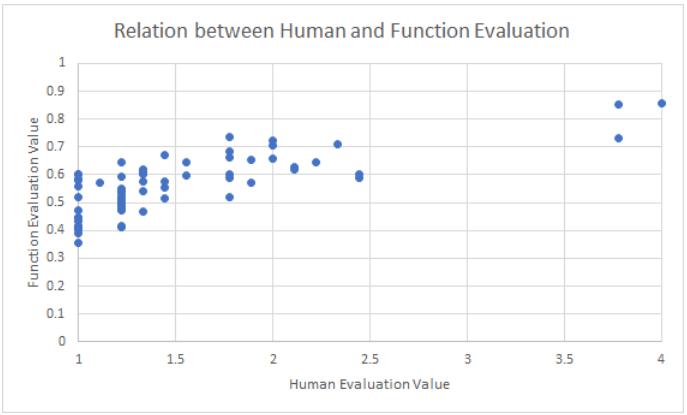


Fig. 6. The Scatter between Human and Function Evaluation

TABLE III
THE MEAN AND STANDARD DIVISION OF EVALUATION

	S_{4v}	$(1-w)S_{4v}$	S_{2p}	wS_{2p}
mean	0.5419	0.1924	0.5780	0.7375
standard division	0.2203	0.0783	0.1043	0.0673

TABLE IV
THE CORRELATION TRANSITION IN w

w	0.0	0.1	0.2	0.3	0.4	0.5
S	0.5782	0.6221	0.6681	0.7116	0.7449	0.7558
w	0.6	0.7	0.8	0.9	1.0	
S	0.7424	0.6987	0.6355	0.5638	0.4954	

of 2 to 3. However, since the standard division of wS_{2p} is small and the average value is large, wS_{2p} is considered to give a rough similarity between sketches. On the other hand, since $(1-w)S_{4v}$ has a large standard division and a small average, $(1-w)S_{4v}$ is similar to sketches. It is thought that a fine evaluation is given regarding a similar degree.

Fig. 6 shows a scatter when the correlation between the human similarity evaluation and the value of (1) is 0.7566. We can see a strong positive correlation in Fig. 6.

Further, we investigate the transition of the correlation coefficient when hyper-parameters except for w were fixed under the condition of the correlation coefficient of 0.7566. TABLE IV shows the transition when w is changed by 0.1. Therefore, the transition is expected to have a convex curve and the maximum value of the transition is 0.6449.

B. Searchability

To confirm the index value obtained by (1), we check the top three sketches and the bottom three ones compared with a target one. The target sketch is the one used in our experiment. On the other hand, the sketches of the candidate for search results are 1600 sketches excluding 12 used in our experiment.

First, we check the search result in 4 kinds of images; star, t-shirt, baseball_bat and boomerang. The number at the bottom of each figure is a value of (1) combined (5) and (8) and assigned $w = 0.6649$ and $\sigma = 1.3082$. The index giving the values has been optimized for each parameter. It is confirmed that the evaluation of both the shape and the stroke-order between sketches are valid. Therefore, it is checked that (1) with optimized the hyper-parameters is highly valid even when (1) evaluates the similarity between an unknown sketch and a known sketch. This result indicates that hyper-parameters have generalization performance.

Next, we check the search result in the same kind of image, boomerang. The first movement is 1, and the final movement is 8. The top three are shown in Fig. 9. The value in (1) is 0.9050 in this case. The bottom three are shown in Fig. 10. The value in (1) is 0.6102 in this case.In each figure, the number is drawing order. The start movement is 1, and the final movement is 8. As we can see from Fig. 9, 2 sketches are almost the same in terms of pen movement and shape. On the contrary, As we can see from Fig. 10, 2 sketches are different in terms of pen movement and direction. The left sketch is clockwise, but the right sketch is counterclockwise. Further, the left sketch is convex upward, but the right sketch is convex to the left. Therefore, if the pen movement and direction in 2 sketches are a little different, the value given (1) will be smaller.

Accordingly, it is valid that searchability with (1) can provide a similar evaluation of hand-drawn sketches as if a human does.

VI. CONCLUSION

This research has dealt with sketches drawn by human hands as a creation that includes human emotions and sensibilities. In order to handle sketches on a computer, an index to quantitatively evaluate the stroke-order was necessary. In previous studies, we proposed a sketch similarity index for pen movement using (2). At the same time, it was found that human cognition strongly depends on the shape drawn. For this reason, we propose (1) that incorporates a shape similarity index, in this paper. As a result, it can be confirmed that (1) composed of (5) and (8) has a strong positive correlation with the human similarity evaluation. Furthermore, the effectiveness of the (1) expression is confirmed by comparing the human evaluation with (1) for similarity evaluation of

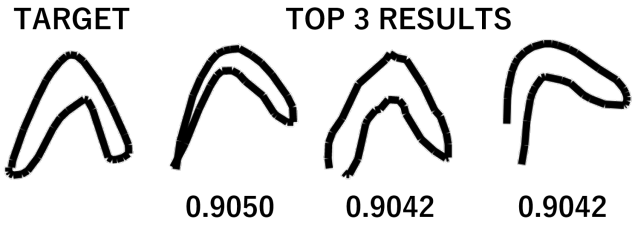


Fig. 7. The Higher 3 Search Result

Fig. 8. The Lower 3 Search Result

unknown sketches. We define the similar degree of not only pen movement but also hand-drawn shapes as a new objective evaluation model for the similarity between stroke-orders, and the effectiveness of the model can be verified by our experiments.

The problem remains in improving the searchability confirmed in the V-B section. We believe that sketch in creative works is to partial strokes in a sketch what a sentence in natural language is to some words in the sentence. So, the mid-term goal of this study is to find out which parts of the stroke are similar between sketches. If that is realized, our proposed index will be able to help AI with generating hand-drawn sketches. Then, the index improves the analysis of human emotions and sensibilities, which are one of the weakest fields of AI, with sketches.

This work was supported by JSPS KAKENHI Grant from Grant-in-Aid for Scientific Research (C), grant number 26330282. We would like to thank Editage (www.editage.jp) for English language editing.

References

[1] Miki Ueno, Kiyohito Fukuda, Akihiko Yasui, Naoki Mori, and Keinosuke Matsumoto. Casook: Creative animating sketchbook. In *Distributed Computing and Artificial Intelligence, 12th International Conference*, pages 175–182, Cham, 2015. Springer International Publishing.

[2] David Ha and Douglas Eck. A neural representation of sketch drawings. In *ICLR 2018*, 2018. 2018.

[3] Ryoske FUJII, Naoki MORI, and Makoto OKADA. A study on evaluation model on stroke order similarity. *Proceedings of the Annual Conference of JSAI*, JSAI2019:2M5J1004–2M5J1004, 2019.

[4] Ryoske Fujii and Naoki Mori. Creative animating sketchbook with sketch-rnn for stroke based visual search picture puzzle. *Proceedings of the Twenty-fourth International Symposium on Artificial Life and Robotics*, AROB 24th 2019:524–527, 2019.

[5] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10, 2012.

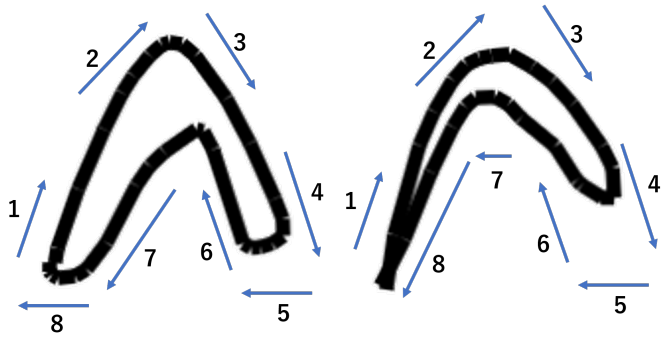


Fig. 9. The Sketch Drawn in the Same Direction

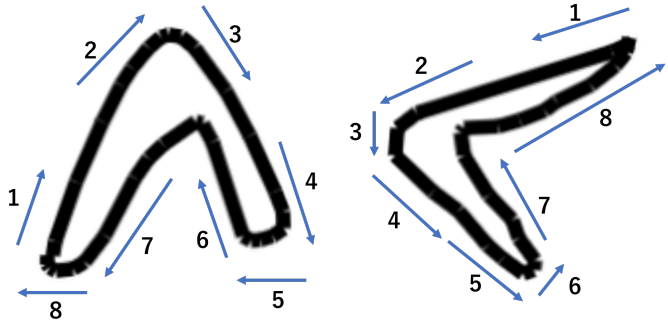


Fig. 10. The Sketch Drawn in the Counter Direction

[6] Hieu V. Nguyen and Li Bai. Cosine similarity metric learning for face verification. In *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part II*, ACCV'10, pages 709–720, Berlin, Heidelberg, 2011. Springer-Verlag.

[7] Patrice Y Simard, David Steinkraus, and John C Platt. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962, 2003.

[8] Yuchen Pu, Zhe Gan, Ricardo Henao, Chunyuan Li, Shaobo Han, and Lawrence Carin. Vae learning via stein variational gradient descent. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4236–4245. Curran Associates, Inc., 2017.

[9] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. *Pearson Correlation Coefficient*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[10] Salman Cheema, Sumit Gulwani, and Joseph LaViola. Quickdraw: Improving drawing experience for geometric diagrams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1037–1064, New York, NY, USA, 2012. ACM.

[11] H.J. Blinchikoff and A.I. Zverev. *Filtering in the Time and Frequency Domains*. Classic series. Institution of Engineering and Technology, 2001.

[12] O. Elbadawy, M. R. El-Sakka, and M. S. Kamel. An information theoretic image-quality measure. In *Conference Proceedings. IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No.98TH8341)*, volume 1, pages 169–172 vol.1, May 1998.

[13] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Trans. Img. Proc.*, 13(4):600–612, April 2004.

[14] George Marsaglia. Evaluating the normal distribution. *Journal of Statistical Software*, 11(5), 2004.

[15] James S. Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 2546–2554. Curran Associates, Inc., 2011.

Optical-based Limit Order Book Modelling using Deep Neural Networks

Pak Laowatanachai
Dept. of Control System and Instrumentation Engineering
King Mongkut's Univ. of Tech.
Thonburi, Bangkok, Thailand
paklaowatanachai@gmail.com

Poj Tangamchit
Dept. of Control System and Instrumentation Engineering
King Mongkut's Univ. of Tech.
Thonburi, Bangkok, Thailand
poj.tan@kmutt.ac.th

Abstract

We used a deep neural network to model an order book's behavior in a stock market using its snapshots as an input. The snapshots were taken from a stock market application in time series format. Google's Tesseract OCR was used to extract price data from these snapshots. A long short-term memory (LSTM) neural network was used to learn the price behaviors in order to predict its future trends, i.e. up, down, or neutral. The result showed that the system achieved an accuracy of 58.98% despite the noise from the OCR and the sampling effect of the snapshots.

Keywords—Predictive models, Time series analysis, Deep Neural networks, Stock markets

I. INTRODUCTION

In technical analysis of stock markets, people make a decision on buying/selling securities based on the past price data. There are roughly two levels of data: level 1 and level 2. Level 1 data includes matched volume and matched prices during a specific time interval in the form of open, high, low, and close prices. Level 2 data adds an addition of order placements by buyers and sellers, which is in form of limited order books (LOBs). Level 2 data is richer in terms of information, but it has access restrictions. Stock markets are regulated in order to protect buyers and sellers. Regulators from different stock markets have different policy on the access of this data. In the stock exchange of Thailand (SET), buyers and sellers can look at an order book via an application. However, the regulator does not provide a function to save data, which is changing rapidly. Investors who want to use this data must watch the application and make their decision accordingly. Some investors have very fatigue eyes because they have to continuously monitor the order book for a long period of time.

Our work wants to implement a system that optically look at the order book as if it is human's eyes, try to extract data, and model the behavior of that order book with a deep neural network. Our system takes snapshots of the application every one second, passes images to Google's Tesseract OCR [1] to extract prices and volumes, then feeds data into a long short-term memory (LSTM) neural network to predict the trend of future prices. Our contribution is to find the effectiveness of a neural network that we can use to model the order book given two obstacles. The first obstacle is the noises from the optical character recognition (OCR) part, which mainly come from flashing numbers in the order book. The second obstacle is the long sampling period of taking snapshots of the order book, which cannot be short due to the delay of application data over the internet.

II. LITERATURE REVIEW

There are many research works in the field of machine learning or deep learning that try to model a stock market. The works can be classified into two types based on two main categories of data used, which are candlestick data and order book data. In candlestick data, one candlestick represents a conclusion of one day data that shows the gap of the prices in that day. Inside it contains open, high, low, and close prices (OHLC). There might be other information tagging along, such as matched volumes. The other type of data is a limit order book which is a list of unexecuted buy/sell orders waiting to be matched from the other side. Order book data contains more information than candlestick data, but its access is usually restricted due to market regulations. Since most research that uses an order book used transactional data that can be accessed by the stock market owner, it is impossible for regular traders to acquire it. On the other hand, traders will see the real data at that time how many people order the stock volume and price at each level which represents the state of the market. But it was not represented every detail in stock markets like the level of the market owner that able to see order cancellation it can say the level of normal users is a level of state of market that still contains a lot of filters in many aspects and only leave the data of bids and offers with the pair of their own volume. This makes a lot of work that revolve around order book are based on level 1 stock data that normal trader cannot have access to and make it difficult for normal trader to make any advantage of it.

A. Candlestick data

There are many research works in candlestick such as the works of Motlagh M. T., & Khaloozadeh H. [2], and Gao T. et al. [3]. They proposed several models and made a comparison with neural network models and statistical models. Most of the results showed that the proposed model performed a little better than neural networks and far better than the statistical models (around 5% at least). There is also notable mention like the work of Lee C., & Soo V [4] and Vargas, M. R. et al. [5]. They used

news data as another input to be fed into neural network models. They performed better than using normal stock data in all aspects, but there was a limitation in news data filter just as in the work of Lee C., & Soo V. They needed to randomly pick only five news topics that still required a proper choice in order to obtain a good result.

B. Order book data

Most of order book data usually relies on historical data dated back many years in the past, because the market regulator do not want to release an updated data due to many reasons. Traders' intentions have to be kept secretly to protect themselves. For example, the work of Tsantekidis, A. et al. [6] [7] was published in the year of 2017, but the data were a long past data of year 2010. There was also work of Doering J. et al. [[8]], who used both the state and event flow of data to combine as an input and achieved a better result. Both of the mentioned works provided predicted market trends. Calvez A. L., & Cliff D. [9]tried to replicate other trading algorithms with neural network models and showed a close result. They could improve the models to better degree with lesser gaps of trading prices than the algorithms that they tried to replicate.

Both of candlesticks and order books are usually used in different timeframe. Candlesticks are usually used with in a sampling time of one day, while order books are a lot shorter. We used a sampling within one second interval, and selected neural network models instead of the others. LSTM neural networks performed better in term of accuracy and processing time especially in the work of Tsantekidis, A. et al., which showed that LSTM performed better than CNN in the same setting.

III. System Overview

Our research tried to understand traders' behaviors from their actions on a limit order book. One of the main problems of using limit order books is that data is kept in various formats. We chose to use snapshots of an order book from a trading application as an input with a sampling period of one second. The snapshots showed five depths of both bid and offer sides. The snapshots were cropped into a specific area that showed the order book part. Then, we used an optical character recognition (OCR) to extract the numbers from this area.

A. Data selection and data collection

To gather this information, we used a Python library call 'pyscreenshot' to take a snapshot of each frame with a sampling period of 1 second. We cropped the images into specific areas that contains the data. Thereafter, we want to do several image processing techniques like binarization, resizing, and border removal before feeding images to OCR. The output from OCR will be in a format of string numbers. The structure of limit order book data that we extracted from an image contains 4 rows and 5 columns. Each row in a column contains four numbers: volume of bids, bid price, offer price, and volume of offers from each depth. Starting from the top of row, the value of bids and offer will be the closest to the middle price. This is depth 1. There are five rows according to five depths visible to users.

Volume1	Bid1	Offer1	Volume6
Volume2	Bid2	Offer2	Volume7
Volume3	Bid3	Offer3	Volume8
Volume4	Bid4	Offer4	Volume9
Volume5	Bid5	Offer5	Volume10

Figure 2. Data contain in one timesteps

B. Optical Character Recognition(OCR)

The OCR we chose was Google Tesseract [8]. We implemented it with Python programming of the module named 'pytesseract' that acts as a program from system's call. Google's Tesseract is OCR with Unicode (UTF-8) support, and can recognize more than 100 languages and able to support many formats of output like plain text, hOCR (HTML), PDF, invisible-text-only PDF, and TSV. The snapshots we took from the trading program showed moving numbers in different color according to market conditions (red for down trend, green for uptrend, and yellow for unchanged). Also, the numbers will flash when they change. This made the OCR have errors. In our work, we used OCR to parse given images into string. In order to get high accuracy, it required additional image preprocessing steps like inverting images, rescaling, binarization, noise removal, rotation, and borders removal. We converted the output string to numbers and fed them to neural networks.

In real practice, there were some downsides from using OCR with images that come from binarization technique. The flicker effect that happened when numbers were changing caused the OCR to fetch correct numbers. Another problem happened with the numbers that has low RGB values. When converted to black and white, these numbers faded and cannot be extracted. Our experiment indicated that the amount these errors happen within an average of 8.5% of the total data.

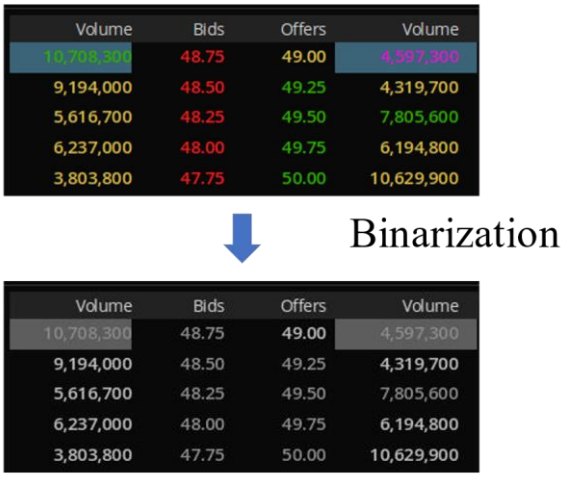
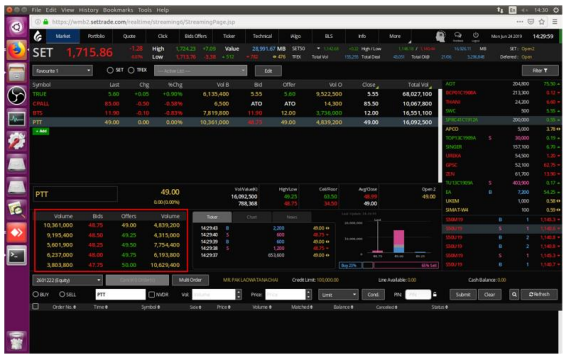


Figure 1. problem in binarization



Cropped

Volume	Bids	Offers	Volume
10,361,000	48.75	49.00	4,839,200
9,195,400	48.50	49.25	4,315,000
5,601,900	48.25	49.50	7,754,400
6,237,000	48.00	49.75	6,193,800
3,803,800	47.75	50.00	10,629,400

Binarization

Volume	Bids	Offers	Volume
10,361,000	48.75	49.00	4,839,200
9,195,400	48.50	49.25	4,315,000
5,601,900	48.25	49.50	7,754,400
6,237,000	48.00	49.75	6,193,800
3,803,800	47.75	50.00	10,629,400

OCR

STRING OF NUMBER

Figure 2. Input and output of OCR

C. *Data arrangement and label generation*

From the number data that is in 2-dimensional 4*5 format, we reshape it to 1-dimension of 20 data. After that, we took 10 of these 20 data and cascade them into one training sample. The training sample spans 10 timesteps, which make our input data to have 2-dimension with shape of 10*20.

Our data represents the state of order book back in past 10 second. These data represent price levels that people prefer or think that the price that suitable for current demand and supply. Next, we generated a label for each training sample by finding an average of mid-price further back 20 seconds in the past. The mid-price in each time step can be calculated by the sum of best bid and best offer divided by two. If the future price is more than the average mid-price, that period is considered as uptrend. If the future price is less than the average mid-price, that period is considered as downtrend. Otherwise, that period is considered as sideways.

$$Mid\ price = \frac{Best\ bid + Best\ offer}{2}$$

$$MA(\mathrm{P}_{1:20}) \begin{cases} < AVG(\mathrm{P}_{21:N}) => uptrend \\ = AVG(\mathrm{P}_{21:N}) => sideways \\ > AVG(\mathrm{P}_{21:N}) => downtrend \end{cases}$$

Figure 3. label generation

D. *Normalization and Class balancing*

Neural networks, prefer the input data to be in a range of 0 and 1 for easy training and stability. We normalized the price and volume data into a range of [0,1] by dividing them with the highest price and volume of that training sample.

From the type of labels that we acquired, most of the labels that we got was sideways (around 97%). The class imbalance will make the neural network stick to the majority class. We oversampling the uptrend and downtrend samples plus some additional noises to equalize the number of sample within each class.

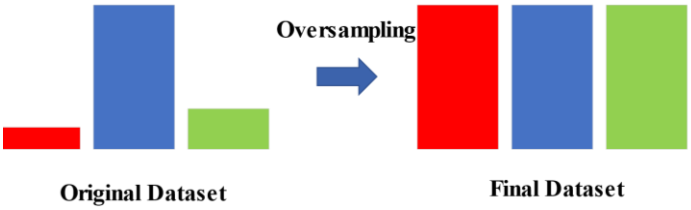


Figure 4. Class balancing by Oversampling

E. *Long Short-term Memory (LSTM)*

Long short-term memory has demonstrated its high efficiency for order book modelling as the best comparing with MLP, CNN [ref]. Order book data can be classified as a time series, which requires a model that can learn time dependency like an LSTM.

LSTM is an adaptive version from the architecture of an RNN model. Its main purpose is to solved the problem of long-term dependencies in RNN models. The structure of LSTM models is nearly the same as that of RNN models. They consist of inputs, outputs, and state cells with an addition of sigmoid gates. The additional gates work as forget gates, which are controlled by numbers ranged of 0 and 1 from the sigmoid gate. The input gate takes the number from the output gate and forget gate to control how much value it will pass to the next state. The main benefit of using LSTMs is in tasks that have sequence data as an input like time-series data prediction, human language processing, or handwriting etc. It can use the previous data as one of variable for prediction.

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i), \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_C), \quad (3)$$

$$C_t = (f_t \times C_{t-1}) + (i_t \times \tilde{C}_t), \quad (4)$$

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t \times \tanh(C_t). \quad (6)$$

Many stock traders made their decision of buy/sell based on the past data. LSTM uses the same mechanic, which is our aim as to make the model work as a human.

IV. Experiments

A. System Overview

We selected a security from the Stock exchange of Thailand (SET) during 24th June to 19th July which is around 4 weeks of data and 19 days (due to exclude holiday and weekends). There are two active sessions on each day: morning 10:00 a.m. to 12:30 p.m. and 2:30 p.m. to 4:30 p.m. There is a total of 4 active hours a day. We chose one of the large cap stock from the stock exchange of Thailand. Its trading value usually stay within top 10. This is because its order book behavior is expected to behave normally, and we can use it as a standard for other stocks.

The system records a video of an active order book, and takes snapshots of it every one second. The images will be preprocessed before being passed to the OCR, which gives outputs in a string format. Due to the blinking numbers caused by active trading, the OCR may misrecognize the numbers. Next, the string output from the OCR is converted to numbers and normalized before being sent to the LSTM neural networks. The output of our neural networks had three classes: uptrend, downtrend or sideways.

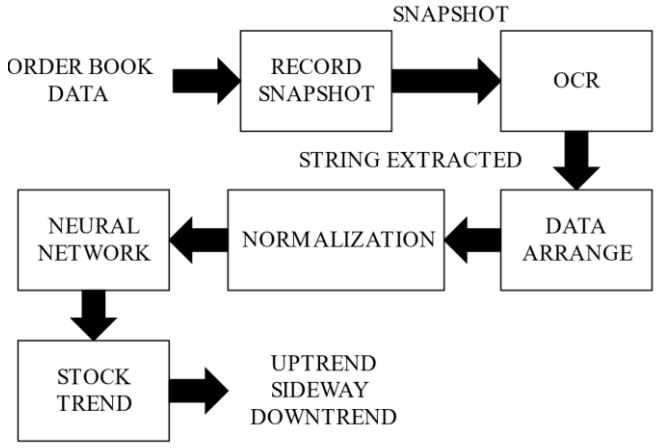


Figure 5. System Diagram

The main objective of this work is to demonstrate that a neural network that has a visual input can understand the behaviors of an active order book by correctly predicting the future trend of it. The work also investigated how the error of OCR and the sampling affects the overall prediction we make it robust enough for those errors.

B. Model Architecture

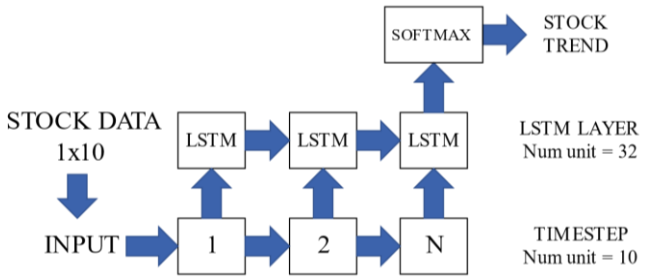


Figure 6. LSTM architect

Our neural network architecture used 10 timesteps of historical data to predict the future trend. We used an LSTM that has 1 layer with 64 number of hidden units inside and has an output of one number for the label of stock trend.

C. Training procedure

The 19-day data that we have consists of a mix of bullish and bearish trading periods for a total of around 300K entries. We divided it into three sets: 12 days of training set, 3 days of validation set, and 4 days of test set. Then, we performed class balancing, because majority of data is sideways. We also performed shuffling and used a batch size of 512.We trained the model with the training set with dropout rate of 0.5 to avoid the overfitting problem and make model more robust, also after training process follow with checked the accuracy of validate set together. When the accuracy of validation set started to drop, we stopped the progress in order to avoid overfitting and tested the result with test set. We reshuffled the data every time after finish one round of all batches.

Table 1. Accuracy from Training and Validation in each class

Class	Training	Validation
Uptrends	82.5%	68.69%
Sideways	81.23%	95.30%
Downtrends	90.02%	50.88%
Total	84.58%	71.65%

V. Result and Discussion

Our experiment predicts the trend of one second ahead from the current time. The results show the accuracy from the test set, in which the model never see it before. The results are divided into three classes, and the accuracy are shown separately as below.

Table 2. Test result in each class

CLASS	Accuracy (Test set)
Uptrends	59.11%
Sideways	68.87%
Downtrends	48.86%
Total	58.98%

Table 3. Confusion matrix of model

Class	Accuracy	Precision	Recall	F1 Score
Uptrends	59.11%	64.67%	59.112%	0.62
Sideways	68.87%	54.53%	68.876%	0.61
Downtrends	48.86%	59.54%	48.863%	0.54

From the test result and the confusion matrix, we achieved an average accuracy of the three classes as 58.984%. The result had an accuracy drop around 10% from that of the validation set. Precision, Recall and F1 score all were above 50%, except the recall of Downtrends which was less than 50%. This showed that the model performed well.

Because of the data has a mix of trends and has class imbalance, uptrend and downtrend are hard to predict than sideways. Also, if we can decrease error in the OCR progress, the result can still be improved in many ways given that many aspects of work can be polished with other methods that can remove noises.

VI. The Effect of Prolong Prediction Time vs. Accuracy

In the previous experiment, we used historical data to predict the trend of one step ahead. A question arose whether how far ahead in the future that we can predict the trend without sacrificing much of the accuracy. We performed an experiment with increasing period in the future to see the trend. The result showed that the model can achieve around the same accuracy within the first five seconds. However, the accuracy started to decreased after that as time goes on. During the first few seconds, the trends were not clear. This is why the model achieved the best accuracy at five-second prediction.

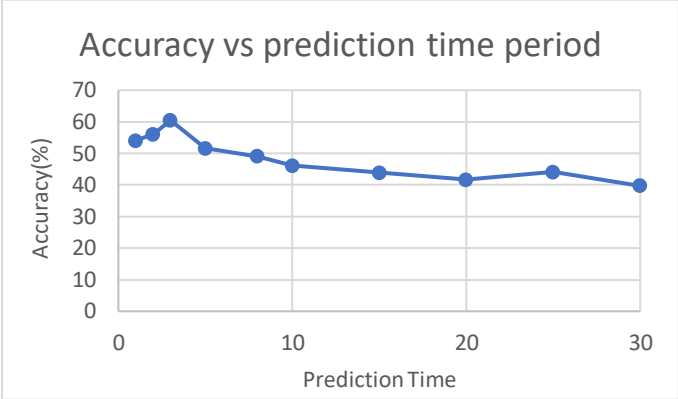


Figure 7. Accuracy vs prediction time period

VII. Conclusions

This work is an attempt to extract information from an order book of a trading stock market via visual information. Stock markets are usually regulated, therefore the access to such information is limited. We used a Google's Tesseract OCR to extract data from snapshots of an order book and passed the numbers to an LSTM neural network. The neural network was trained to learn traders' behaviors expressed on the order book and predict the future trends for the next few seconds.

We used one month of order book image snapshots of a big cap stock, captured from a trading application by the stock exchange of Thailand. The result showed that it can achieve a reliable prediction in most of the cases are above 50%. There were high prediction errors in the class downtrends. We expected that this may be due to the lack of variety in data. Our work had to deal with big data (images), while other similar work of stock prediction used numbers (text). This research is also restricted by hardware limitation, data sampling, and errors from the OCR algorithm. We tried to make the model more robust by adding dropouts to make the model less sensitive to noises.

VIII. References

[1] S. R. , An Overview of the Tesseract OCR Engine, Parana, Brazil: IEEE, 2007.

[2] H. Khaloozadeh and M. T. Motlagh, A new architecture for modeling and prediction of dynamic systems using neural networks: Application in Tehran stock exchange, Qazvin, Iran: IEEE, 2016.

[3] T. Gao, Y. Liu and Y. Chai, Applying long short term momory neural networks for predicting stock closing price, Beijing, China: IEEE, 2017.

[4] C.-Y. Lee and V.-W. Soo, Predict Stock Price with Financial News Based on Recurrent Convolutional Neural Networks, Taipei, Taiwan: IEEE, 2017.

[5] M. R. Vargas, C. E. M. dos Anjos, G. . L. G. Bichara and A. G. Evsukoff, Deep Leaming for Stock Market

Prediction Using Technical Indicators and Financial News Articles. 2018 International Joint Conference on Neural Networks, Rio de Janeiro, Brazil: IEEE, 2018.

[6] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj and A. Iosifidis, Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks, Thessaloniki, Greece: IEEE, 2017.

[7] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj and A. Iosifidis, Using deep learning to detect price change indications in financial markets, Kos, Greece: IEEE, 2017.

[8] J. Doering, M. Fairbank and S. Markose, Convolutional Neural Networks Applied to High-Frequency Market Microstructure Forecasting, Colchester, UK: IEEE, 2017.

[9] A. l. Calvez and D. Cliff, Deep Learning can Replicate Adaptive Traders in a Limit-Order-Book Financial Market, Bangalore, India, India: IEEE, 2018.

Hierarchical Attention Model for Acquiring Relationships Among Sentences

1st Hiroki Teranishi
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
teranishi@ss.cs.osakafu-u.ac.jp

2nd Makoto Okada
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
okada@cs.osakafu-u.ac.jp

3rd Naoki Mori
Graduate School of Engineering
Osaka Prefecture University
Osaka, Japan
mori@cs.osakafu-u.ac.jp

***Abstract*—In this paper, we propose a hierarchical attention model for summarization. Normally, sentences have relations among another sentence and it is important to consider these relations in summarizing. Our proposed model can make each sentence vectors from document composed of multi sentences and get relations among sentences from these vectors by the incorporated operation. As an operation of taking relations, we use self-attention and gated convolutional neural network. It has been reported that these operations can get dependencies among words, and self-attention is particularly powerful. Therefore we adopted these operations expecting the same work in sentences. We conducted an experiment of title generation by using Japanese news articles. We evaluated the performance of our proposed model by Rouge and visualized the relations among sentences.**

***Index Terms*—nueral network, attention mechanism, generative summarization**

I. Introduction

Sequence to sequence model represented by machine translation [1] has been actively conducted, and high accuracy has been reported in various tasks such as text summarization [2] and speech recognition [3]. In the sequence to sequence model, the input sequence is first encoded by some method, and then generate outputs in the decoder side. In most cases, the input is time-series data, and Recurrent Neural Networks(RNN) or Long Short Term Memory(LSTM) [4] is used. Recently, it has been reported that attention mechanism [5] and convolutional neural network [6] are effective for sequence models.

Attention mechanism makes it possible to generate outputs while focusing on important input words for a generation. Attention mechanism learns which input word is important by calculating the weight between the input sequence and target. The main issue with the previous sequence model is that a neural network needs to compress all the necessary information of an input sentence into a fixed-length vector. However, in attention model, it can choose a subset of input vectors, and this frees sequence model from having to compress all information of an input sentence, regardless of its length. Attention model achieves comparable to the existing state-of-the-art phase-based system on the English-to-French translation at the time.

Convolutional Neural Network(CNN) is a method of acquiring dependencies between inputs vectors with parallelization. Convolutional networks do not depend on the computations of the previous time step and therefore allow parallelization over every element in a sequence. By using CNN to words, it can obtain a feature representation capturing relationships within a window of k words by applying convolutional operations for kernels of width k. Gehring et al. achieve state-of-the-art in WMT' 16 English-Romanian by using Gated CNN and attention mechanism [6].

Self-attention learns dependencies between distant positions of input vectors. In obtaining the dependencies, while the convolutional neural network is limited to the kernel size of k, self-attention is not limited due to averaging attention-weighted position. Self-attention has been used successfully in a variety of tasks including reading comprehension, abstractive summarization, textual entailment. By using multi-layer attention, Vaswani et al. achieve state-of-the-art in WMT' 14 English-to-French [5].

However, all these operations are limited to acquiring dependencies among words, it is unclear whether they can capture relationships among sentences which position higher dimensions than words.

In this paper, we first propose the architecture of hierarchical attention model which obtains sentence vectors of input document which composed of multi-sentence. By making the attention model hierarchical, the Rouge score is improved compared to a previous single attention model. Furthermore, we execute operations of convolution and self-attention to the sentence vectors in our model and compared the transition of Rouge score and attention-weight in order to confirm that our model obtains the relationship between sentences.

II. Models

A. Attention Model

Attention model is based on encoder-decoder architecture. In the attention mechanism, it makes context bector c_t from attention weight at time t [7]. Attention weight of state i, a_{it} is computed by dot-product between each vectors $H = (h_1, ..., h_m)$ of input sentence $X = (x_1, ...x_m)$ from encoder and internal state $\overline{h_t}$ from the decoder [8].

$$\alpha_{it} = \frac{\exp\left(\overline{h_t} \cdot h_i\right)}{\sum_{j=1}^{m} \exp\left(\overline{h_t} \cdot h_j\right)} \tag{1}$$

The context vector c_t is a weighted sum of attention weights and encoder outputs.

$$c_t = \sum_{i=1}^{m} \alpha_{it} h_i \tag{2}$$

Using this context vector, the sentence $S = (s_1, ..., s_n)$ is generated as follows. Here, W_c and W_o is weight matrics.

$$\tilde{h_t} = \tanh W_c[c_t; \overline{h_t}] \tag{3}$$

$$p(s_t|s_{<t}, X) = \text{softmax}(W_o \tilde{h_t}) \tag{4}$$

B. Gated CNN Model

Gated CNN(GCNN) is a model that incorporates a gate structure into CNN [6]. As a gate structure, Dauphin et al. use gated linear units (GLU) [9]. GCNN has a block structure that computes intermediate vectors based on a constant number of input elements. The constant number is defined by kernel and number of blocks. In GCNN, convolution kernel is parameterized as $W \in R^{2d \times kd}$, $b_w \in R^{2d}$ and takes as input $X \in R^{k \times d}$ which is a concatenation of k input elements embedded in d dimensions and maps them to a single output element $Y = ([A\ B]) \in R^{2d}$. GLU is applied to Y and finally outputs vector $v([A\ B])$ like (5).

$$v([A\ B]) = A \otimes \sigma(B) \tag{5}$$

$\otimes$ is the point-wise multiplication and the output $v([A\ B])$ is half size of Y. The gates $\sigma(B)$ control which inputs A of the current context is relevant. GCNN takes residual connections from the input of each convolution to the output of the block [10]. Subsequent blocks operate over the k output elements of the previous block.

C. Self-Attention Model

Attention can be described as mapping a query and a set of key-value pairs to an output [11]. Normally, query is target from decoder, and key-value is source from the encoder. However, in self-attention, source and target are same and all come from the under hidden layer [5]. Self-attention can refer to all positions in under hidden layer, and obtain dependencies of the input sequence. In the self-attention, query is given at the same time and each attention is computed, and the same number of output vectors same to query can be obtained. The queries are put together into a matrix Q, and the keys and values are also put together into matrices K and V. Self-attention is computed like (6).

$$\text{Self-attention}(Q, K, V) = \text{softmax}(QK^T)V \tag{6}$$

$$Q = W_1 H, K = W_2 H, V = W_3 H \tag{7}$$

Here, W_1, W_2 and W_3 is weight matrix, and H is vectors from under hidden layer.

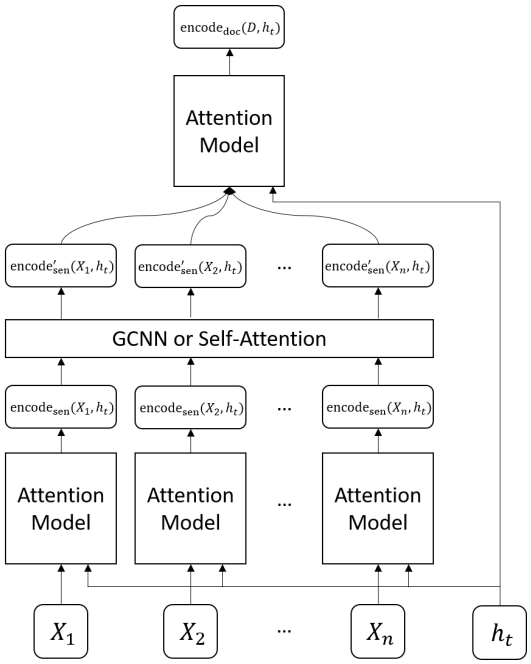


Fig. 1. Outline of proposed model

III. Proposed Method

In this research, we propose a hierarchical model that can obtain sentence vectors of the input document. Our proposed model is based on encoder-decoder model. The encoder side is a hierarchical structure of the attention model, and the decoder side is a language model. Each attention model is based on attention based summarization (ABS) model of Rush et al. [2]. As a language model, we use LSTM [4]. Input of our model is document $D = (X_1, ...X_n)$ which composed of multi sentences, and each sentence $X_i = (x_1, ..., x_m)$ is composed of words. Each sentence X_i are applied to attention model separately, and make sentence vectors $\text{encode}_{\text{sen}}(X_i, h_t)$ as section 2.1. Here, h_t is the hidden vector of the decoder at a time of t. Our model use scaled-dot attention [5] in conputing attention weight. After make sentence vectors, each sentence vectors are applied to subsequent attention model, and make document vector $\text{encode}_{\text{doc}}(D, h_t)$. Finally, summarized sentence is generated from this document vector in the same way as (3) and (4).

Hierarchical attention model can make each sentence vectors, however, it can not obtain the relations among sentence vectors. Therefore, we adopt GCNN and self-attention in order to obtain relations of sentences. As stated in section 2.1 and 2.2, these operations have been used for capturing words dependency. We add each operation layer before subsequent attention layer. Each oparation is done to the sentence vectors $\text{encode}_{\text{sen}}(X_i, h_t)$, and output $\text{encode}'_{\text{sen}}(X_i, h_t)$ which have relational information among sentences. Fig. 1 is the outline of our proposed model in the encoder.

IV. Experimental Setup

We generated titles of multi-sentence documents using the proposed model. We report a result of three methods, proposed model using GCNN or self-attention or nothing. We also show the results of the Rush et al. model [3] for comparison. Rush et al. model does not have a hierarchical structure, therefore we input multiple sentences regarded as one sentence. The dataset and parameters are the same for either method. The details of these settings are shown below.

A. Datasets

We use Mainichi-Sinbun-Dataset [1] as dataset. In this dataset, news articles consisting of titles and contents are organized by article genre, social, international, and economic. We used from 2008 to 2012 of this dataset. Furthermore, in order to increase dataset, we collect news articles from the internet. For these articles, we extracted only articles that contained 3 to 20 sentences and 7 to 20 words of the title. Finally, we got 171,060 news article dataset. Of these, 166,000 were used as training data and 5,160 as test data.

For Japanese morphological analysis, Mecab was used, and all numbers were replaced with ”num” tokens. We replaced words with low frequent occurrences to unknown words, and the threshold value of low frequent was 30,000.

B. Parameters and Optimization

We use 200 hidden units for embeddings in encoders and decoders, and dimensionalities of all linear layers are same. LSTM set the initial state to zero and the internal state dimension size was set to 200. In GCNN layer, we use 2 to 5 as kernal, and 1 to 5 is used as stacked blocks. When the kernel is 5 and stack 5 blocks, GCNN can get dependencies up to 21. In this experiment, we use articles less than 20 sentences, therefore we can get the dependencies between all sentences in this experimental setup.

Adam was used as an optimizer, and learning rate was 0.001. We learned to generate a title from articles and finished learning when the validation error became the minimum. Dropout was applied to the embeddings and the decoder output, and 0.1 was used as values. The batch size was set to 10.

C. Evaluation

For the evaluation, we used recall-based Rouge [12]. Rouge is evaluated by the number of overlapping units such as n-gram, word sequence, and word pairs between the ideal summary created by human and the computer-generated summary. We use Rouge-1 (unigrams), Rouge-2 (bi-grams), and Rouge-3 (tri-grams) of Rouge. As a maximum length of the generated title, we set 20. In order to compare each method, we evaluate by Rouge translation. We take Rouge score for every epoch up to 30. Each epoch score is computed by Rouge average of all test data.

[1]http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html

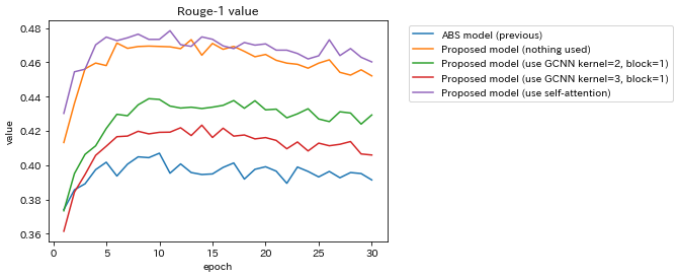


Fig. 2. Transition of Rouge-1 score

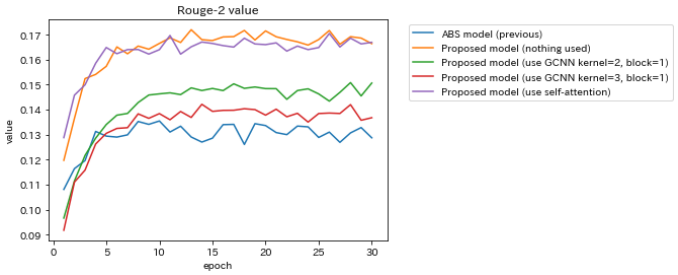


Fig. 3. Transition of Rouge-2 score

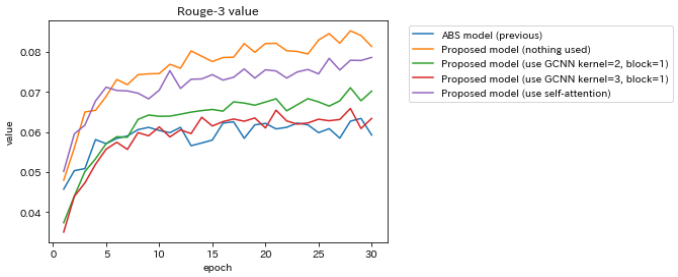


Fig. 4. Transition of Rouge-3 score

V. Experimental Results

We first show the result of the translation of Rouge score. About model with GCNN, We show the result of kernel is 2 and 3 when the stacked block is 1. Another result of using GCNN is lower. Proposed model with nothing used got the highest score. When we incorporated the layer of GCNN or self-attention, the Rouge score became low. Focusing on the difference between GCNN and self-attention, GCNN score was lower than self-attention. Convolution of word vectors by GCNN works well since the adjacent words have relationships. On the other hand, the convolution of sentence vectors by GCNN does not always work well since the adjacent sentence do not always have relationships. Consecutive sentences may have completely different contents. Self-attention can take dependencies among distant sentence vectors, therefore it seems that self-attention score performed better.

Next, we show the weight-value of the subsequent source-target-attention layer. Fig. 5 is proposed model with nothing used and Fig. 6 is proposed model with self-attention. It can

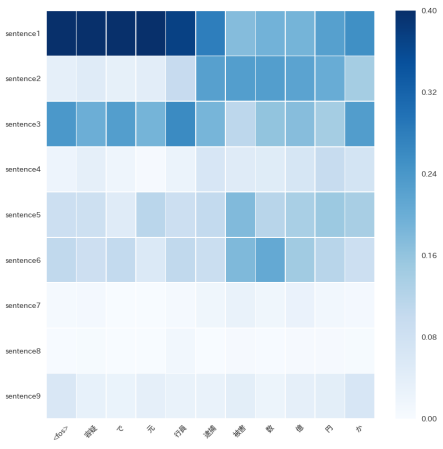


Fig. 5. Heatmap of attention-weight (nothing used)

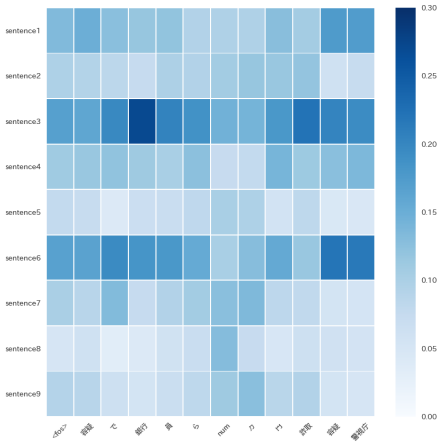


Fig. 6. Heatmap of attention-weight (use self-attention)

be confirmed from Fig. 5 that our proposed model can capture the feature that the headline is often used in the title generation of the news article. Focusing on Fig. 6, we can see that the weight-value is smoothed. As mentioned above, normally headline information is often used when generating titles from news articles. Therefore, the score drop of incorporating self-attention was attributed to taking dependencies more than necessary for summarization.

Finally, we show one example of the result of relations of sentence vectors of self-attention. Fig. 7 shows the relationship among sentences. The highest weight-value among Query and Value is connected by a line, and the top is Query and the bottom is Value. From Fig. 7, many sentences have a relation with the headline, it can be seen that the relations between sentences are taken. It can be said that the model with self-attention is more versatile considering that the summary is

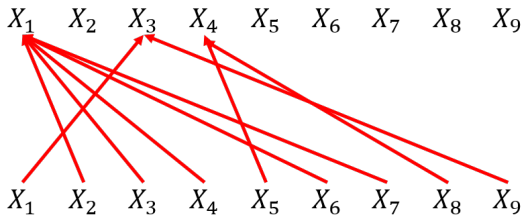


Fig. 7. Visualized relations among sentences

made by consideration of the relation between sentences.

VI. Conclusion

In this research, we proposed hierarchical attention model in order to make each sentence vectors, and we operate sentence vectors to taking relationship among sentences. As an operation to the sentence vectors, we use GCNN and self-attention. We find that self-attention is more suitable for taking relationships among sentences than GCNN. Moreover, we confirm that self-attention can take relationships among sentences by checking weight-value. Finally, we conclude that hierarchical attention with self-attention is versatile summarization model.

In this experiment, we use soft attention in all attention layer. However, it seems more suitable to restrict the use of attention-weight-value in the summarization. Therefore, in the future task, we try to incorporate hard attention in attention layer. Moreover, we would like to apply hierarchical architectures to other tasks which have dependencies among multi inputs.

References

[1] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. “Sequence to sequence learning with neural networks.” Advances in neural information processing systems. 2014.
[2] Chorowski, Jan K., et al. “Attention-based models for speech recognition.” Advances in neural information processing systems. 2015.
[3] Rush, Alexander M., Sumit Chopra, and Jason Weston. “A neural attention model for abstractive sentence summarization.” arXiv preprint arXiv:1509.00685 (2015).
[4] Hochreiter, Sepp, and Jurgen Schmidhuber. “Long short-term memory.” Neural computation 9.8 (1997): 1735-1780.
[5] Vaswani, Ashish, et al. “Attention is all you need.” Advances in neural information processing systems. 2017.
[6] Gehring, Jonas, et al. “Convolutional sequence to sequence learning.” Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.
[7] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. “Neural machine translation by jointly learning to align and translate.” arXiv preprint arXiv:1409.0473 (2014).
[8] Luong, Minh-Thang, Hieu Pham, and Christopher D. Manning. “Effective approaches to attention-based neural machine translation.” arXiv preprint arXiv:1508.04025 (2015).
[9] Dauphin, Yann N., et al. “Language modeling with gated convolutional networks.” Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.
[10] He, Kaiming, et al. “Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[11] Miller, Alexander, et al. “Key-value memory networks for directly reading documents.” arXiv preprint arXiv:1606.03126 (2016).
[12] Lin, Chin-Yew. “Rouge: A package for automatic evaluation of summaries.” Text summarization branches out. 2004.

Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO

Kwanruan Rusmee
Department of Industrial Technology Management,
Faculty of Industrial Technology
Muban Chombueng Rajabhat University,
Ratchaburi, Thailand
krusmee@hotmail.com

Narumol Chumuang
Department of Digital Media Technology,
Faculty of Industrial Technology
Muban Chombueng Rajabhat University,
Ratchaburi, Thailand
Lecho20@hotmail.com

***Abstract*—Current consumer behavior traders are used to create decision-making tools for entrepreneurs to produce products or services that meet consumer needs. For this reason, therefore focusing on data analysis using SMO techniques for predict system for the behavior of consumer buying personal car decision based on a total of 1,110 data obtained, with a total of 6 relevant car trading information features, consisting of income customers, type of car, down payment/cash booking, decision results requires that the answer in the forecast is divided into two classes, namely the class of buying cars and not buying cars When dividing the data into 50% training set and using 50% test set, 555 training set will be used and 555 test sets can be used to enter the learning and testing process. By random 50/50 in the selection of algorithms to be tested and the accuracy of 95.13%**

Keyword*— *predicting, behavior, consumer, personal-car decision, SMO

I. INTRODUCTION

The automotive industry of Thailand has been continuously growing [1], [2]. Observed by many cars, the launch of new models into the market and the use of various marketing strategies to compete because production volume car sales in Thailand are increasing in car production and sales every year [3]-[5]. Although the overall economy of European countries will be in a slowdown but the Thai economy is still active consistent with the image of increasing the production capacity of many cars to meet the demand both increase production capacity and add more product models to choose [6].

Toyota Motor Thailand Company Limited was registered in 1962 in addition to the assembly of cars. Toyota Motor Thailand Co., Ltd. has imported and exported cars [7]. Finished and auto parts to various countries around the world in more than 90 countries throughout the 48 years of operation and dedication of the company is an attempt to meet the highest demands of customers both in the world standard production process cutting-edge technology, environmental awareness, service quality and personnel development by including business expansion [8].

Mr. Mihinobu Tsukata, President of Toyota Motor Thailand Co., Ltd., announced the statistics of car sales in January-June 2019 in the passenger car market, No.1, Toyota, sales of 60,350 units, up 12.8%. 29.2% market share, 2nd place, Honda sales 48,889 units, increased 5.6%, market share 23.7% which reinforces the quality of the Toyota car [6]. In the after-sales service center, prices, promotions and special offers Each car is competing in sales. Current problems in consumer behavior traders are used to create decision-making tools for entrepreneurs to produce products or services that meet consumer needs. Therefore making various behaviors in deciding to buy all kinds of products is very important and must be worth the amount paid to provide effective products as shown in Fig 1.

Fig.1. Decision to buy a personal car [6]

In this paper, we focused on predicting consumer behavior of personal car products with SMO technique. By Toyota brand cars type of passenger car is a popular model with 5 models, namely Altis, Camry, C-HR, Vios and Yaris from total of 1,100 people.

II. RELATED THEORIES AND RESEARCH

A. Data Mining

The data mining [9] is a technique for analyzing data from different perspectives and can summarize results as useful information. Data mining process uses many principles such as machine learning, statistics, visualization techniques to discover and present knowledge in an easily understandable format another meaning is a process is done with a lot of data to extract information including the form and the relationship hidden in that big data set.

Knowledge Discovery in Database (KDD) or Data Mining work processes

The Knowledge Discovery in Database (KDD) process [10] refers to the process of searching for knowledge from a large number of data groups. Which has the process of data mining as the main process to find the interesting characteristics of these data. thus obtaining a rational data model useful and easy to understand. The implementation of data mining techniques will have different methods. Depending on the purpose that will be used to use what to do. What is needed to know? Therefore, various methods are proposed for different goals in order to obtain the appropriate results for the needs. After that, the data mining process can be shown as follows.

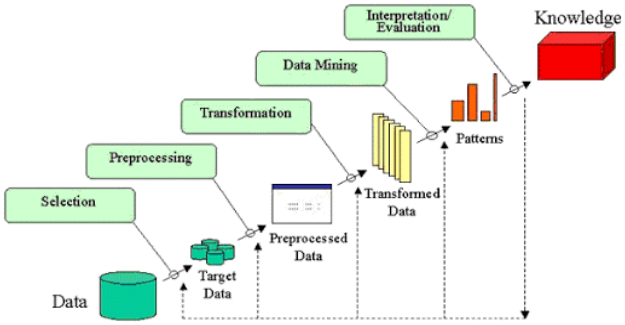


Fig.2. Data mining process [10]

B. Sequential Minimal Optimization (SMO) techniques

Overview of SMO is an algorithm based on Support vector machine techniques (Suport vector Machine). Used for training. The full algorithm is described in John Platt's paper. [11] However, the full SMO algorithm consists of customizations designed to speed up the algorithm. On large data sets and can detect algorithms If you want to use SVM in a real-world application Should use the full SMO algorithm or search for a software package for SVM. After using the algorithms described here, it should be quite easy to use the full SMO algorithm described in the paper of John Platt [12].

SVM optimization problems

Recall from the lecture notes that Support Vector Machine can analyze linear data [12].

$$f(x) = w^T x + b. \qquad (1)$$

The problem of binary identification is predicted y = 1 finally f (x) □ 0 and if y = 1, if f (x) <0 which considers the function f (x) can be expressed as well.

$$f(x) = \sum_{i-1}^{m} a_i y^{(i)} < x^{(i)}, x > + b \qquad (2)$$

Can replace the kernel K (x (i) , x)

SMO algorithm is a method that increases the efficiency of the vector machine. Or problems that must be solved

$$\max_a \quad W(a) = \sum_{i=1}^{m0} a_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y^{(i)}y^{(j)}a_i a_j \langle x^{(i)}, x^{(j)}\rangle \qquad (3)$$

$$\text{subject to} \qquad 0 \le \alpha i \le C, i = 1, \dots, m \qquad (4)$$

$$\sum_{i=1}^{m} a_i y^{(i)} = 0 \qquad (5)$$

Can use KKT condition to check the convergence of the most suitable spot For this problem, KKT conditions are.

$$a_i = 0 \Rightarrow y^{(i)}\left(w^T x^{(i)} + b\right) \ge 1 \qquad (6)$$

$$a_i = C \Rightarrow y^{(i)}\left(w^T x^{(i)} + b\right) \le 1 \qquad (7)$$

$$0 < a_i < C \Rightarrow y^{(i)}\left(w^T x^{(i)} + b\right) = 1 \qquad (8)$$

In other words, αi that meet these qualities will be the best solution for the optimization problems listed above. The SMO algorithm will repeat until all these conditions are met. Thus ensuring the reliability in forecasting.

Simple SMO algorithm as described, the SMO algorithm [13], [14] can select two α parameters, namely αi and αj, and adjust both common objectives of these α and can adjust the α parameter b depending on the new α's. This process is repeated until the convergence of α can explain these 3 steps more thoroughly.

Selection of α parameters, the most of full SMO algorithms perform behavior analysis to select αi and αj to optimize to increase the objective function as much as possible. For this large data set is very important to the speed of the algorithm because there are possible options for m (m - 1) for αi and αj, and some will result in less development than others.

However, for the new version of SMO use a much easier solution just repeat it more than αi, i = 1, ... m if αi does not meet KKT conditions within some numbers tolerance. We choose αj randomly from the remaining m-1 α and try to adjust to αi and αj. If there is no α, there is more change than every α. It is important to realize that the use makes this easy to understand. Because we didn't try to increase efficiency. All possibilities αi, αj have the possibility that some pairs can be optimized. Do not consider However, for the data set used in the problem set. This method will receive the same results as the selection process of the full SMO algorithm.

Optimizing αi and αj performance, the selecting the Lagrange αi and αj multipliers. In order to optimize, we will calculate the limits of the values of these parameters. Then we will solve the problem of maximizing the limit.

First of all, we want to search for the L and H boundaries such that L ≤ αj ≤ H so that αj is a response to the limit of 0 ≤ α αj ≤ C. It can show that these things are derived from the following:

$$If\ y^{(i)} \neq y^{(j)}, L = max(0, a_j - a_i), H = min(C, C + a_j - a_i) \quad (9)$$

$$If\ y^{(i)} = y^{(j)}, L = max(0, a_i + a_j - C), H = \min(C, a_i + a_j) \quad (10)$$

Now we want to find αj to add the most objective functions outside the L and H boundaries. We just cut the values of αj in this range. Can show (Try to find yourself using the contents of the class notes or see [12]) that αj is given by using the following equation.

$$a_j := a_j - \frac{y^{(i)}(E_i - E_j)}{\eta} \quad (11)$$

$$E_k = f(x^{(k)}) - y^{(k)} \quad (12)$$

$$\eta = 2\langle x^{(i)}, x^{(j)} \rangle - \langle x^{(i)}, x^{(i)} \rangle - \langle x^{(j)}, x^{(j)} \rangle \quad (13)$$

You can think of Ek as an error between the SVM output in the kth sample and the actual label. Y (k) can be calculated using the equation (2) when calculating the parameter. H Can use the K kernel function instead if needed anyway. αj is in the range [L, H]

$$a_j = \begin{cases} H & if a_j > H \\ a_j & if L \leq a_j \leq H \\ L & if a_j < L. \end{cases} \quad (14)$$

Finally, there is a fix for αj. We want to find the value for αi. The value obtained from the next equation.

$$a_i := a_i + y^{(i)} y^{(j)} \left(a_j^{(old} a_j\right) \quad (15)$$

Where old αj is the value of αj before optimization by equation (11) and (14).

The full SMO algorithm can handle rare cases that that = 0 for our purposes. If η = 0 can treat this as if we were not able to do with this pair's α.

Calculation of criteria b.

After optimizing αi and αj, we will select the criteria b, which is in accordance with KKT conditions. For example, ith and jth, if after adjustment, αi will not be bounded (eg 0 <αi <C), so the limit b1 The following is correct because it forces SVM to export $y^{(i)}$ when that input $x^{(i)}$.

$$b_1 = b - E_i - y^{(i)} \left(a_i - a_i^{(old}\right) \langle x^{(i)}, x^{(i)} \rangle - y^{(j)} \left(a_j - a_j^{(old}\right) \langle x^{(i)}, x^{(j)} \rangle. \quad (16)$$

In the same way, the b_2 criteria is correct if 0 <αj < C

$$b_2 = b - E_j - y^{(i)} \left(a_i - a_i^{(old}\right) \langle x^{(i)}, x^{(j)} \rangle - y^{(j)} \left(a_j - a_j^{(old}\right) \langle x^{(j)}, x^{(j)} \rangle. \quad (17)$$

If both <αi <C and 0 <αj <C when both of these criteria are correct and they are equal, if both new α are in the scope (such as αi = 0 or αi = C and αj = 0 or αj = C) So all the limits between b1 and b2 are in accordance with KKT conditions. We give b: = (b 1 + b 2) / 2. This will help make the equation more complete for b.

$$b := \begin{cases} b1 & if\ 0 < a_i < C \\ b2 & if\ 0 < a_j < C \\ (b1 + b2)/2 & otherwise \end{cases} \quad (18)$$

C. *Related research*

Saichon Sinsomboonthong [15] Comparison of the efficacy in predicting diabetes outcomes is a prediction of the outcome of diabetes in a hospital By comparing the efficiency of predicting the results of the six methods of group classification, which is the most similar method, using IBK type algorithm, decision tree method using J48 type algorithm, artificial neural network method by using a multi-layered perceptron algorithm, how to support vector machines using SMO algorithms, polynomial kernel, logistic regression methods, Two groups and methods Naif Bay showed how neural networks are accurate. Average absolute error value and the average mean square error is 95.54%, 0.0491 and 0.0396. Therefore, the neural network method is effective in predicting the best results.

Saichon Sinsomboonthong [16] Comparison of the efficiency in predicting game addiction of children and adolescents in Bangkok with the method of classifying all 7 methods, the nearest neighbor k method is using IBK type algorithm, the decision tree method using the J48 type algorithm, artificial neural network method using multi-layer algorithms, how to support vector machines using SMO algorithms, polynomial kernel methods, rules-based methods using decision table type algorithms, binary logistic regression methods and the method of Naive Bay It was found that the most effective decision making method was 92.17%.

Samorn Lekkla and Jaree Thongkam [17] Forecasting foreign exchange rates using time series in this research, the four methods of group classification are linear regression (LR), Multi-Layer 9 Perceptron (MlP), Support Vector Machine Regression (SVMR) and Sequential Minimal Optimization Regression (SMOR). Better than LR, MlP, and SMOR without the MAE value and the lowest RMSE value to 1.11 + -2.10 and 1.13 + -2.14 respectively.

Pataya Boonraksa and Jaree Thongkam [18] Comparison of the efficiency of the road accident model using time series techniques using 5 powerful techniques to create models, namely linear regression (LR), Artificial Neural Network (ANN), Support Vector Machine Regression (SVMR), Sequential Minimal Optimization Regression (SMOR) and Gussian Process (GP) the results showed that SVM technique is effective in creating models for predicting the number of valuable road accidents. Which is the lowest error value compared to the model created from LR, ANN, SMOR and GP techniques.

Choubey, Mishra and Pandey [19] Studied the SMO model, a technique in the Support Vector Machine (SVM) technique family, to predict the amount of runoff. And rainfall which is a basin area of 39,372 square kilometers using the principle of sliding windows, each month the colors of the Narmada River, Madhya Pradesh Province,

India with information on rainfall, sediment temperature, rainfall, rainfall levels from water discharge is the parameter value is a variable for use in classification. Information used from the years 1975 - 2010 with the SMOreg technique. dividing the data into two sets: Training and Testing. The teaching set is from 1975 to 2010. To measure the performance of the model, he used Mean Absolute error (MAE)., Root Mean Square Error (RMSE), Relative absolute Error (RAE) and Root Relative Square Error (RRSE). Test results show that the SMOreg model has RMSE equal to 2.3731. The RAE value is 65.28% and the RRSE is 62.491% which is That compares with the actual value Will see that the difference between the predicted value and the actual value is significantly different with significance.

Harsh Valecha, Aparna Varma, Ishita Khare, Aakash Sachdeva, Mukta Goyal [20]. In the ultramodern age of technology, anticipation of market trend is very important to observe consumer behaviour in this competitive world as trends are volatile. Building on developments in machine learning and prior work in the science of behaviour prediction, we construct a model designed to predict the behaviour of Consumer. The aim of this research paper is to examine the relation between consumer behaviour parameters and willingness to buy. First we investigate to find relationship between consumer behaviour to buy products on changing parameters such as environmental factor, organizational factor, individual factor and interpersonal factor. Thus this paper proposes time-evolving random forest classifier that leverages unique feature engineering to predict the behaviour of consumer that affect the choice of purchasing the product significantly. Results of random forest classifier are more accurate than other machine learning algorithm.

K.Maheswari, P.Packia Amutha Priya [21]Customer buying behavior is identified by people's personality and character. These personality characters vary from person to person. The character includes quality, motivation, occupation and income level, perception, psychological, personality, reference groups and demographic reasons learning, beliefs, attitude, Culture and social forces.

Nowadays, data mining normally used to investigate the customer activities on shopping by using various algorithms and methods. Data mining has gradually raised and it gains numerous industries which applies this technology. Each and every activity of a customer is stored as a byte of data in a database to collect information such as how the customer spends their valuable time, day in buying decision. Most frequents items bought and quantity of buy is also considered. These data are collected without the knowledge of the customer. In this paper, the dataset is used to analyze and categorize the customer based on their purchase behavior. The classification is performed by SVM algorithm. The inventory data set and sales data set which is available in the internet is used in this work and the performance is evaluated by using the algorithms. The experimental results are analyzed and it shows that the proposed methodology analyze a customer behavior in a better way.

III. METHODOLOGY

In this research, the research team has set the steps into 5 main steps, as shown in Figure 3, the process of conducting research.

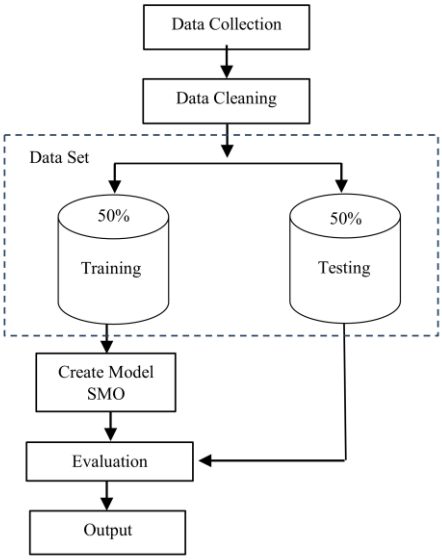


Fig.3. The overview of the predicting system for the behavior.

A. Data Collection Steps

The researcher used information on the behavior of buying a car, Toyota brand, type of passenger car. There are 5 popular models, namely Altis, Camry, C-HR, Vios and Yaris from the Toyota Service Center. Data collection is 1,110, with 6 relevant car trading information features, including group, income, car_type, down, cash, class with meaning and possible values as shown in Table I.

TABLE I. ATTRIBUTE USED IN RESEARCH

No.	Attribute	Explanation	Note
1	group	There are 3 groups of customers.	M = meal W = female Trade = Merchant
2	income	income	25,000 - 80,000 Lessthan 80,000 Morethan 80,000
3	car_type	Car_type	Altis Camry C-HR Vios Yaris
4	down	down	Yes = Reservation No = Don't Reservation
5	cash	cash	Yes = Yes (cash) No = No (cash)
6	class	Decision results	buy = buy don't buy = don't buy

B. Data Cleaning

Procedure due to the data set used in this experiment Originally in paper format. Therefore, data conversion must be digitized with human labor. Causing the problem of human error, therefore having to correct the data with errors to be accurate, such as converting data from Thai language to be in English eliminate various noise in the data set, such as the space between characters. as shown in Table II.

TABLE II. DATA SET THAT HAS BEEN CLEANED UP

group	income	car_type	down	cash	class
m	25,000-80,000	Altis	Yes	No	Don't buy
m	25,000-80,000	Altis	Yes	No	Don't buy
m	25,000-80,000	Altis	Yes	No	Don't buy
m	25,000-80,000	Camry	Yes	No	Don't buy
m	25,000-80,000	C-HR	Yes	No	Don't buy
w	lessthan80,000	Yaris	Yes	No	Buy
w	lessthan80,000	Yaris	Yes	No	Buy
w	lessthan80,000	Yaris	Yes	No	Buy
w	lessthan80,000	Yaris	Yes	No	Buy

C. Data set

This experiment has a total of 1,110 items from a total of 2 classes. In this research, the data set will be divided into 2 sets : training set is used to create a model to find patterns of car consumer behavior. Personal and testing set for evaluation of models created by both sets of data set by random method.

D. Create Model Using SMO

Information used in modeling a total of 555 items were used to create models with SMO algorithm by writing a pseudo-code as shown in Fig.4.

Input :
C: regularization parameter
tol : numerical tolerance
max_passes: max # of times to iterate over α's without changing
$(x^{(1)}, y^{(1)}), \ldots, (x^m, y^m)$: training data

Output:
$\alpha \in R^m$: Lagrange multipliers for solution
b ∈ R : threshold for solution

◦ Initialize $\alpha_i = 0, \forall i,\ b = 0$.
◦ Initialize passes = 0.
◦ while (passes < max_passes)
◦ num_changed_alphas = 0.
◦ for i = 1, …m,
◦ Calculate $E_i = f(x^{(i)}) - y^{(i)}$ using (2).
◦ if $((y^{(i)}E_j < -tol \;\&\&\; a_i < C) \,||\, (y^{(i)}E_i > tol \;\&\&\; a_i > 0))$
◦ Select j ≠ i randomly.
◦ Calculate $E_j = f(x^{(j)}) - y^{(j)}$ using (2).
◦ Save old α's: $a_i^{(old)} = a_i$, $a_j^{(old)} = a_j$.
◦ Compute *L* and *H* by (10) or (11).
◦ if (*L* == *H*)
continue to next i.
◦ Compute η by (14).
◦ if (η >= 0)
continue to next i.
◦ Compute and clip new value for αj using (12) and (15).
◦ if $(|a_j - a_j^{(old)}| < 10^{-5})$
continue to next i.
◦ Determine value for a_i using (16).
◦ Compute b_1 and b_2 using (17) and (18) respectively.
◦ Compute *b* by (19).
◦ num_changed_alphas := num_changed_alphas + 1.
◦ end if
◦ end for
◦ if (num_changed_alphas == 0)
passes := passes + 1
◦ else
passes := 0
◦ end while

Fig. 4. The peseudo-code for simplified SMO.

IV. EXPERIMENT RESULTS

From the experiment found that model for consumer behavior forecasting for personal car products with SMO. There are six attributes of the relevant car trading information. The forecast answer is divided into two classes: car purchase class and not buying cars when dividing the data into the training set of 50% and using the test set 50% will use the training set 555 items and using the test set of 555 items. The prediction results are accurate to 528 items, with an average value of 95.13%, 27 false predictions, representing an average. 4.86% as shown in Table 3.

TABLE III. PREDICTION ACCURACY TEST RESULTS

Data Set	**Accuracy Rate**	**Error Rate**	**RMSE**
Testing	95.13	0.05	0.22

TABLE IV. MATRIX OF CONFUSION FROM FORECASTING CONSUMER BEHAVIOR OF PERSONAL CAR PRODUCTS WITH SMO

Group classification techniques	**Real situation**	**Consumer behavior classification of personal car products**	
		Buy	**Don't buy**
SMO	Buy	528	27
	Not buy		

From table IV, it was found that the forecast results were accurate to 528 items, accounting for 95.13%, 27 false predictions, accounting for 4.86%. This is because class = don't buy contains only 54 items, which is considered very small. While the data purchased is as high as 1,056 items, representing the data purchased only 0.05 percent of the set data purchased. Thus resulting in a set of predictions that do not buy all errors.

V. CONCLUSIONS

In this research the researcher focused on consumer behavior of personal car products. Based on the data obtained from all 1,110 cases, with the characteristics of the relevant car trading information, all 6 attributes, consisting of income customers, type of car, down/cash booking, the decision result is divided into 2 classes: car buying class and not buying 555 cars. Information that is entered into the learning process and tested by a random 50/50 method for selecting algorithms to be tested. The researcher chose to use SMO technique. The test results showed that the accuracy was 95.13%. The researcher found from this research found that the Error 4.87% accounted for 27 items from the experiment due to the data set of consumer behavior. Who decided not to buy a car there is a small amount of information that makes the pattern search. The behavior that is decided not to buy is still not performing well, resulting in the SMO classifying the wrong class. Should add information that is not buying behavior into the data group will make the data more efficiency.

REFERENCES

[1] Z. Liu, F. Zhao and S. Zhao, "Basic Principles of Technology Transformation in Long Value Chain in the Manufacturing Industry and Key Technology Innovation Issues in China-A Case Study of the Automotive Industry," 2019 8th International Conference on Industrial Technology and Management (ICITM), Cambridge, United Kingdom, 2019, pp. 257-264.

[2] S. M. M. Pisal et al., "Moving towards sustainable human capital development for Malaysian automotive industry," 2015 10th Asian Control Conference (ASCC), Kota Kinabalu, 2015, pp. 1-4.

[3] C. Schöll, S. Möller and A. Longino, "Assessing the impact of inaccurate decision support systems on experts' behavior and decisions," 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, 2017, pp. 1-3.

[4] Bo He, Hongen Ren and Wei Kan, "Design and simulation of Behavior-Based Reactive Decision-making Control System for autonomous underwater vehicle," 2010 2nd International Conference on Advanced Computer Control, Shenyang, 2010, pp. 647-651.

[5] T. T. Zin, P. Tin, T. Toriu and H. Hama, "A Human Behavior Analyzer Framework for consumer product search engines," 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), Tokyo, 2014, pp. 138-139.

[6] Mulla Salim Husen, D. Yadav Sanjay, D. Shinde and G. Deshpande, "Strength enhancement of 2010 Toyota Yaris Passenger Sedan driver seat as per Federal Motor Vehicle Safety Standards 207/210-A FEA approach," 2013 International Conference on Energy Efficient Technologies for Sustainability, Nagercoil, 2013, pp. 174-177.

[7] D. Sato, K. Lee, K. Araki, T. Masuda, M. Yamaguchi and N. Yamada, "Design and Evaluation of Low-concentration Static III-V/Si Partial CPV Module for Car-rooftop Application," 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), Waikoloa Village, HI, 2018, pp. 0954-0957.

[8] Chinrit Ratanaruj, Ajarn Prathanporn Sopa Chitthawattana. (2010). Study of problem and formulation of marketing strategies to increase potential Competition of passenger cars Toyota Camry Hybrid Toyota Motor Thailand Company Limited. Faculty of Business Administration, University of the Thai Chamber of Commerce.

[9] S. Thaiparnit, N. Chumuang and M. Ketcham, "A Comparitive Study of Clasification Liver Dysfunction with Machine Learning," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-4.

[10] J. Ming, L. Zhang, J. Sun and Y. Zhang, "Analysis models of technical and economic data of mining enterprises based on big data analysis," 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, 2018, pp. 224-227.

[11] N. Chumuang, "Comparative Algorithm for Predicting the Protein Localization Sites with Yeast Dataset," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 369-374.

[12] Platt, John. Fast Training of Support Vector, In Advances in Kernel Methods - Support Vector Learning, B. Scholkopf, C. Burges, A. Smola, eds., MIT Press 1988.

[13] P. Peng, Q. Ma and L. Hong, "The research of the parallel SMO algorithm for solving SVM," 2009 International Conference on Machine Learning and Cybernetics, Hebei, 2009, pp. 1271-1274.

[14] N. Chumuang and M. Ketcham, "Model for Handwritten Recognition Based on Artificial Intelligence," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-5.

[15] Saichon Sonsomboonthong. Comparison of efficacy in predicting diabetes outcomes. Journal of Science and Technology, Year 26, Issue 2, March-April 2018.

[16] Saichon Sinsomboonthong. A comparison of the efficiency in a predicting game addiction of children and Adolescents in Bangkok. Journal of Science and Technology, Year 26, Issue 3, May-June 2018.

[17] Samorn Lekkla and Jaree Thongkam Journal of Information Technology Management and Innovation Department of information technology Sarakham Rajabhat University Year 5, Issue 2 July - December 2018.

[18] Pataya Boonraksa and Jaree Thongkam By using time series techniques. Journal of Information Technology Management and Innovation Department of information technology Sarakham Rajabhat University Year 4, Issue 2 July - December 2017.

[19] S.M.V. Choubey, S.K.Pandey, "Time Series Data Mining in Real Time Surface Runoff Forecasting Support Vector Machine," International Journal of Computer Applications (0975-8887), vol. Volume 98- No.3 ,, July 2014.

[20] Harsh Valecha, Aparna Varma, Ishita Khare, Aakash Sachdeva and Mukta Goyal, “Uttar Pradesh Section Intenational Conference on Electrical Electronics and Computer Engineering (UPCON)” IEEE Gorakhpur, India, Nov 2018.

[21] K.Maheswari, P.Packia Amutha Priya, “ International Conference on Intelligent Techniques in Control, Optimization” IEEE Sriviliputhur, India, Marcj 2018.

An Efficiency Comparison for Predicting of Educational Achievement Based on LMT

Sudarat Thiennoi[1, 2] ,Narumol Chumuang[3], Chairit Siladech[4]
[1]*Dep. of Industrial Technology Management,* [1,3]*Faculty of Industrial Technology,*
[2]*Dep. of business computer,* [2]*Faculty of Management Science,*
[4]*Dep. of Educational Research and Evaluation, Faculty of Education,*

[1,2,3,4]*Muban Chombueng Rajabhat University, Ratchaburi, Thailand.*

[1,2]sudarut_th@hotmail.com, [3]lecho20@hotmail.com, and [4]siladech9@gmail.com

***Abstract*—This paper focuses on efficiency comparison for predicting of educational achievement by using the decision tree technique. The data are used in this work from graduation of Muban Chombueng Rajabhat University (MCRU) during the academic year A.D. 2013 - A.D. 2016. The data set consist of 2,437 records with 15 attributes such as debut consisting of gender, race, nationality, religion, number of siblings, number of siblings who have studied, career, father, occupation, mother, income, father, income, mother, status, father, mother, district, semester, graduation and educational achievement. The class separated into two classes which is "complete" and "incomplete". The group of decision tree used for comparing. In our experimental, the data are divided into 50/50 % for training and testing set by random. The accuracy rate of LMT is the highest that showed 100%.**

Keywords— Data mining, Predictions, Decision trees, LMT

I. INTRODUCTION

In the rapidly changing world of the 21st century of Thailand gives importance to education [1] because of education is an important tool in building people building society and building a nation as the main mechanism for quality manpower development around the world, it is important and dedicated to the development of education in order to develop their human resources to be able to keep pace with changes in the country's economic and social systems creating an educational system for academic excellence.

Problems or termination of operation, graduate soon causing loss of financial and time costs for both institutions and students with the cumulative GPA used as a contributing factor in the evaluation [2] but the cumulative GPA of only one factor is not able to properly analyze the status of future students. Because there are other factors involved such as cumulative GPA of major majors etc. Which the student data in the database has a lot of storage but lacking the management of data sets complicated and the information is not clear making improvements in techniques and tools for analysis in large data sets one of the techniques is the making data mining that has received the attention of many paper [3] called educational data mining to be able to solve problems that are directly specific [4] educational data mining involves the extraction of useful things from a large educational database to evaluate the learning process of students improved. [5] student registration system information has collected data at a large database but those data have not been utilized as they should both the information is interesting and can be used to search for useful knowledge by applying student data in the past to analyze to create a prototype for forecasting or predicting future trends such as forecasting opportunities for graduation or predicting the trend of freeing [6] is a technique used to handle large data. [7] there are a variety of techniques including association techniques technical clustering classification and modeling for use in forecasting as an example in the paper of Gulati H [8] has predicted the termination of students by using data mining techniques Omkar and Parag's paper [9] predicted the condition of students using data mining techniques J48, RandomForest and REPTree's paper and the faculty [10] have applied data mining to predict the status of students using the J48 algorithm.

The organization of this paper begin with introduction, the next one is theoretical and related work. The 3rd topic is proposed methodology experimental results summary of paper results.

II. THEORETICAL AND RELATED WORK.

A. Data Mining

Data mining [11] is a process that deals with a large amount of data to find hidden patterns and relationships. In that data set data mining has been applied in many types of jobs both in the business that helps the management's decision in science and medicine including economic and social aspects.

Data mining is like an evolution to store and interpret data stored in a database that can retrieve information from information to data mining that can discover hidden knowledge in data the purpose of data mining use to find important information that is mixed with other information. In a database that is not just random called KDD (Knowledge Discovery in Database) or knowledge search there are 5 styles Namely 1) searching for relationship rules 2) classification and forecasting 3) grouping of data 4) finding abnormal values occurring 5) trend analysis.

In this paper the forecasting technique is used [12] which is an important technique of searching knowledge on a large database purpose is the creation of one feature separation model based on other features the model obtained from data classification will be able to consider classes in data that has not yet been divided into future groups.

B. Decision Tree Techniques

Decision tree [13] is a technique that results in the appearance of tree structures which when there is information that needs to be grouped will bring various features of that data compared to the path in the tree until the destination class which is the same data group the tree will consist of nodes which each node has a feature as a test tree branch shows the possible values of the selected feature and the leaf is the

bottom of the decision tree represents a group of (class) is the result of the prediction the node at the top of the tree is called root node decision trees have a measure of the ability of each group of attributes or factors. as follows:

Gini index values that indicate which features or factors should be used as a grouping feature of the J48 and CART algorithms as in Eq.1.

$$Gini(t_i) = 1 - \sum_{i=1}^{N}[p(t_i)]^2 \qquad (1)$$

Entropy the estimate of data is a separate value using Identification of the ID3 algorithm

$$Entropy(t_i) = 1 - \sum_{i=1}^{N}[p(t_i)] \log_2 p(t_i) \quad (2)$$

when t_i is features that are used to measure entropy

$p(t_i)$ is the proportion of the number of members of the group *i* with the number all members of the sample group. Which each algorithm will give different results of the decision tree structure decision tree techniques used in paper have this [14].

1) DecisionStump technique is a tree-based decision that is the basis in the system one of the decisions the money will be split at the root of the foundation by identifying an attribute and value pairs.

2) Technical J48 is a technique that has been modified from C4.5. The algorithm that applies to the decision of this decision tree the growing use called depth-first strategy by which the information should therefore not be tested will be divided into a series and the test set based on the best information gain.

3) Logistic Model Trees (LMT) technique [15] is a highly effective technique for learning with instructors for normal class predictions and numerical values consists of a decision tree structure that has a regression function in the leaf section as with any general decision tree, testing one of the node-related features within this algorithm is a combination of linear logistic regression and tree induction forces decision tree with a linear regression model that leaves as follows Eq.3 – Eq.6 as follows:

$$S = \bigcup_{teT} S_t\,, S_t \cap S_t, = \theta for t \neq t' \qquad (3)$$

Unlike ordinary decision trees, the leaves t ∈ T have an associated logistic regression function ft instead of just a class label. The regression function ft takes into account a subset Vt ⊆ V of all attributes present in the data (where we assume that nominal attributes have been binarized for the purpose of regression), and models the class membership probabilities as

$$P_r(G = i|X = x) = \frac{e^{f\,j(x)}}{\sum_{k=1}^{J} e^{Fk(x)}} \qquad (4)$$

where

$$F_j(x) = a_0^j + \sum_{k=1}^{m} a_{U_k}^j \bullet U_{,} \qquad (5)$$

or, equivalently

$$F_j(x) = a_0^j + \sum_{k=1}^{m} a_{U_k}^j \bullet U_k \qquad (6)$$

if $\alpha^j_{vk} = 0$ for $v_k \in / V_t$. The model represented by the whole logistic model tree is then given by

$$f(x) = \sum_{t \in T} f_t(x) \bullet I(x \in S_t) \qquad (7)$$

where $I(x \in St)$ is 1 if $x \in S_t$ and 0 otherwise.

Both standalone logistic regression and ordinary decision trees are special cases of logistic model trees, the first is a logistic model tree pruned back to the root, the second a tree in which $V_t = \emptyset$ for all $t \in T$.

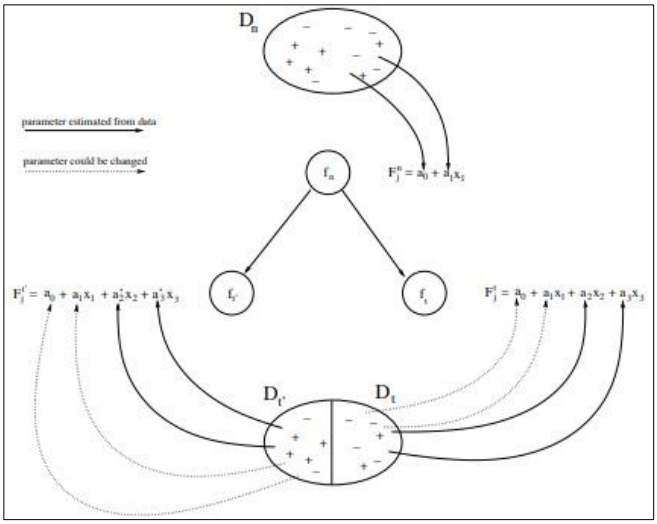


Fig. 1. Building logistic models by incremental refinement. The parameters a_0, a_1 are estimated from the training examples at *n*, the parameters a_2, a_3 and a_2, a_3 from the training examples at *t* and *t* respectively. Attribute x_1 has a global influence, x_2, x_3 have a local influence. [16].

As an Fig. 1, consider a tree with a single split at the root and two successor nodes [16]. The root node n has training data D_n and one of its children t has a subset of the training data $D_t \subset D_n$. Fitting the logistic regression models in isolation means the model f_n would be built by iteratively fitting simple regression functions to D_n and the model f_t by iteratively fitting simple regression functions to D_t. In contrast, in the 'iterative refinement' approach, the tree is constructed as follows. We start by building a logistic model f_n at *n* by running LogitBoost on D_n, including more and more variables in the model by adding simple regressions f_{mj} to the F_{nj} (the linear class function for each class *j* at node *n*). At some point, adding more variables does not increase the accuracy of the model,2 but splitting the instance space and refining the logistic models locally in the two subdivisions created by the split might give a better model. So we split the node *n* and build refined logistic models at the child nodes by proceeding with the LogitBoost algorithm on the smaller set of examples D_t, adding more simple regression functions to the F_{nj} to form the F_{tj} . These simple linear regressions are fit to the response variables of the set of training examples D_t given the (partial) logistic regression already fit at the parent node. Fig.1 illustrates this scheme for building the logistic regression models.

4) RandomForest techniques [17] is a technique of classification data without any cutting data out or called Is a

regression tree which is defined from the Bootstrap group that comes from the test set data and use random selection tree decision-making process this technique focuses on improving the efficiency of single tree classification, such as CART and C4.5, which is a method that can handle disturbing data well.

5) RandomTree [18] techniques is the drawing of the initial decision randomly generated from the set of trees defined In this tree each tree in the set of all trees is likely to be randomized random tree methods this was put to use as much about science disciplines neck computing machine learning.

6) REPTree techniques is a tree-based decision that has the speed to create tree-based models and tree regression by relying on information gain as a basis for separation and cutting branches to reduce mistakes that will occur which will be arranged extinguish numeric values only the error value will be handled by using the method of C4.5 [19].

7) HoeffdingTree Techniques [20] is a decision algorithm the decision tree increases every time that can be learned from a large data stream by assuming that the distribution creation example does not change over time HoeffdingTree take advantage of the fact that small samples are often enough to choose the best separation feature.

C. Performance measurement

For the method of measuring the effectiveness of the proposed method is as follows. [21]

1) Accuracy is a measure of efficiency overall model forecasting as the Eq. 8

2) The sensitivity or true-positive rate is probability or predictive ratio in the group interested as Eq. 9.

3) Specificity or true-negative rate is the probability or the correct forecast ratio in other groups as the Eq. 10.

$$Accuracy = \frac{TP+TN}{TR+FP+FN+TN} \times 100 \qquad (8)$$

$$Sensitivity = \frac{TP}{TP+FP} \times 100 \qquad (9)$$

$$Specificity = \frac{TN}{TN+FN} \times 100 \qquad (10)$$

by

TP is valuable predictions for interested groups,

FP is wrong forecasting values in the interested group,

TN is accurate predictions in other groups,

FP is wrong predictions in other groups,

D. Related work

Pichai Rawengwan and Dr.Watsadisiri Saengtrakul [22] studied factors affecting the education status of students to create a model to predict the educational status of students by using the data of factors that are expected to affect the educational status of 2,272 people it was found that the factors affecting the status of the students were 9 factors used to create models to predict students' status with the decision tree technique Bayian Beach Liv Network and Green Creation Logistics found that the accuracy of 82.85%, 78.98% and 78.50% with the decision tree technique Christie's Grace And learning Bayan Netshot found that the model created with the decision tree method gives the best accuracy.

Pattanapong Dolrat and Jari Thongkham [23] Comparison of the effectiveness of models in forecasting student achievement this paper uses 6 effective techniques for modeling: C4.5, Random Forest, Random Tree, Reduced Error Pruning (REP Tree), k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM). It was found that the C4.5 model had the highest predictive efficiency of 95.36%.

Security at the General Medicine Ward Palmas [24] optimization techniques to tree on a series of imbalances by means of random sampling, adding small sample groups for data on internet addiction by developing the model with the decision tree technique , J48, ID3, LMT, CART and Random Forest. Results of the forecasting efficiency of the model showed that the Random Forest technique can predict better than J48, ID3, LMT and CART with accuracy percent 87.15.

Yauvana the successor and faculty [25] she used to bring the search factor that affects the level of academic performance by reducing introduced into each variable data the water of 4,591 people at 17 attribute the results to find the most efficient model C4.5 percent 73.55 higher than the k-Nearest Neighbor and Naïve Bayes, which is 66.63 and percentage 49 Results for factors affecting the level of academic performance variables that are critical of the selected sorted by descending a grade binders in the brothers offloaded crates of old memories were all brothers and branch department

Kamal Bunkar and colleagues [26] presented a model, making the grades of students using the classification of making data mining. In the study, the paperer used features consisting of each course score history of education study behavior using decision tree techniques C4.5, ID3, CART found that students with learning criteria must improve validity is equal to 0.22% of the criteria passed 0.856%

Ratchapol Kladchuen and Charan Saenrarat [27] conducted a paper to compare the efficiency of the predictive algorithm on the learning achievement of 5,100 students using 3 techniques for data classification, including J48graft, Naïve Bayes and Rule Induction compare the predictive model performance During the use of all features and the forward select feature selection, the results of the performance tests with the highest accuracy of 2 values were compared with the T-Test method it was found that the forward selection J48graft technique and all feature selection with the accuracy of 83.08% and 81.71%. both types of tests are significantly different at the level of .05

III. PROPOSED METHODOLOGY

This is paper work the paper team has set the steps into 5 main steps, as shown in Fig. 2, the process of conducting paper.

A. Step input dataset

The paper data set has been collected during the academic year 2013 - 2016. There are a total of 3,625 students who have graduated and not graduated person it features personal information on the total consists of 15 attribute as shown in Table I.

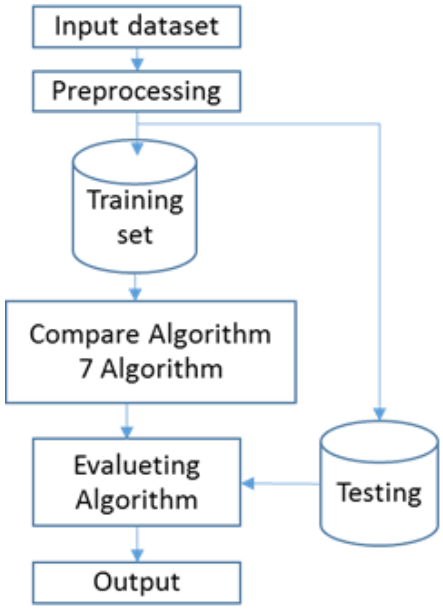


Fig. 2. Shows our system the framework process.

B. *Pre-processing steps*

Cleaning data to ensure the integrity of the data to ensure data integrity and quality of water data to analyze correctly a process to clean the data below.

1) If the information is missing on your part personal characteristics more than 3 characteristics, such as lack of career characteristics, father, career, mother, father's income and the mother's income will be cut off

2) Please fill in the missing word non in order to fully complete the information can be used to operate it.

3) Convert data from Thai Language to be in English

4) Eliminate various noise in the data set such as the space between characters.

When the procedure to clean the data and make the data of all 2,431 person as detailed in Table II and Table III.

C. *Training set steps*

By bringing all the data cleaning process successfully which has a total of 2,431 items, consisting of 15 personal attributes and requires the answers to be divided into two classes: graduate and non-graduate classes. After that, take the data set into the learning process, perform data analysis based on the 10-fold cross validation method

D. *Compare the algorithm*

This step is a comparison of all 7 algorithms from each test result the algorithm provides different accuracy values as detailed in Table IV.

TABLE I. SHOWS DETAILS INDIVIDUAL CHARACTERISTICS OF EACH STUDENT

Attributes	Personal characteristics of students
1	sex
2	race
3	nationality
4	Religion
5	Number of brothers
6	Number of brothers who studied
7	Father's career
8	Mother's occupation
9	Father's income
10	Mother's income
11	Status of parents
12	District
13	province
14	Graduation semester
15	Educational achievement

TABLE II. SHOWS A DETAILED DESCRIPTION OF INFORMATION RELATED TO THE STUDENT'S PERSONAL STATUS

order	Attributes	explanation	note
1	Sex	sex	Sex show f = male , m = female
2	Race	race	Show value Ethnicity
3	Nation	nationality	Show value Nationality fee
4	Rell	Religion	Show value Religious display
5	Num_Bro	Number of brothers	The numerical value shows the total number of brothers.
6	Bro_Stu	Number of brothers who studied	The numerical value shows the number of brothers who have studied.
7	Fa_career	Father's career	Show value Father's career
8	Mo_career	Mother's occupation	Show value Mother's occupation
9	Inc_Fa	Father's income	Show value Father's income
10	Inc_Mo	Mother's income	Show income mothers
11	Sta_par	Status of parents	The present status of the parents.
12	Dis	District	Show value District display fee
13	Pro	province	Show value Province display fee
14	Suc_term	Semester ends	Graduation Fee for Graduation
15	Class	Achievement results Educational	Graduation = graduation No = No graduate

TABLE III. EXAMPLE OF DATA SETS RELATED TO STATUS OF STUDENTS AND CONVERT THE DATA ALREADY

Sex	Race	Nation	Reli	Num_Bro	Bro_Stu	Fa_creer	Mo_creer	Inc_Fa	Inc_Mo	Sta_par	Dis	Pro	Suc_term	Class
m	T	Thai	Buddhism	3	3	agriculture	agriculture	Moderate	Low	Live_togethe	Chombueng	Ratchaburi	2	Gradution
f	T	Thai	Buddhism	2	2	electrician	agriculture	Moderate	Low	Live_togethe	Bangphae	Ratchaburi	2	Gradution
f	T	Thai	Buddhism	3	1	ivate busine	agriculture	Moderate	Low	Live_togethe	Chombueng	Ratchaburi	3	Gradution
m	T	Thai	Buddhism	4	1	agriculture	agriculture	Moderate	Low	Live_togethe	Chombueng	Ratchaburi		No
f	T	Thai	Buddhism	2	2	agriculture	agriculture	Moderate	Low	Live_togethe	Chombueng	Ratchaburi		No
m	T	Thai	Buddhism	1	1	agriculture	agriculture	Moderate	Low	Live_togethe	Bankha	Ratchaburi		No
f	T	Thai	Buddhism	5	3	Employment	agriculture	Moderate	Low	Live_togethe	Photharam	Ratchaburi		No
f	T	Thai	Buddhism	1	1	Employment	agriculture	Low	Moderate	Separated	Paktho	Ratchaburi		No
f	T	Thai	Buddhism	2	2	Civil servants	agriculture	Moderate	Moderate	Separated	Muang	Ratchaburi	2	Gradution
m	T	Thai	Buddhism	3	3	Employment	agriculture	Moderate	Moderate	divorce	Chombueng	Ratchaburi	1	Gradution
m	T	Thai	Buddhism	2	2	agriculture	agriculture	Low	Moderate	Live_togethe	Suanphueng	Ratchaburi		No
m	T	Thai	Buddhism	3	1	Employment	agriculture	Low	Moderate	Live_togethe	Paktho	Ratchaburi		No
f	T	Thai	Buddhism	3	1	Civil servants	agriculture	Moderate	Moderate	Live_togethe	Suanphueng	Ratchaburi	2	Gradution
f	T	Thai	Buddhism	1	1	Employment	agriculture	Moderate	Moderate	Live_togethe	Suanphueng	Ratchaburi	1	Gradution
m	T	Thai	Buddhism	2	2	agriculture	agriculture	Moderate	Moderate	Live_togethe	Muang	Ratchaburi	3	Gradution
f	T	Thai	Buddhism	2	2	agriculture	agriculture	Moderate	Moderate	Live_togethe	Chombueng	Ratchaburi	2	Gradution
m	T	Thai	Buddhism	3	3	agriculture	agriculture	Moderate	Moderate	Live_togethe	Chombueng	Ratchaburi	1	Gradution
f	T	Thai	Buddhism	4	3	Merchant	agriculture	Moderate	Moderate	Live_togethe	Paktho	Ratchaburi	2	Gradution
f	T	Thai	Buddhism	1	1	agriculture	agriculture	Moderate	Moderate	Live_togethe	Banpong	Ratchaburi	2	Gradution
f	T	Thai	Buddhism	3	1	agriculture	agriculture	Moderate	Moderate	Live_togethe	Paktho	Ratchaburi		No
f	T	Thai	Buddhism	1	1	agriculture	agriculture	Moderate	Moderate	Live_togethe	Bangphae	Ratchaburi		No

TABLE 4 SHOWS THE COMPARISON OF THE ACCURACY OF THE ALGORITHM. USED IN PAPER

order	Algorithm	Accuracy rate (%)
1	DecisionStump	100
2	HoeffidingTree	100
3	J48	72.94
4	LMT	100
5	RandomForest	67.44
6	RandomTree	62
7	REPTree	71

E. Procedure Testing and Evalueting

By using a total of 2,431 items consisting of 15 personal attributes and requiring the answers to be divided into 2 classes: 975 graduated classes and 1,456 non- graduating classes. Have entered the testing process the division's learning 50% and tests 50% by way of random step test this using an algorithm LMT they can compete with ledger advances such as to make better decisions the tree while more easily interpreted if you want to take the time to create a tree structure which is mostly the cost modeling regression logistic regained the node but the structure is not complicated test to be 100%

IV. EXPERIMENTAL RESULTS

This paper study has brought the students during the academic year 2013 – 2016 of 2,431 items with features individual involved 15 attribute set to answer in the forecast into 2 classes is the class graduate and graduating class brings a simple belt with data to compare performance. 7 Technical Decision Tree technique shown accuracy in Fig. 3.

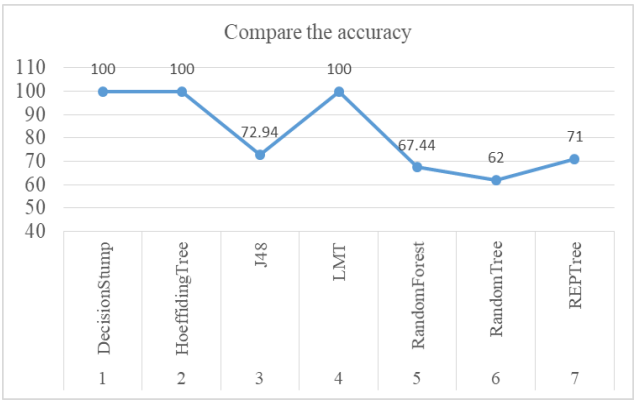


Fig 3. showing the comparison of the accuracy of the algorithms used in paper.

Compare the forecasting efficiency Test with algorithms The test results showed that the DecisionStump technique Provide 100% accuracy. HoeffidingTree technique for 100% accuracy, J48 technique gives 72. 94% accuracy. LMT technique gives 100% accuracy need 71%

From figure 2 shows that there are 3 ways to accuracy as well, including HoeffdingTree, the LMT and DecisionStump is 100% of RandomTree the accuracy least 62% , RandomForest the accuracy 67.44% REPTree provides 71% accuracy, J48 gives 72.94% accuracy

V. CONCLUSIONS

This paper, we focused on comparing the effectiveness of learning achievement of undergraduate students. By bringing information from the office of academic promotion and registration which is the information of students of Muban Chombueng Rajabhat University during the academic year 2013 - 2016, based on the data obtained, a total of 3,625 people were brought through the data cleaning process. Therefore, there are 2,431 remaining information and

compiling all 15 related personal attributes, including gender, race, nationality, religion, number of siblings, number of siblings, education, career, father, occupation, mother, income, father, income, mother, status, parents Graduation semester And educational achievement Has determined that the answers in the forecast are divided into 2 classes, which are 975 graduated classes and 1,456 non-graduated classes that have passed the paper process. Perform data analysis on the basis of the 10-Fold cross validation method. Bring the information into the learning process and test by the random method 50/50. Compare the forecasting efficiency. Test with algorithms The test results showed that the DecisionStump technique Provide 100% accuracy. HoeffidingTree technique For 100% accuracy, J48 technique provides 72.94% accuracy. LMT technique provides 100% accuracy. RandomForest technique provides 67.44% accuracy. RandomTree technique gives 62% accuracy and REPTree technique provides accuracy. Need 71%

As for testing, the paper team selected an algorithm to test the Logistic Model Trees (LMT) technique. The reason that the paper team chose LMT is because LMT technique is a highly effective technique for learning with Instructor For normal class predictions and numerical values Consists of a decision tree structure that has a regression function in the leaf section As with any general decision tree, testing one of the node-related features within this algorithm Is a combination of linear logistic regression and tree induction forces From the test results, LMT technique provides 100% accuracy.

REFERENCES

[1] Ministry of Education. Educational Development Plan of the Ministry of Education (No. 12, 2017 - 2036)

[2] Shovon, M., and Haque, M. (2012). An Approach of Improving Student's Academic Performance by Using K-means Clustering Algorithm and Decision Tree. International Journal of Advanced Computer Science and Applications, vol. 3(8), pp. 146-149.

[3] Piatetsky- Shapiro, G. (2007). Data Mining and Knowledge Discovery 1996 to 2005: Overcoming the Hype and Moving from "University" to "Business" and "Analytics". Data Mining and Knowledge Discovery, vol. 15 (1), pp. 99-105.

[4] Romero, C., Ventura, S., and Garcia, E. (2008). Data mining in Course Management Systems: Moodle Case Study and Tutorial. Computers & Education, vol. 51 (1), pp. 368-384.

[5] Chan, AY, Chow, KO, and Cheung, KS (2008). Online Course Refinement through Association Rule Mining. Journal of Educational Technology Systems, vol. 36 (4), pp. 433-444.

[6] J. Han, M. Kamber, J. Pei, Data mining concepts and techniques, 3rd ed., ElSevier , USA, 2011.

[7] The S. Sinsomboon, Data Mining, JamjureeProduct, Bangkok, Thailand, 2015.

[8] The H. Gulati, Predictive analytics using data mining techniques Computing for Sustainable Global Development (INDIACom), IEEE Conference Publications, 11-13 March, pp.713- 716. New Delhi, India, 2015.

[9] S. Omkar, M. Parag, Dropout Students Using Data- Mining Techniques, IJR. 2 (1) (2015) 365-375.

[10] M. Songsee, C. Palaman , W. Wuttisak , Application to Predict Student Status Using Data Mining for Southern College of Technology, SCT.3(1) (2010) 73-89

[11] B. Kijsirikul. *Data Mining Algorithm*. Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand, 2002.

[12] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.

[13] C. Kaewchinporn *Data Classification with Clustering Techniques*. Thesis in Computer Science, King Mongkut's Institute of Technology Ladkrabang Thailand, 2010.

[14] R. Quinlan. "Induction of decision trees." Machine Learning, Vol. 1, No. 1, pp. 81-106, 1986.

[15] Preecha Khakham, Narumol Chumuang and Maha sak Ketcham "Isan Dhamma Character Recognition on Palm Leaves based on Logistic Model Tree "IEICE Technical Report, vol.114, no.286, pp.109- 114, 5-6 November, 2014.

[16] Landwehr, N., M. Hall, and E. Frank: 2003, 'Logistic Model Trees'. In: Proc 14th European Conference on Machine Learning. pp. 241–252, Springer-Verlag.

[17] L. Breiman and R. Forests. *Machine Learning*, Vol. 45, No. 1, pp.5-32, 2001.

[18] S. Thaiparnit, N. Chumuang and M. Ketcham, "A Comparitive Study of Clasification Liver Dysfunction with Machine Learning," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-4.

[19] N. Chumuang, " Comparative Algorithm for Predicting the Protein Localization Sites with Yeast Dataset," 2018 14th International Conference on Signal- Image Technology & Internet- Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 369-374.

[20] Albert, Bifet, Eibe, Frank, Geoffrey, Holmes & Bernard, Pfahringer. (2007). Accurate Ensembles for Data Steams Combining estricted Hoeffding Trees Using Stacking. Journal of Machine Learning Paper, 225-240.

[21] J. Thongkam, G. Xu, Y. Zhang and F. Huang. "Toward breast cancer survivability prediction models through improving training space." Expert Systems with Applications, pp.12200–9, 2009

[22] Pichai between Weng days and Dr. Wednesday EL good Pastor's family of light. The model for forecasting educational status of students. Graduate Paper Proposal Conference National and international levels 2017.

[23] Pattanapong Dolrat and Jari Thongkham. Comparison of the effectiveness of models in forecasting success in education of the vocational certificate students. The Thirteenth National Conference on Computing and Information Technology. pp. 8 –13. 2017.

[24] T. at the General Medicine Ward palm. Optimization techniques, decision trees on a data set that imbalance by means of random samples add minorities for Internet addiction as a disease. Journal of Information Technology. Vol. 12, pp 54-63, 2016.

[25] Yaowapharatsuen, Jirattha Bhubobob and Virat Phongsiri. Comparison of data mining algorithms for factor analysis that affects the level of student performance. Journal of Science and Technology Mahasarakham University. 2013.

[26] Bunkar, K., Bunka, R., Umesh, r., & Bhupendra, P. (2012). Data Mining: Prediction for Performance Improvement of Graduate Students using Classification. Ninth International Conference on Wireless and Optical Communications Networks (WOCN), 20-22 Sept. (pp.1-5). India: Indore.

[27] Ratchapol Kladchuen and Charan Saenraj. Compa rison of algorithm performance and selection of appropriate features for predicting academic achievement of vocational students. Paper Journal Rajamangala University of Technology Thanyaburi , ISSN 1686- 8420, Vol 17, Issue 1, 2018.

Prediction Model for Amphetamine Behaviors Based on Bayes Network Classifier

[1,2]Kumnung Vongprechakorn, [3]Narumol Chumuang, [4]Adil Farooq

[1] *Dep. of Industrial Technology Management Program,* [1,3]*Faculty of Industrial Technology*
[1,3]*Muban Chom Bueng Rajabhat University, Ratchaburi, Thailand.*
[2] *Suan Phueng Police Station, Suan Phueng District,* Ratchaburi , Thailand
[4] *The BioRobotics Institute, Sant'Anna school of advanced studies, Italy*

[1,2]kumnung_vong@hotmail.com, [3]Lecho20@hotmail.com, [4]adil.farooq@santannapisa.it

***Abstract*— This paper focus to present a prediction model for drug addiction of the accused in type 1 drug abuse case as amphetamine. Case studies in the Suan Phueng Police Station area Ratchaburi province. The data set that used for modeling is obtained from the collection from Suan Phueng Police Station from 2016 - 2018. The data set 1,598 items consist of gender, age, number of offenses, education status, nationality, occupation, and non-drug abuse. For our contribute a prediction model into two classes as "take" and "untake" by using data mining techniques namely Bayes Network classifier. In our experimental, a group of bayes are used to comparison such as Bayes Network, Naive Bayes and Naive Bayes Updateable. The results displayed the Bayes Network classifier shown the highest accuracy rate, Naive Bayes and Naive Bayes Updateable with 81.53, 80.85 and 80.85 respectively.**

Keywords—Prediction model, Amphetamine, Behaviors, Bayes Network Classifier

I. INTRODUCTION

Drug problems are a global problem [1]-[3]. The drug situation began to intensify and has a tendency to increase significantly that the drug problem continues to intensify continuously which the statistics of the arrest are constantly increasing [1]. In considering drugs to society the concealment of drugs in the human body should be investigated strictly [2] while the current automatic detection method is missing and the examination rate with the human's eye is low efficiency [3]. Therefore, the introduction a new method using separation of fractional surface properties, direction and vector machine support. (SVM) to identify x-ray images [4]-[5]. Many people think that drug addiction reliance and patience is the same. In fact, each word means something very different about how the drug affects the body and the brain of a person learning the difference is important together [6]. During that time, increasing economic and social development also causes The expansion of drug problems as well both in terms of population and drug use [7], there are new methods of drug addiction, including hemp amphetamine and volatile substances [8]. The social situation family and challenging academic often during this time, children will experience various substances such as smoking and alcohol for the first time when they enter high school, teenagers may encounter drug availability [9]. Drug use of older adolescents and social activities that use drugs taking a level of risk It's normal for teenage development. Desire to try new things and having more freedom is good for health but may also increase teenagers' tendency to try different drugs of the brain that controls the decision and the decision was not fully developed until people were in the early or mid 20s [2], [4]. This limits the ability of adolescents to accurately assess the risk of trial drug use and making young people vulnerable to pressure from friends [10]. There are possible methods to prevent the broad use of substances, both theoretical and practical. The effectiveness of these methods, mainly in reducing the use of substances unknown because the research evidence does not exist or is inconclusive in some methods, because the research evidence does not exist or is inconclusive. We have no evidence effectiveness, including many popular control strategies, such as the policy of not accepting security measures such as searching for lockers and the presence of police in schools pulling advancement in toxicology, molecular biology, genetics and clinical medicine such as parents' efforts to protect their children through the use of test kits at home. Increased to detect drug use or vaccination for high-risk children. More research is needed on prevention activities than has been studied today [11].

Fig. 1. Show event of a crime caused by amphetamines.

This paper interested for finding a solution to the problem by using amount of data set 1,598 cases of drug addicts from 2016 - 2018 of Suan Phueng Police Station to study predicting model by using data mining techniques. In the classification of information in order to increase the accuracy of data classification to be as accurate as possible.Therefore, it is necessary to

choose the technique of data classification that is appropriate for the data and obtain the accuracy at the acceptable level with the use of narcotics characteristics or drug abuse effects of drug users received from the arrest results in terms of sex, age, number of times, other offenses and drug addiction cases occupational status of drug users. For substance abusers, those who encounter substance abuse This research presents three techniques for data classification, which are data classification techniques by Bayes Network classifier, Naive Bayes classifier and Naive Bayes Updateable for analysis of values prediction model accuracy and apply techniques that gives the most accuracy to be used as a model in developing the forecasting system for further career guidance.

The paper is organized further in five sections. Section II explains the related theories. Section III presents the implementation, Our experimental and results are discussed in section IV, conclusion and reference are discussed in section V and VI.

II. RELATED THEORIES

A. Bayes Network Classifier

Classification is a basic task in analyzing data and recognizing patterns that need to create a classifier. Function defined class label to the instance described by the data set of attributes induction of the classifier from the dataset of a pre- classified instance is an important issue in the process machine learning [12]. One of the most effective classifiers in the sense that the predictive efficiency competes with the modern classifier is the so- called Bayes Network classifier as described, for example [13], [14]. This classifier learns from data on the conditional probability practice of each attribute. A_i received a class label C classification is done using Bay rules. To calculate the probability of C that defines a specific example of $A_1, \ldots, A_n$ and predict the class with the highest probability. This calculation is possible by creating strong independence. Hypothesis: all features A_i there is conditional independence by determining the value of the class C [15].

B. Learning Bayes Networks

Considering the limited set U = $\{X_1,\ldots,X_n\}$ of discrete random variables at each variable X_j of discrete random variables in which each variable may use values from the finite sets expressed by val X_j [16]. We use capital letters like this. such as X, Y, Z for variable names and lowercase letters such as x, y, z to show only the values taken by those variables. The set of variables is represented by capital letters such as X, Y, Z and assignments to variables in these sets are represented by lowercase letters in bold type. x, y, z. we use val (X) clearly. Finally, give P is a joint probability distribution of variables in U and let X, Y, Z is a subset of U. We say X and Y conditionally independent Z if for x all $\in$ Val(X), $y \in$ Val(Y), $z \in$ Val(Z), $P(x \mid z, y) = P(x \mid z)$ whenever $P(y, z) > 0$.

Bay network is an annotated graph that encodes the probability of distribution across a set of random variables. U official Bayesian network for U is a pair $B = \langle G, \Theta \rangle$. The first element is G, is an acoustic graph directed which has vertices corresponding to the random variable $X_1, \ldots, X_n$ and at the edges showing direct dependency between graph variables G independent hypothesis coding: Each variable X_i is independent from the non-discriminatory person of his parents in grams, the second element of pair is Θ refers to a set of parameters that determines the network volume. It has parameters $\theta_{xi} \mid II_{xi} = PB\ (x_i \mid II_{xi})$. For each possible value x_i of X_i and I_{Ixi} of II_{Xi} by II_{Xi} referring to the guardian group of X_i in G. A network Bayesian B determine a unique connection, more probability distribution U determined by

$$P_B(X_{1,\ldots,}X_n) = \prod_{i=1}^{n} P_B(X_i|\Pi_{x_i}) = \prod_{i=1}^{n} \theta_{X_i} \mid \Pi_{x_i} \quad (1)$$

We know that we may associate with the concept of minimization and definition of Bayesian Network operated by Pearl (1988) But this link is not related to the content in this paper.

give $B = hG$; £ i is a Bay network and provides D = fu1; :::; uNg is a training set where each UI will assign values to all variables in U rating function MDL of the network B receive training data sets D written by MDL (BjD)

$$MDL\left(B \left| D = \frac{\log N}{2} \right| B \right| - LL(B|D) \quad (2)$$

by jBj is the number of parameters in the network. The first term represents the length of the network description. B in it, it counts the bits needed to encrypt a specific network. B by 1 = 2 ¢ $logN$ B its are used for each parameter in £. The second phase is the denial of the probability of the record of B giving D:

$$LL(B|D) = \sum_{i=1}^{N} \log(P_B(u_i)), \quad (3)$$

Which measures how many bits are needed in the explanation, depending on the probability distribution PB. Record opportunities also have statistical interpretations.: the higher the likelihood of saving information, the closer B is the modeling of probability distributions in data D give ^ PD (¢) is an empirical distribution determined by the frequency of events in D that is to say ^ PD $(A) = 1$ no messages yet P j $1A$ (uj) for each event A μ Val (U) by $1A$ $(u) = 1$ if you 2 A and $1A$ $(u) = 0$ if u 62 A. Using the Eq. 1 with the opportunity to record and change the order of the totals, the well known degradation of the probability is recorded according to the structure of B:

$$LL(B|D) = N \sum_{i=1}^{N} \sum_{\substack{x_i \in Val(x_i) \\ \Pi x_i \in Val(\Pi x_i)}} \widehat{P_D}\left(x_{i,\Pi x_i}\right) \log(\theta x_i|\Pi x_i) \quad (4)$$

It's easy to show when this expression is maximize

$$\text{ed}.\theta_{x_i} \mid \Pi x_i = \widehat{P_D}\,(x_i|\Pi x_i)\,. \quad (5)$$

C. Bayes Network as Classifiers

When using the methods described, we can make the network Bayesian *B* that encrypts the distribution $P_B\,(A_1, \ldots, A_n, C)$ from the set of training set. Then we can use the results model to make the set of attributes *a1, ... , an,* Classifier according to *B* return label *c* that increases the probability of the back $P_B\,(\,a_1, \ldots, a_n\,)$ as much as possible please note that by stimulating the classifier in this manner, we are talking about the main concerns expressed in the introduction. Removing the bias presented by the independent assumptions embedded in the naive Bayes classifier. This method is justified by the exactness of the Bayesian learning process with large data sets, learned networks are closely assessed for the domain controller probability distribution. Assume that the instance is sampled independently from the static distribution. Although this argument gives us a theoretical basis in practice, we may find cases in which the learning process returns networks with scores. *MDL* quite good, which does not work as well as classifier.

In order to understand, the possible differences between good predictive accuracy and *MDL* At a good score, we have to check the score. *MDL* again remember that the word "opportunity" is recorded in Eq. 2 is a measure of the quality of learning models and $D = \{u_1, \ldots, u_N\}$ show that training sets in each classification mission U_I are tuple of styles $<a^i_1, \ldots, a^i_n, c^i> I$ say that assigning values to attributes $A_1, \ldots, A_n$ and class variables *C* can write the probability log function as Eq. 3.

$$LL(B|D) = \sum_{i=1}^{N} \log P_B\left(c^i \left| a^i_1, \ldots, a^i_n \right.\right) \sum_{i=1}^{N} \log P_B\left(a^i_1, \ldots, a^i_n\right) \quad (6)$$

D. Related work

S. Sriprayaya and S. Sinsomboonthong [17] proposed a comparison of efficacy of classification methods for chronic kidney disease. The case study of a hospital in India by choosing the method that is most similar how the tree decides artificial neural network methods supporting vector machines, methods, rules, logistic regression methods and the method of Naive Bay to measure group classification efficiency by using chronic kidney disease data from apollo hospital, India by dividing the data into a set of model creation and the test set of models in ratio 70 and 30 respectively by comparing the efficiency of the method of classification of chronic kidney disease patients by comparing the accuracy and mean square error. The group classification method that has the most effective classification is how the tree decides which gives the accuracy value is 100 % and the average square error is 0.0059.

D. Chakraborty et al. [19] developed a model for predicting the repeated treatment of schizophrenia patients using data mining techniques. From the database of Phra Si Mahapho Hospital Ubon Ratchathani Province, year 2007 - 2012 amount of 2,831 record by dividing the data into two classes as: class 0 is a group of patients who come to receive repeated treatment in 1 - 28 days, while class 1 is a group of patients who come to receive repeated treatment from 29 - 90 days. This model can help the doctor's treatment plan. The results showed that the data had outliers and there is an imbalance of classes in the data, with a number of classes more than another class. They solved the problem by screening out abnormal values by SVM and adjust the balance of the data by synthetic minority over- sampling technique (SMOTE)and developed the models with C4.5, Naïve Bayes and PART decision list. In addition, they also used 10- fold cross validation in dividing data into teaching data sets and test data sets and used the accuracy, sensitivity, and specificity to show the model's forecasting performance as well. The results of the forecasting efficiency of the model showed that the PART decision list can predict better than C4.5 and Naïve Bayes which has accuracy rate 92.98% Sensitivity rate 92.95% and specificity rate 93.02% respectively.

S. Sinsomboonthong [20] presented a prediction model of game addiction among children and adolescents in Bangkok. The data obtained from the study are also collected. Questionnaire for addiction to games of children and teens which consists of questions about gender, age, education level income per month using social media average duration per game played during Monday – Friday. The average duration per game played during Saturday - Sunday / holiday parental opinions about the addiction of the game to children obsession in the game loss of responsibility, loss in self-control, mood, behavior and time and game addiction from January to April 2017 In Lat Krabang, Min Buri and Bang Khen districts, 500 sets person total 1,500 sets collected additional 20%, equivalent to 300 sets, a total of 1,800 sets of comparison criteria, considering the accuracy Found that the percentage accuracy rate show 87.83%.

D. Vigneswari, et al. [21] focus to comparison of the effectiveness of predicting diabetes results using Naive Bay method is an algorithm that uses the principles of probability in screening each class. There are two classes. The results of the data analysis are evaluated to compare the effectiveness of predicting the diabetic results of the physical examination using the 6 methods of classification. By using the data from the analysis to compare which methods to classify data groups, which methods have the most accuracy. The average absolute error and the mean square error with

the least value gave the best result in predicting the best results. The average square error is a good evaluation gauge. Due to the average power error value, consisting of both variance and bias, the average absolute error value found that the accuracy 91.06 %.

III. IMPLEMENTATION

For the research on predictive behavior model of amphetamines based on Bayes Network Classifier, can be described as an overview as shown in Fig. 2.

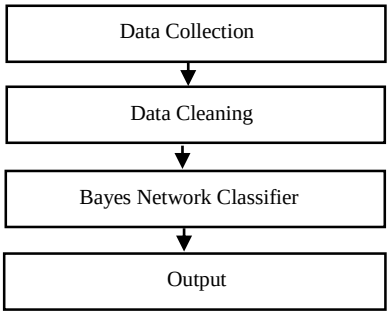


Fig. 2. shows an overview of the work system.

A. Data Collection

The first step of our work begin with data collection. This step is an important process for collecting the data set. The data set used for this work are about offenders arrested at Suan Phueng police station from 2016 – 2018. Our data set consist of eight attributes such as, gender, age, number of offenses, status, education, nationality, occupation, and class, which has 2 statuses, namely “take” and untake”. The amount of 1,598 items by collecting data from the offender's log file which has details as shown in Table I.

TABLE I. DESCRIPTION OF ATTRIBUTES

no	attribute	Description
1.	sex	Man, Woman
2.	age	The unit is the year starting from the age of 13 years - 78 years
3.	case	Number of offenses From 1 - 8 times
4.	education	Bachelor Degrees, HighVocCert, M1, M2, M3 M4, M5, M6, MS_degree, Not_studying, P1, P2, P3, P4, P5, P6, P7, Studying_M2 Studying_M3, Studying_M3, VocCert
5.	status	Divorce, Husband_died, Married, Single, Wife_died
6.	nationality	Cambodia, India, Karen, Laos, Mon, Myanmar No_nationality, Thai
7.	career	Student, animals_farm, animals_farm, animals_farm animals_farm, animals_farm, employee, Escape_workers Headman, Lawyer, Merchant, No_career, nurse, official, Pensioner, Pensioner, plant_farming, private_business Soldier, Student, teacher
8.	class	Take, untake

B. Data Cleaning

Data cleaning Is the process of updating information and correct the list of incorrect information which is the cornerstone of the database because it means imperfection, inaccuracy not related to other information, etc. Must therefore replace the update or removing these inaccurate data to ensure data quality, data cleaning happened because there is a lack of data canal which may be caused by an error of data recording, data transmission, or the definition of data stored differently. Therefore, has a high chance of being born " unclean information" in which information is used for decision making accuracy of the data is therefore important to avoid false conclusions. For example, redundant or missing information can result in incorrect or misleading statistics due to the large amount of data. The amount of information that is inconsistent is therefore very much as well. Cleaning is therefore the most important aspect of the data analysis process.

C. Classification

Classification at this stage, the researchers use. Bayes Network classifier for creating a model for predicting the behavior of users and non- users by using the data of 1,598 items offenders who were arrested in Suan Phueng Police Station, dividing the data into 10 sets and calculating the error value of 10 cycles. Selected out as test data and another 9 sets of data will be used as learning data as in Fig 3.

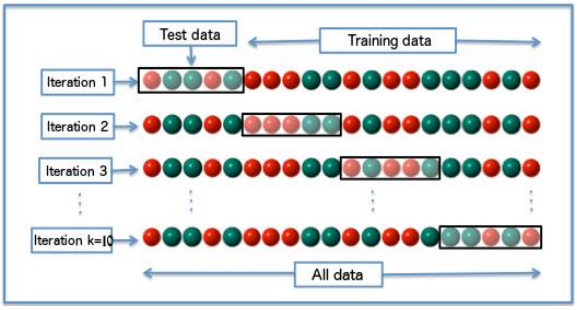


Fig. 3. Data Method 10-Fold cross-validation.

D. Model Performance Evaluation

Model performance evaluation by evaluating the accuracy of predictions for percentage accuracy TP Rate, FP Rate, Precision, Recall and F-Measure. The sample of results of our model for predicting the person who has the amphetamine behaviors show in TABLE II. In this step, we used 10-folds cross validation to do the performance which we show the results of step of cross validation in next section. That are the reason of why we used 10-folds cross validation for this work. Moreover, we show the line graph of the results cross validation to easy for understanding.

TABLE II. SAMPLE DATA SETS

no	sex	age	case	education	status	nationality	career	class
1.	man	18	1	P6	single	thai	enployee	take
2.	man	29	1	M3	single	thai	enployee	take
3.	man	30	1	M3	married	thai	plant_farming	take
4.	man	46	1	P6	single	thai	enployee	take
5.	man	48	2	M3	divorce	thai	enployee	take
6.	man	15	1	P6	single	karen	enployee	take
7.	man	38	1	P6	married	thai	enployee	untake
8.	man	13	1	P6	single	karen	enployee	untake

IV. EXPERIMENTAL AND RESULTS

In this experiment, we designed the experiment by using classifier in the Bayes group. The results showed that method Bayes Network classifier get the highest precision 81.53. Followed by is Naive Bayes classifier get accuracy 80.85% and Naive Bayes Updateable get accuracy 80.85% as shown in Table III.

TABLE III. ACCURACY VALUES WHEN ANALYZING 10 – 100 FOLDS

Fold	Accuracy rate
10	81.53
20	81.41
30	81.42
40	81.35
50	81.28
60	81.28
70	81.41
80	81.35
90	81.35
100	81.35

From Table III, the accuracy when analyzing 10 – 100 folds cross validation. The best performance for our predicting model found that 10 folds cross validates show the highest accuracy with 81.53 that illustrate as Fig. 4.

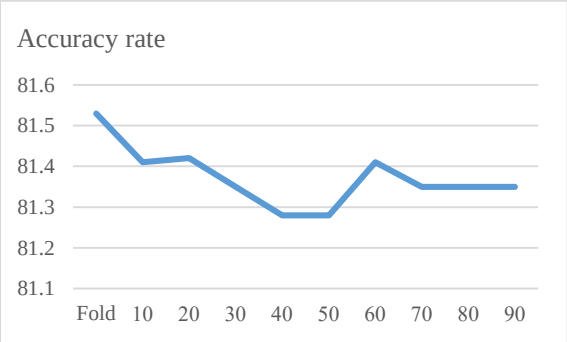


Fig. 4. Analyzing results of 10 – 100 Folds cross validation.

TABLE IV. ACCURACY RATE WITH 10 FOLDS CROSS VALIDATION

NO.	Methods	Accuracy rate
1.	Bayes Network Classifier	81.53
2.	Naive Bayes Classifier	80.85
3.	Naive Bayes Updateable	80.85

From Table 4, precision when analyzed as well Bayes Network Classifier, Naive Bayes Classifier, Bayes Naive Bayes Updateable 10 Fold style Found that Bayes Network classifier with the highest accuracy, with a value of 81.53, followed by Naive Bayes classifier and Naive Bayes Updateable Which has the same accuracy of 80.55 as the graph 2

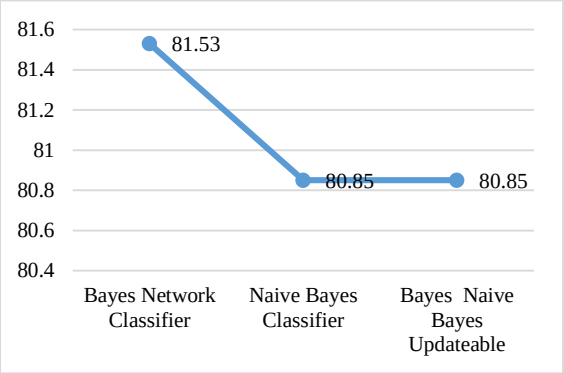


Fig. 5. Show the accuracy rate between a group of Bayes with 10 – folds cross validation.

TABLE V. MODEL TEST RESULTS WITH TECHNIQUES BAYES NETWORK CLASSIFIER 10 FOLDS STYLE

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.522	0.093	0.638	0.522	0.574	take
	0.907	0.478	0.858	0.907	0.882	untake
	0.815	0.386	0.806	0.815	0.809	
weighted Avg.	0.815	0.386	0.806	0.815	0.809	

From Table V, the results for testing model with Bayes Network classifier 10 folds when considered F-Measure It was found that the take class or the addicts had the value 0.574, while the untake class or the addicts were 0.882.

V. CONCLUSION

This paper proposed a prediction model for amphetamine behaviors based on a group of Bayes. From the experimental results The model for the behavior of amphetamine users on the basis of Bayes Network Classifier collected data about the perpetrators arrested at Suan Phueng Police Station since the year 2016 – 2018 to consist of 8 attributes Which are sex, age, number of offenses, status, education, nationality, occupation, and class, which have two classes, which are “take” and “untake”, amount of 1,598 items. The collecting data from the offender log files clean the data by updating the data and correcting the incorrect data list. We divided the data set into 10 equal sets and calculate 10 error values. Each calculation cycle, one set of 10 data sets will be selected as test data and the other 9 sets will be used as data for learning how to find Bayes Network classifier. The highest accuracy is 81.53%, followed by the Naive Bayes classifier and Naive Bayes Updateable, which has the same accuracy of 80.55%. The prediction results are known of the behavior of users and ways to precisely prevent and suppress the addicts. The experiment to assess the risks of adolescents in the use of drugs and make young people

continue to press pressure from their friends correctly. parents' efforts to protect their children through the use of additional home test kits to detect drug use or vaccination for high-risk children.

REFERENCE

[1] T. Yang, M. Chen and Y. S. Sun, "An Investigation of the Influence of Drug Addiction on Learning Behaviors in a Game- Based Learning Environment," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), Timisoara, 2017, pp. 158-162.

[2] M. Temurhan, R. Meijer, S. Choenni, M. van Ooyen-Houben, G. Cruts and M. van Laar, "Capture- Recapture Method for Estimating the Number of Problem Drug Users: The Case of the Netherlands," 2011 European Intelligence and Security Informatics Conference, Athens, 2011, pp. 46-51.

[3] R. Eshleman, D. Jha and R. Singh, "Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 849-854.

[4] N. Chumuang, P. Chansuek, M. Ketcham, A. Silsanpisut, T. Ganokratanaa and P. Selarat, "Analysis of X-ray for locating the weapon in the vehicle by using scale- invariant features transform," 2017 Fourth Asian Conference on Defence Technology - Japan (ACDT), Tokyo, 2017, pp. 1-6.

[5] Y. Nasser, M. E. Hassouni, A. Brahim, H. Toumi, E. Lespessailles and R. Jennane, "Diagnosis of osteoporosis disease from bone X- ray images with stacked sparse autoencoder and SVM classifier," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, 2017, pp. 1-5.

[6] Y. Luo and D. K. Mills, "Chitosan-Halloysite Hydrogel Drug Delivery System," 2016 32nd Southern Biomedical Engineering Conference (SBEC), Shreveport, LA, 2016, pp. 76-77.

[7] J. Zhao, T. Zhou and W. Dai, "Convolutional Neural Network-Based Joint Extraction of Adverse Drug Events," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1-5.

[8] Maneerat Theeraviwat and Nirat Imani. "Organization of drug prevention programs in secondary schools Karnchanaburi". Journal of Health Education 22 (83), (Sep-Dec 2542) : 37-51.

[9] H. S. Cho, D. Lee, S. Lim and M. Hahn, "SoTong: An Aware System of Relation Oriented Communication for Enhancing Family Relationship," 2008 International Symposium on Ubiquitous Virtual Reality, Gwangju, 2008, pp. 71-74.

[10] X. Huang and K. Li, "Influences of the Qigong upon the bodybuilding of the physically vulnerable undergraduates," Proceedings 2011 International Conference on Human Health and Biomedical Engineering, Jilin, 2011, pp. 736-739.

[11] J. L. Gastwirth, "The need for careful evaluation of epidemiological evidence in product liablility cases: a reexamination of Wells v. Ortho and Key Pharmaceuticals," in Law, Probability and Risk, vol. 2, no. 3, pp. 151-189, Sept. 2003.

[12] C. Liu, J. Ying, F. Han and M. Ruan, "Abnormal Human Activity Recognition using Bayes Classifier and Convolutional Neural Network," 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP), Shenzhen, 2018, pp. 33-37.

[13] B. Lerner, H. Guterman, I. Dinstein and Y. Romem, "A comparison of multilayer perceptron neural network and Bayes piecewise classifier for chromosome classification," Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Orlando, FL, USA, 1994, pp. 3472-3477 vol.6.

[14] M. Sangeetha and K. B.N., "Conditional Mutual Information based Attribute Weighting on General Bayesian Network Classifier," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 558-563.

[15] S. D. Thepade and M. M. Kalbhor, "Novel data mining based image classificatoin with Bayes, Tree, Rule, Lazy and Function Classifiers using fractional row mean of Cosine, Sine and Walsh column transformed images," 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, 2015, pp. 1-6.

[16] Nittaya Muangnak, Wannapa Pukdee and Thapani Hengsanunkun, " Classification students with learning disabilities using Naïve Bayes Classifier and Decision Tree," The 6th International Conference on Networked Computing and Advanced Information Management, Seoul, 2010, pp. 189-192.

[17] Surawachat, Sri Prayaya and Saichon Sinsomboonthong. 2560. "Comparison of efficiency of classification methods for chronic kidney disease: A case study of a hospital in India " Journal of Science and Technology, Year 25, Issue 5, September - October 2560. 839-853.

[18] Wirayut Mayuasiri, Jari Thongthong and Watinee Sookak. 2557. "Developing a model for predicting repeated treatment of schizophrenia patients using data mining techniques. " Journal of Science and Technology Academic conference Mahasarakham Research Time 10. 144-153.

[19] D. Chakraborty et al., "Prediction of Negative Symptoms of Schizophrenia from Objective Linguistic, Acoustic and Non-verbal Conversational Cues," 2018 International Conference on Cyberworlds (CW), Singapore, 2018, pp. 280-283.

[20] Saichol Sinsomboonthong. 2561. "A comparison of the efficiency of predicting game addiction among children and adolescents in Bangkok. "Journal of Science and Technology, Vol. 26, No. 3 May - June. 2561. 405-416.

[21] D. Vigneswari, N. K. Kumar, V. Ganesh Raj, A. Gugan and S. R. Vikash, "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 84-87.

The Development of a Model to Predict Marbling Score for fattening Kamphaeng Saen Beef Breed Using Data Mining

Watchara Ninphet[1,2] · Narumol Chumuang[3], Chairit Siladech[4] and [5]Mahasak Ketcham

[1]*Department of Industrial Technology Management, Faculty of Industrial Technology*
[2]*Department of Animal Science, Faculty of Science and Technology,*
[3]*Department of Digital Media Technology, Faculty of Industrial Technology*
[4]*Dep. of Educational Research and Evaluation, Faculty of Education,*
[5]*Dep. Of Management Information System, Faculty of Information Technology,*

[1,2,3,4]*Muban Chombueng Rajabhat University, Ratchaburi, Thailand*
[5]*King Mongkut's University of Technology North Bangkok, Bangkok Thailand.*

[1,2]way.n50@hotmail.com, [3]lecho20@hotmail.com, [4]siladech9@gmail.com and [5]mahasak.k@it.kmutnb.ac.th

***Abstract*— The purpose of this research was to create a model for beef marbling score in Kamphaeng Saen beef breed using data mining techniques. The data of the fattening cattle were transformed from Kamphaeng Saen Beef Co-operative Co., Ltd. during 2015-2019. For 5 years, 1 ,5 6 8 head, including the fattening period (month), the age of the cattle from the shedding and wear of the teeth The weight is alive when slaughter, the carcass weight into a fresh, carcass weight, cool carcass percentage and the marbling score in the meat. To create a forecast model by using data classification techniques with Decision Tree method with C4 .5 algorithm, then test the predictive model with percentage split method by dividing the data into training set 3 0 % and test set 7 0 % . It is found that the accuracy is 5 7 .1 9 4 9 % both because there are many factors related to the formation of fat in the muscles such as varieties, age, sex and feed, especially feed is the main factor affecting the accumulation of fat in the muscles. The growth of fat that is controlled by Stearoyl-CoA desaturase (SCD) shows that the prediction of the marbling score in fattening cow muscle with data mining techniques can be used as a means of forecasting without having to wait. The cattle were finished and assessed from the remains of the cattle.**

Keywords*— *Predict, Marbling Score, Kamphaeng Saen Beef Breed, Data Mining

I. INTRODUCTION

marbling score in the muscles (intramuscular fat) is the type of fat that is inserted in the muscle band clearly visible to the naked eye. Is a small line spread inside the muscles is a marble-like pattern called marbling. The amount of marbling score in this muscle bundle makes the meat softer. Helps to lubricate while chewing and swallowing the meat stimulates saliva secretion. Thus causing the feeling of being wet in the mouth. Therefore, the marbling score will result in the meat being wet and soft. At present, consumers have selected beef that has highly marbling score to achieve satisfaction in consumption. The level of marbling score in the cow's muscles is therefore necessary and important for consumers' purchasing decisions. The quality assessment of beef is used to evaluate the fat layer inserted in the muscle. With reports that affect taste [1] and acceptance from meat consumers [2] and the market that buys live cattle for privatization will determine the purchase price of fattening cattle The higher the fat level, the higher the price of the cattle that will be purchased.

Kamphaeng Saen Beef Cattle breed is a cow with 2 5 % native blood, 2 5 % Brahman, and 5 0 % Charolais with characteristics and qualifications that meet the standards of excellence of Kamphaeng Saen cattle breed, which the Kamphaengsaen beef cattle association set up. Kamphaeng Saen beef cattle is a result of 35 years of continuous research and development from Kasetsart University. Kamphaeng Saen Campus in collaboration with the Kamphaengsaen Beef Cattle Association. Until being acknowledged as 1 in 5 of the research and development results of Kasetsart University that farmers are most respected. At present, it is estimated that there are at least four hundred thousand Kamphaeng Saen beef cattle in the country. The main guideline for improving and developing beef cattle according to international principles. There are 2 approaches: selection and breeding plan. The selection of breeder cows with excellent characteristics gives the child a good appearance. Because of the important economic characteristics in beef cattle can be transmitted to the offspring. Therefore collected data of cattle breeder, Kamphaeng Saen breed, growth rate, food consumption, including beef quality. In order to use the information obtained to analyze the relationship for the knowledge that has been disseminated to promote and encourage farmers to use as the criteria for choosing to use cattle breeders to improve the production efficiency of the farm [3]

Fig. 1. Kamphaeng Saen Beef Cattle Breed [3]

The beef marbling score levels is based on the expert's visual assessment of the specimens [4] and has developed image processing techniques to help save time. [5] However, this method has not yet been found. Predict the amount of beef marbling score in advance, which has not yet fattened the cows and evaluated from the carcasses.

This research The researcher focused on the development of a prediction model for marbling scores in Kamphaeng Saen beef breed with data mining techniques. And if the relationship of the cattle breeders and the economic characteristics of the calves.

II. LITERATURE REVIEW

A. Marbling

The marbling scores means the fat that is inserted in the muscle bundle, visible with the naked eye clearly. The fat that is inserted in the muscle bundle indicates the texture of the meat. Which fats will help stimulate saliva secretion, resulting in a feeling of wetness in the mouth [6].

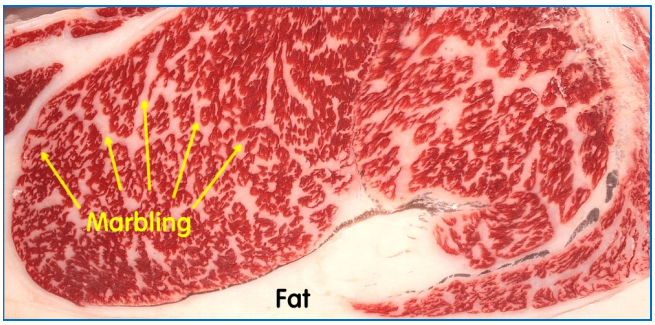


Fig. 2. Beef Marbling [5]

The amount of marbling scores in the muscles, more or less depends on the genetic influence of the animal. And it is well known that Japanese Black cows are more capable of accumulating fat in the muscles than other breeds [6] and tropical cows have less fat deposits than that of the European Cattle. Charolais beef cattle have higher fat content in muscle than American Brahman beef cattle. Though raised and fed as well. [7] Angus beef cows have a higher fat content in the muscle than the American Brahman. Under raising with the same food.In addition, the age of mature cows will allow the cows to accumulate more fat. When fully grown, the fat can be inserted into the meat up to 40 times after birth. [8]

Beef marbling score level in cow muscle evaluated from the area of the tenderloin (Longgissimus dorsi muscle) between the ribs 12 and 13, with the level of beef marbling score according to different associations or countries with the minimum beef marbling score, level 12 is the level with the most beef marbling score. In the United States of America, used at 6 levels. In Thailand, the standard beef marbling score is used at 5 levels, showing the sample image of each marbling score (Fig. 3.).

B. C4.5 Decission Tree

Decision Tree is a process of data mining that uses tree structures to help decide various aspects of work. Whether in business or other aspects Usually consists of rules in the form "If the condition is the result" such as "If Income = High and Married = No THEN Risk = Poor." The first node is the root of the tree. Each node represents an attribute. Each branch shows the result of the test and the leaf node displays the class as shown in Fig. 4.

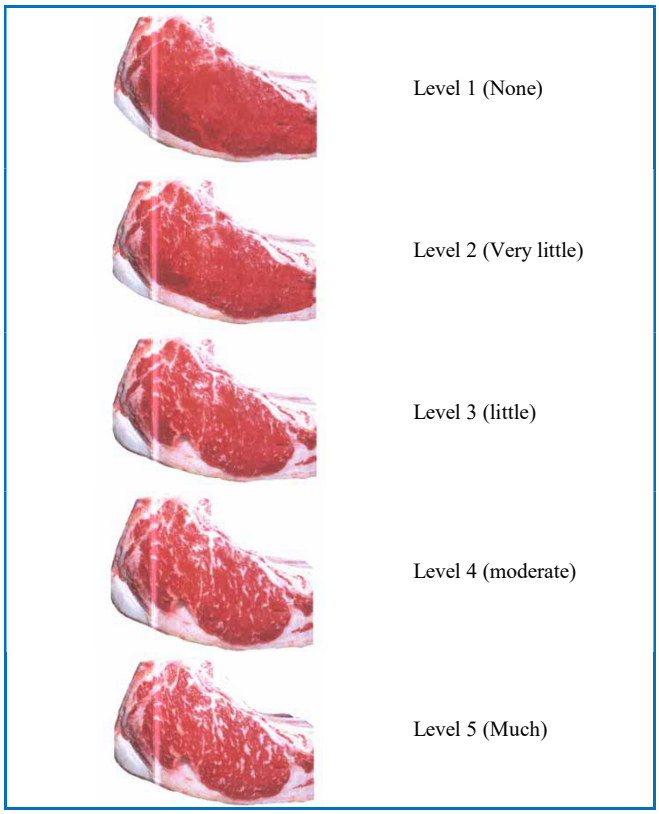


Fig. 3. Thai beef marbling score 5 levels standard [4]

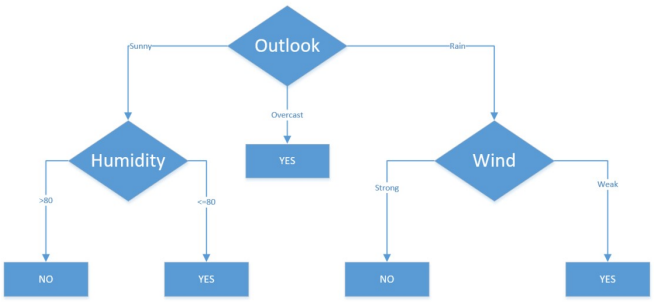


Fig. 4. Decision tree generated by C4.5 [9]

C4 . 5 Algorithm developed by Ross Quinlan [1 0] developed from ID3 algorithm [1 1] which is an algorithm used to create decision trees. By applying the principles of news theory The measured values will be used to decide which variables to use in dividing the data. The method of determining the structure of the decision tree will select the data according to the order of the indicators or the highest gain as the initial data. And the next data that has a descending order, respectively. For example, considering the two classes of data is P and N, where the number of samples in class P is p, and the number of samples in class N is n. Specifying that the sample must use the number of bits to separate class P and N by definition according to equation 1

$$I(p,n) = -\frac{p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right) \quad (1)$$

The estimate of the data (Entropy) is the value that is separated by the use of the characteristic A, which defines A is the routine that divides S into {S1 , S2 ,, Sv}. Class N, number n1, as equation 2

$$E(A) = \sum_{i=1}^{v} \frac{p_1+n_1}{p+n} I(p_1,n_1) \quad (2)$$

Therefore, the data gain (data gain) obtained from the separation of data with the characteristic A will be as follows.

$$Gain(A) = I(p,n) - E(A) \quad (3)$$

C4 . 5 Algorithm will have additional parts from the important ID3 algorithm as follows:

1 . Can use both data with continuous and non-continuous (Discrete) features. In the continuous data section, algorithm C4 .5 will create a threshold and separate that feature. Into 2 parts, which are more and less valuable And equal to the value used to create the starting point

2 . Can be used with training data that does not have a feature value Which will mark that feature as "?" And do not use that value to calculate the data predictions (Entropy)

3 . Can be used with values that are abnormal or damaged.

4 . able to customize the tree to make decisions (Pruning Trees) while creating

C. literature review

The beef marbling score is evaluated from the area of the tenderloin. (Longgissimus dorsi muscle) between the ribs 12 and 1 3 , the national agricultural and food products standards. Determine the amount beef marbling score in the muscle into 5 levels. The beef marbling score level rating is based on the expert eye assessment from the sample image of each level of the marbling score. [4]

Forecasting the level of fat in the beef by the amount of fat in the skin (subcutaneous (s.c.) adipose tissue lipids) in 79 male Angus cows found that the level of fat in the beef was not related to the level of fat in the meat. [12]

There is research that uses image processing techniques to increase the efficiency of lipid evaluations, helping to save more time. [5] There is also a study using Visible infrared (VIS) to assess fat from the surface. Touch the top of the cow's meat [5]

Application of decision tree to determine fat level in beef according to 12 Japanese benchmarks using hyperspectral imaging, it was found at 440 nm. wavelength has an accuracy of 99.92%, which is an efficient and fast technique. [13]

The study of the method of estimating fat level in marbling cows using dynamic ultrasound live cow photos .It consists of four processes, namely surface analysis, time series, dynamic property extraction, principal component analysis and estimation of fat levels in cattle by artificial neural network technique. Found to be highly accurate. The correlation coefficient between actual fat level and fat level is evaluated by using the dynamic image feature by the one-off method, r = 0.75 (P <; 0.01) and the average estimation error is 1.09. These results suggest that the dynamic image properties extracted from the dynamic ultrasound image have the potential to accurately estimate fat levels inserted in the beef. [14]

The study of methods for estimating the level of fat in meat by determining the relationship between the impedance of the bio-dielectric properties of the sirloin, the amount of raw fat in the boliolose and the inserted fat level in beef. The results show that the relationship (r = 0.61, RSME = ± 20.6, P <; 0.01) shows that the analysis of biological resistance has a potential impact on BMS evaluation which is useful for the beef production management industry. [15]

From this method, there has not been a way to predict the level of fat in the cattle in advance without having to fatten the cows and assess the carcass.

III. METHODOLOGY

The research process consists of 4 steps asfollows

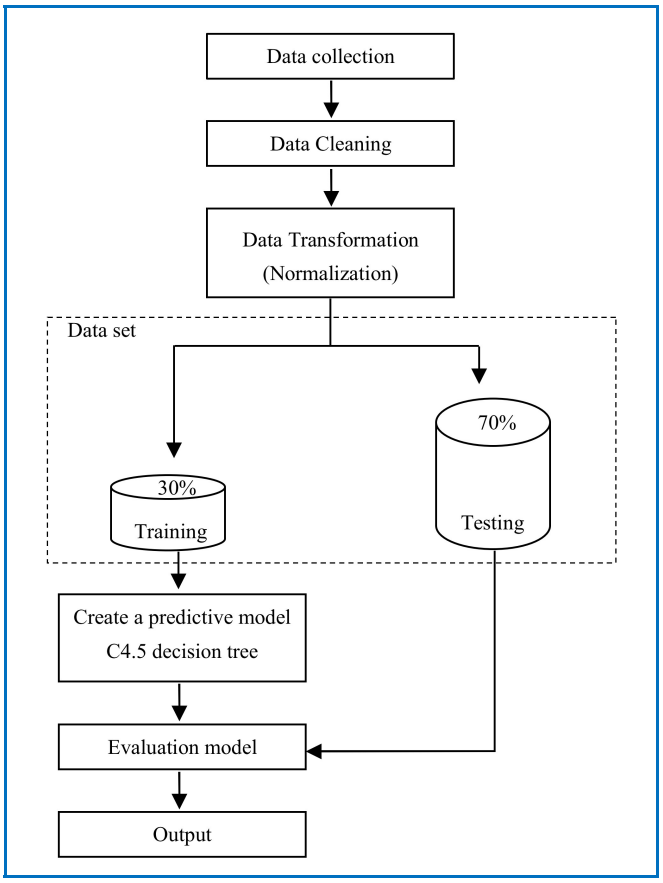


Fig. 5. Flow Chart and Proposed Set-Up.

A. Data collection

The data of the fattening cattle that were transformed from the Kamphaeng Saen Beef Co-operative Co., Ltd. during the year 2 0 1 5 -2 0 1 9 for 5 years, totaling 2 ,0 8 9 characters, including the fattening period (month), the age of the cattle from the shedding and wear of the teeth. The weight is alive when slaughter weight, fresh, carcass weight cool carcass percentage and the marbling score in the meat

B. Data Cleaning

Make the information accurate delete or repair missing data and correct the information with errors

C. Data Transformation

Normalization by Max-Min normalization method of aging, live carcass weight carcass weight

age	live carcass weight
9	498
7	550
12	612

→

age	live carcass weight
0.27	0.32
0.19	0.42
0.38	0.54

Fig. 6. Data Transformation

TABLE 1 RESEARCH ATTRIBUTE

No.	attribute	meaning	Description
1.	age	Fattening period (month)	Max-Min normalization method in the range 0-1
2.	teeth	Age of the cattle from the shedding and wear of the teeth	M = all milk teeth (Cattle aged not more than 2 years) N = 1 pair of permanent teeth (cattle last about 2 years) O = 2 pairs of permanent teeth (cattle last about 3 years) P = 3 pairs of permanent teeth (cattle last about 4 years)
3.	live	Life weight when slaughter (kg)	Max-Min normalization method in the range 0-1
4.	fresh	Fresh carcass weight (kg)	Max-Min normalization method in the range 0-1
5.	class	Beef marbling score (1-5)	A = 1 , B = 2 , C = 3 , D = 4 , E = 5

TABLE 2 SAMPLE DATA SET THAT HAS BEEN CLEANED UP

age	teeth	live	fresh	class
0.27	M	0.32	0.37	A
0.15	M	0.22	0.24	A
0.15	M	0.25	0.27	A
0.31	N	0.33	0.24	B
0.31	N	0.28	0.22	B
0.31	M	0.13	0.11	B
0.19	N	0.34	0.30	C
0.19	O	0.31	0.28	C
0.23	N	0.24	0.27	C
0.23	M	0.26	0.21	D
0.31	N	0.22	0.17	D
0.19	N	0.30	0.27	D
0.19	O	0.41	0.43	E
0.23	N	0.39	0.42	E
0.23	N	0.39	0.39	E

D. Create a forecast model

Create a forecast model using Decission Tree C4.5 The method of randomly dividing data with percentage split is to divide the learning set data and test sets by random methods. By specifying the size of the test set data as a percentage, for example, if dividing the data into 70.00%, it means choosing a random data out of 70 sets for teaching. And using the remaining 30 sets of data in the test. For this research, the researcher determined the size of the test set data to 30.00 percent.

E. Performance measurement

Evaluate the accuracy of predictions by using percentage accuracy (Accuracy Rate) and the root mean squared error (RMSE) . If the closer to 0 means will be more accurate

$$\text{Accuracy Rate} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

by

1 . True Positive (TP) is the amount of data that is predicted to be a class that is currently interested.

2 . True Negative (TN) is the amount of data that is predicted to be a class that is not interested.

3 . The False Positive (FP) is the amount of data that is predicted to be a class that is currently interested.

4 . False Negative (FN) is the amount of information that is wrongly predicted as a class that is not interested.

IV. Experiment Result and Discussion

The results showed that when dividing the data into 30% training set and 70% testing set for the highest accuracy rate with accuracy rate 5 7 . 1 9 4 9 and Root Meansqure Error ; RMSE = 0.3423 as TABLE 3 and Fig.7.

TABLE 3 ACCURACY RATE AND AND ROOT MEANSQURE ERROR

Percentage Split	Accuracy Rate	Root Meansqure Error ; RMSE
10	55.7052	0.3614
20	54.0670	0.3682
30	57.1949	0.3423
40	52.3911	0.3670
50	51.4031	0.3719
60	51.8341	0.3690
70	53.6170	0.3577

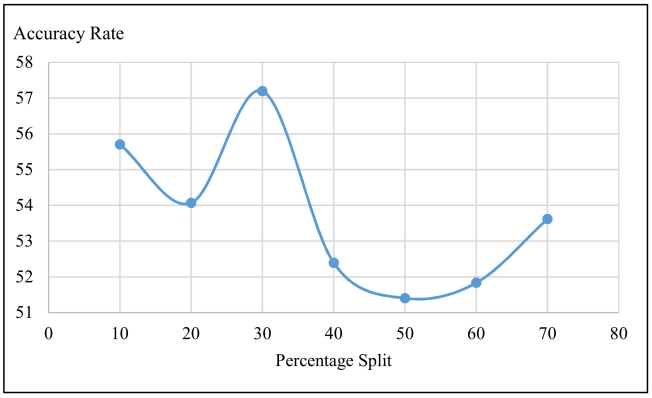


Fig. 7. Accuracy Rate

Forecasting by using data classification techniques with Decision Tree method by C4.5 method, providing 57.1949% accuracy. There are many factors related to the formation of fat in cattle, such as age, breed, sex and feed, which is a factor. Principles that affect the accumulation of fat in the muscles [1], [17], [18]. The expansion of fat is controlled by Stearoyl-CoA desaturase (SCD) [1 9] . This study does not use feed factors. Because Kamphaeng Saen Beef Co-operative Co., Ltd. does not record the type of feed used

therefore cannot be used to predict resulting in low accuracy. But this technique will help to predict the beef marbling score in the fattening beef in advance. Without having to wait for the fattening of the cattle to transform and incubate the carcasses, which takes at least 8-10 months and also allows cattle farmers to use as a consideration for the selection of cattle breeders in the farm to be desired The fattening beef with a beef marbling score as planned

V. Conclusion

Forecasting By using data classification techniques with Decision Tree method by C4.5 method, providing 57.1949% accuracy. There are many factors related to fat formation in muscles such as age, breed, sex and feed which feed is the main factor affecting the accumulation of fat in the muscles [17], [18]. The expansion of fat is controlled by Stearoyl-CoA desaturase (SCD) [19] . beef marbling score in fattening beef with tech The statistics, data mining can be used as a guide for forecasting without waiting for fattening cattle and finished cattle carcass evaluation.

Suggestions to make predictions more accurate feed information should be added to accompany the forecasting as well

Acknowledgment

Thank you to the Kamphaeng Saen Beef Cooperative Ltd. for the support of this research. Help the research to proceed quickly and smoothly until accomplished well.

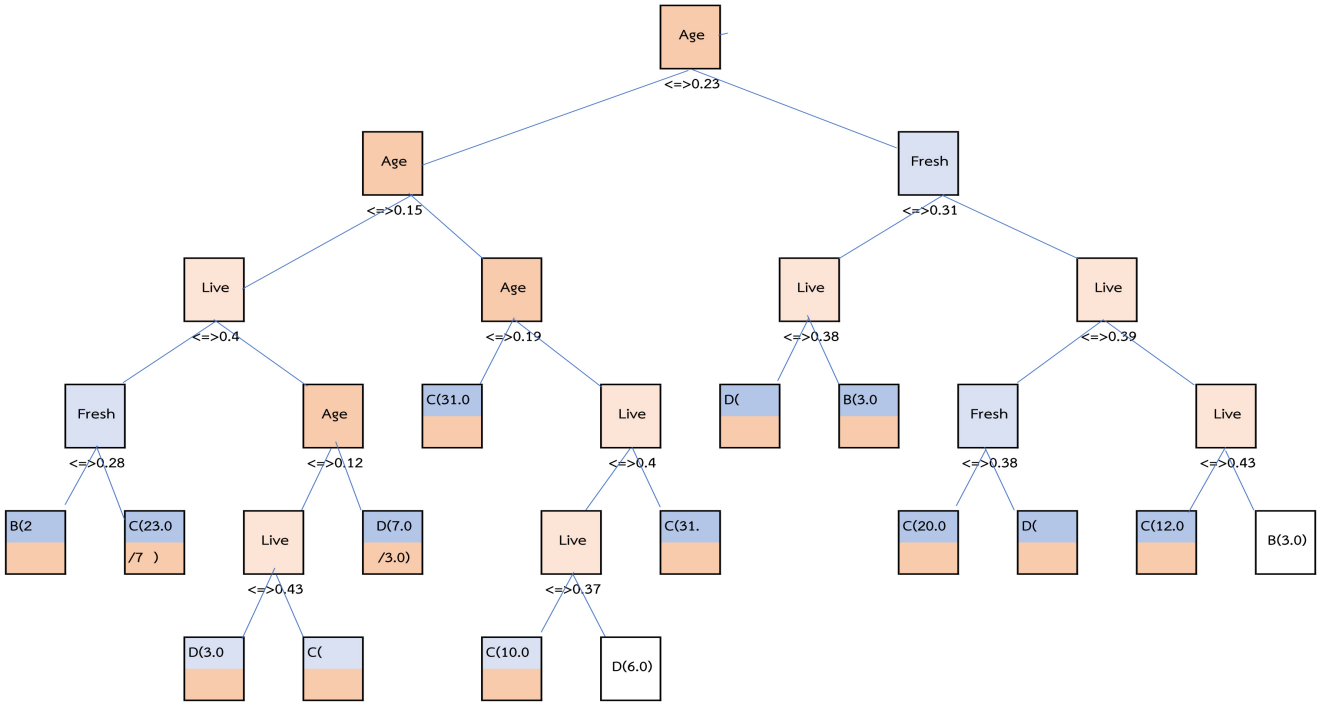


Fig. 8. Some of the data mining techniques are using the decision tree method.

References

[1] Calkins, C. R. , and J. M. Hodgen. "A fresg look at meat flavor" 2007. Meat Sci. 77. pp. 63-80

[2] Platter, W. J., J. D. Tatum, K. E. Belk, P. L. Chapman, J. A. Scanga, and G. C Smith. "Relationships of consumer sensory ratings, marbling score, and shear force value to consumer acceptance of beef strip loin steaks." 2003, J. Anim. Sci. 81 pp. 2741-2750.

[3] Prattana Prucsasri. "Interesting information about beef cattle: set 3, breeding and selection of cows." 2015 Neon Book Media, Bangkok.

[4] National Bureau of Agricultural Commodity and Food Standards. "Thai Agriculture Commodity and Food Standard ; TACFS 60001-2004". 2004. pp.1-22.

[5] Ziadi, A., X.Maldague, and L. Saucier. "Image analysis in computer vision: A high level means for Non-Destructive evaluation of the marbling in the beef meat." 2010. 10.21611/qirt.2010.149.

[6] Albrecht, E., Teuscher, F., Ender, K., and J. Wegner. "Growth- and breed-related changes of marbling characteristics in cattle." 2006. J. Anim. Sci. 84: pp. 1067-1075.

[7] Yamada, T. and N. Nakanishi. "Effects of the roughage/ concentrate ratio on the expression of angiogenic growth factors in adipose tissue of fattening Wagyu steers." 2012. Meat Sci. 90: pp. 807-813

[8] McKeith, F. K., J. W. Savell, G. C. Smith, T. R. Dutson, and Z. L. Carpenter. "Tenderness of major muscles from three breed-types of cattle at different times-on-feed." 1985. Meat Sci. 13: pp. 151-166.

[9] Sefik Serengil (May 13, 2018) "A Step By Step C4.5 Decision Tree Example" [Online]. Available: http://sefiks.com/2018/05/13/a-step-by-step-c4-5-decision-tree-example/[Accessed: Sep. 27, 2019].

[10] J. R. Quinlan. "C4.5: program for machine learning, California : Morgan Kaufmann," 1993.

[11] J. R. Qui "ductio of Decisio Trees " Machine Learning, 1986. vol. 1, no. 1, pp. 81-106,

[12] Victor V. Carvalho, Stephen B. Smith, Slip points of subcutaneous adipose tissue lipids do not predict beef marbling score or percent intramuscular lipid, Meat Science, Volume 139, 2018, Pages 201-206,

[13] Lía Velásquez, J.P. Cruz-Tirado, Raúl Siche, Roberto Quevedo, An application based on the decision tree to classify the marbling

of beef by hyperspectral imaging, Meat Science, Volume 133, 2017, Pages 43-50,

[14] O. Fukuda, N. Nabeoka and T. Miyajima, "Estimation of marbling score in live cattle based on dynamic ultrasound image using a neural network," 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP), Vienna, 2012, pp. 276-279.

[15] O. Fukuda, D. Hashimoto and I. Ahmed, "Bioelectrical impedance analysis for estimating marbling score of live beef cattle in Japan," 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bali, 2016, pp. 1508-1512.

[16] Stephen, B., K. David, Y. Chung, B. Chang, K. Tume, and M. Zembayashi. “Adiposity, fatty acid composition and delta-9-desaturase acitivity during growth in beef cattle.” 2006. J. Anim. Sci. 77: pp. 478-486

[17] Hood, R. L. “Relationships among growth, adipose cell size, and lipid metabolism in ruminant adipose tissue” 1982. Fed. Proc.41 pp. 2555-2561

[18] Chilliard, Y., and J. Robelin. “Activity of lipoprotein lipas in different adipose deposits and its relation to adipocyte size in the cow during fattening or early lactation” 1985. Reprod. Nutr. Dev. 23 pp.287-293

[19] Martin, G. S., D. K. Lunt, K. G. Britain, and S. B. Smith. “Postnatal development ofstearoyl coenzyme A desaturase gene expression and adiposity in bovine subcutaneous adipose tissue” 1999. J. Anim. Sci. 77 pp.630-636

Parameterized Minutiae Analysis for Generating Secured Fingerprint Template

Md. Mijanur Rahman
Dept. of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh-2224, Bangladesh
Email: mijanjkkniu@gmail.com

Tanjarul Islam Mishu
Dept. of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh-2224, Bangladesh
Email: tanjarul26@gmail.com

***Abstract*— This paper presents a parameterized minutia-based approach for generating secured templates from fingerprint images. The research has a great contribution in the field of security of the fingerprint template database. The fingerprint minutia features and their related parameters have been analyzed and hence, proposed a method that can hide fingerprint features by adding chaffs or fake minutiae and changing real minutia information to generate secured templates or vaults. The proposed method achieved highest accuracy for verification and has diversity that creates dissimilar vaults to resist correlation attacks.**

***Keywords*— Biometric features, correlation attacks, fingerprint minutiae, fuzzy vaults, secured template**

I. Introduction

Fingerprint has some unique biometric features that can be used in the authentication system [1]. Past few decades, it has been studied to improve the security of the biometric-based system [2-5] and to analyze severe attacks on fingerprint authentication system [6-7]. A lot of studies have been done to solve all the vulnerabilities of the model [8] from its different points and hence, the template is a great matter of concern for fingerprint security. Several methods proposed [9-11] to stop getting real or closely fingerprint image from a template, such as fuzzy-vault techniques [12]; but they are vulnerable to correlation attacks [13]. When same fuzzy vault technique is applied to a bulk of fingerprint images, two different generated vaults (templates) can be used to get the real minutiae information (Fig. 1) [14] and hence, the lack of security. Clancy [15] and Uludag [16] proposed a fuzzy fingerprint vault, which is not realistic [17] because of its limitation. The method needs pre-alignment before the operation.

In this paper, the fingerprint minutiae features and their related parameters have been analyzed and proposed a method that can hide fingerprint features by adding chaffs (fake minutiae) and changing real minutiae information to generate secure template or vault. Since the template pattern is different for every new fingerprint, there is no way to get vault information and it resists correlation attacks.

II. Fingerprint Minutiae Analysis

A fingerprint feature, known as minutia, m_i is defined by x-, y-coordinates, angle and minutiae type: $m_i = (x_i, y_i, \theta_i, t)$ [$i=1,2,3\ldots\ldots, n$ (number of minutiae in the vault)]. For each minutia m_i, we draw a circle centered on m_i with r radius (the value could be on the range of 15 to 40). Then we choose $8*k$ equidistance points (EPs) on the circumference of the circle, where $k=1,2,3,\ldots\ldots, n$. The first point is marked with the angle of minutia. The others are chosen equidistantly in clockwise (see Fig.2, where k=2). The standard range of EPs would be 8 to 64. The angular distance (AD) between each EP on the circumference is calculated as follows:

$$AD = \frac{2\pi}{No.of\,EPs\ (8*k)} = \frac{2\pi}{16} = \frac{\pi}{8} \qquad (1)$$

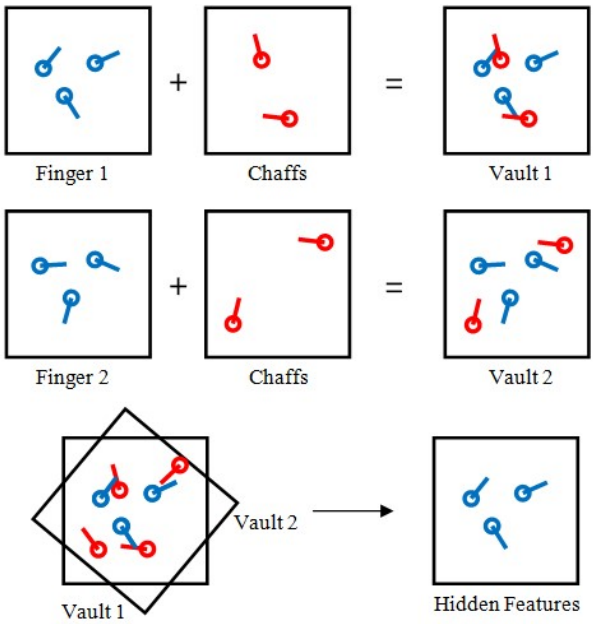


Fig. 1. Illustrating the correlation attacks.

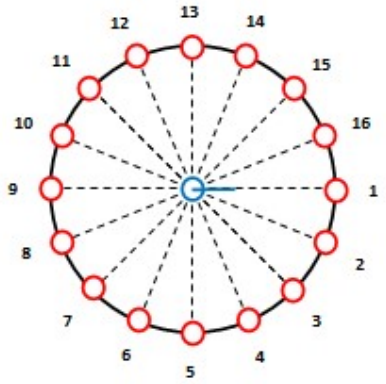


Fig. 2. Drawing circle and placing EPs.

For the minutia m_i, we randomly choose one or more points (for chaff minutiae) from the $8*k$ EPs set on the circumference, where the chaff or false minutia will be added; these points are known as selected chaff points

(SCPs) or choosing chaff points (CCPs), as shown in Fig.3. The chaff minutia is rotated by an angle (α) and the chaff angle value is determined by the following equation:

$$\alpha = \begin{cases} \theta + \delta - 2\pi & ;(\theta + \delta) \geq 2\pi \\ \theta + \delta & ;else \end{cases} \quad (2)$$

where θ is the angle of real minutiae and δ is the angle offset between 0 to 2π, but the standard range is 0.5 to 5.78.

Then we choose another point (for real minutia) from the *8k* EPs set, except the SCPs; this point is called real minutia moving point (RMMP) and we move the real minutia m_i, which now place in the center of the circle, to the RMMP, as shown in Fig.4. The real minutia m_i' is then rotated by an angle (β) and the angle value is determined by:

$$\beta = \begin{cases} \theta + (2\pi - \delta) - 2\pi & ;(\theta + (2\pi - \delta)) \geq 2\pi \\ \theta + (2\pi - \delta) & ;else \end{cases} \quad (3)$$

The chaffs can be added for either termination or bifurcation (Fig.5) or for both type minutiae. If the fingerprint contains very few bifurcations, then chaff minutiae should added for both type. As more chaff added, more the computational complexity and more secure the template. The pattern of template or vault varies depending on the values of above mentioned parameters, such as (radius), (angle of real minutia), (angle offset), SCP (selected chaff point), RMMP (real minutia moving point), and *t* (minutia type); and hence, so called parameterized minutiae templates.

III. Proposed Method

The proposed approach includes the methods of adding chaffs to generate secured template and removing these chaffs from template for verification. The overall scenario of the proposed technique is shown in Fig. 6.

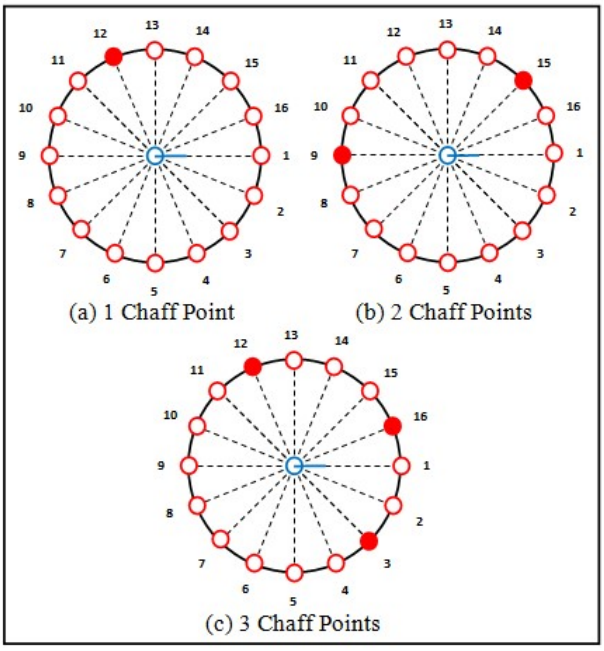


Fig. 3. Randomly chosen SCPs from 8*k EPs.

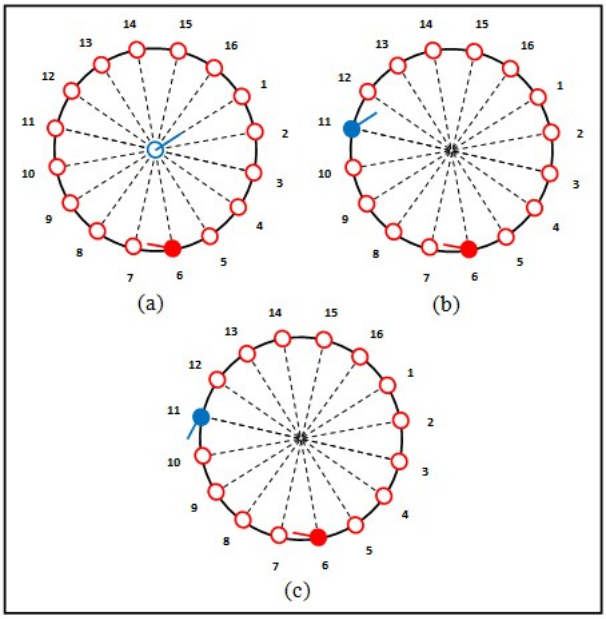


Fig. 4. Selecting the SCPs, RMMP and placing minutiae: (a) Adding a chaff minutia at SCP 6, (b) Moving the real minutia to RMMP 11, and (c) Rotating the moved real minutia at RMMP 11.

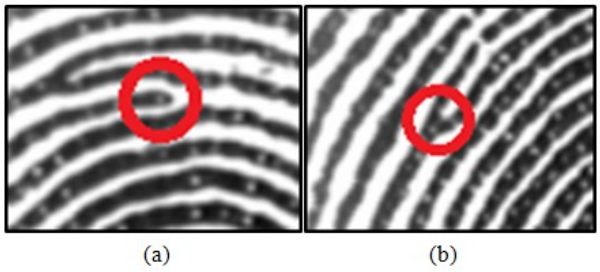


Fig. 5. Minutia type: (a) Termination or ridge ending; and (b) Bifurcation.

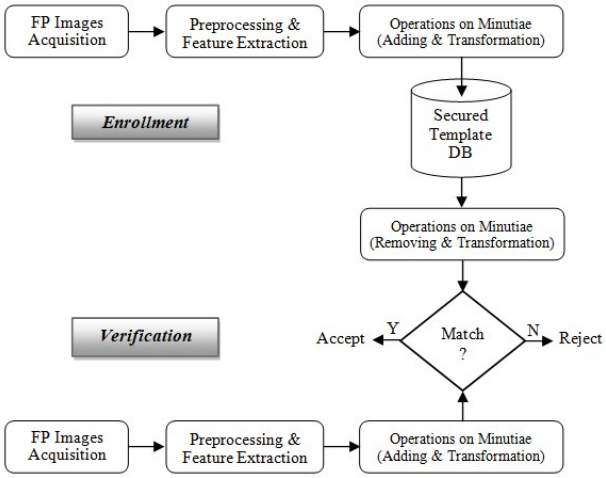


Fig. 6. Overall scenario of the proposed system.

A. Adding Chaffs to Template

A minutia feature m_1 is detected from the fingerprint image. A circle is drawn with the center on m_1 and the radius *r*. There are 16 (8*2) EPs are marked on the circumference of the circle. For illustrating enrollment

technique (Fig. 7), the SCP 6 is chosen for adding chaff minutia. Assume that angle θ is 1.9 and angle offset δ is 2.5. Then the angle (α) of chaff minutia m_2 is 4.4 (1.9+2.5) using equation (2). Now choose the RMMP 11 for moving the real minutia m_1. The angle (β) of moved minutia is 5.6832 (1.9 + 6.2832-2.5) using the equation (3). The same process is remained for both termination and bifurcation minutiae. The moved real minutia m_1 and the added chaff minutia m_2 both stored in the new generated template and hence, the template is secured. Fig. 8 shows the secure template generation operation done on the image "103_1.tif" from FVC 2002 DB1 [18] and each step of the proposed approach is summarized in the following.

Algorithm for generating secure template:

Step 1.	Draw a circle with center on the minutia.
Step 2.	Choose 8**k* EPs on the circumference of the circle.
Step 3.	Choose a SCP from 8**k* EPs for the chaff minutia and add angle offset (δ) to real minutia's angle (θ) to determine chaff minutia's angle (α) and then rotate chaff.
Step 4.	Choose a RMMP and move the real minutia on the circumference except previously chosen SCP for chaff minutia.
Step 5.	Calculate the angle offset (2π - δ) for the moved real minutia and add to its angle to determine moved real minutia's angle (β) and finally, rotate the real minutia with angle β.

B. *Removing Chaffs from Template for Verification*

Assume that only two minutia m_1 and m_2 in the secured template (Fig. 10). As discuss earlier, m_1 is a real and m_2 is a chaff or fake minutia. But in this phase we don't know which one is real or fake. We take all the minutiae and perform operations on each minutia. First we take m_1 minutia and remove its angle offset, and the angle (γ) of the m_1 is determined by the following equation.

$$\alpha = \begin{cases} \theta + \delta & ; \theta < (2\pi - \delta) \\ \theta + \delta - 2\pi & ; else \end{cases} \quad (4)$$

The angle (θ), offset (δ) and radius (r) have been discussed earlier. Then we draw a circle with r radius centered on the minutia and mark the EPs. Now we move the minutia from the center of circle to a point on the circumference, called real minutia returning point (RMRP). RMMP and RMRP are the two endpoints of a diameter of the circle (see Fig. 9).

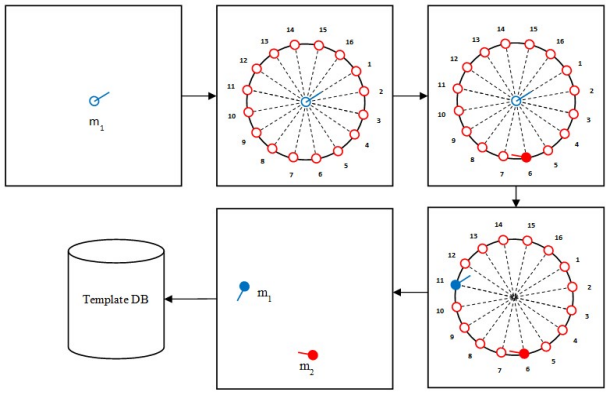


Fig. 7. Adding chaff and minutia transformation to generate secure template.

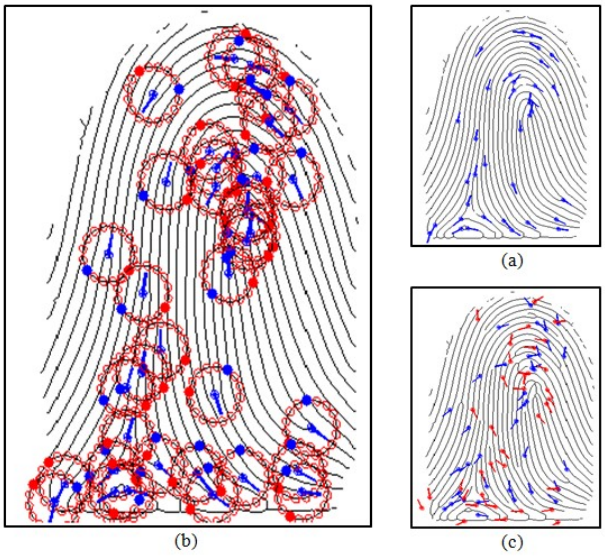


Fig. 8. (a) Real minutiae detected from the fingerprint image, (b) Adding chaffs (red dots) and moving real minutiae (blue dots), and (c) Showing chaffs (red) and real minutiae (blue) stored in the template.

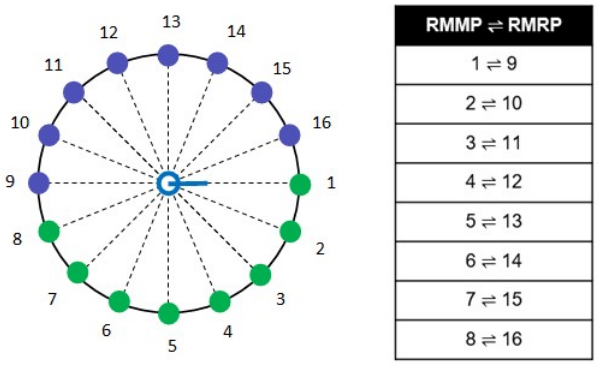


Fig. 9. RMMP and RMRP are two endpoints of a diameter of same circle.

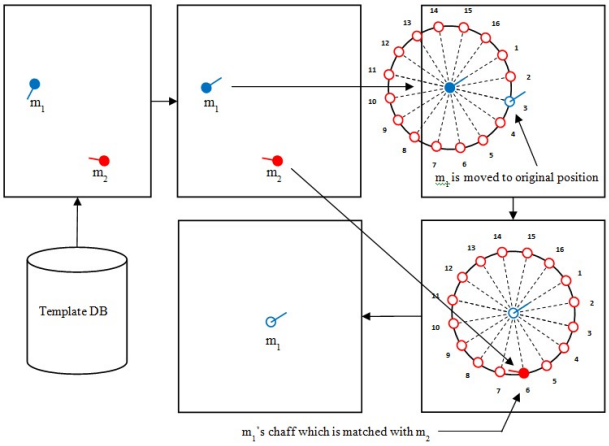


Fig. 10. Removing chaff from the template and minutia transformation for verification.

After returning m_1 to RMRP (m_1's original position), again we draw a circle centered on the moved m_1 and mark the EPs. Then we generate a chaff minutia of m_1 at the same SCP and determine the chaff's angle using by the equation (3). Finally, we search for minutia in the template similar to

m_1's chaff minutia. If it is matched with any minutia, then the minutia in the template is a fake or chaff minutia and m_1 is a real minutia. Here m_1's chaff is matched with the minutia m_2 at the point 6, and hence m_2 is a fake minutia and it is removed from the template, as shown in Fig.10. If it is not matched with any minutia, m_1 is a fake minutia and m_1 is removed from the template to regenerate the original fingerprint template. Each step of the proposed approach for verification is summarized in the following.

Algorithm for regenerating original template:

Step 1. Remove the angle offset from the minutia's angle.
Step 2. Draw a circle with same 8**k* EPs and radius *r* centered on the minutia.
Step 3. Move the minutia to its original point by choosing the RMRP.
Step 4. Draw again a circle centered on the moved position (RMRP) of minutia and mark the EPs.
Step 5. Generate a chaff minutia with the same SCP.
Step 6. If the new generated chaff is matched with any minutia in template, then the minutia is a real minutia; otherwise it is a chaff minutia.
Step 7. Remove all the chaff minutiae from the template.

IV. Experimental Results

In this research work, the FVC 2002 DB1 set B database [18] has been used to evaluate the performance of the proposed method. The database contains 8 (eight) impressions of each of 10 (ten) fingers; 5 (five) impressions were used for generating secured template database and other 3 (three) were used for matching at the verification phase. The research analyzed the parameterized minutia features to implement the method; and different values of the parameters: radius of circle (r), number of EPs, number of CCPs, angle offset (δ), and termination (*Ter*) or/and bifurcation (*Bif*) minutiae were tested in the experiment. The method was implemented using MATLAB tools in this experiment. The performance of the proposed approach is shown in Tab.1. From the table, it is seen that the proposed method maintains its performance compared to existing methods with false acceptance rate (FAR) 0.06% and genuine acceptance rate (GAR) 96.67% , i.e., it does not degrade the performance.

TABLE I. Performance of Proposed Method

Existing Method			Proposed Method		
Minutia type	***FAR (%)***	***GAR (%)***	***Minutia type***	***FAR (%)***	***GAR (%)***
Not defined	0.06	96.67	Bifurcation	0.06	96.67
			Termination & Bifurcation	0.06	96.67

Tab.2 shows the average number of chaffs added per template in the proposed technique. It also shows the efficiency of removing chaff minutia from each template at the verification phase and achieves the highest accuracy.

TABLE II. Efficiency of Chaff Removing Process

Minutia Type	Avg. No. of Chaff Added Per Template	Chaff Removing Efficiency (%)
Bifurcation	19	100
Termination & Bifurcation	44	100

Tab. 3 shows the membership values of similarity (fuzzy values within the range of 0 to 1) among secured templates or vaults. For illustrating the similarity, Vault 1, 2, and 3 are generated by proposed method and General Template is generated by existing method from the same fingerprint image 103_1.tif with different parameters in this experiment. The higher fuzzy value represents more similarity and lower value represents more dissimilarity between two vaults. From the table, it is seen that the same fingerprint's template is varied vault-to-vault with different parameters and hence, it resists the correlation attacks.

TABLE III. Similarity Comparison among Vaults to Resist Correlation Attack

Vault-to-Vault Fuzzy Membership				Vaults	Corresponding Vault Parameters
General Template	***Vault 1***	***Vault 2***	***Vault 3***		
0.27	1	0.22	0.275	***Vault 1***	*r* = 20; EPs = 16 CCP = 1; δ = 2.5 *t* = Ter+Bif SCP 6; RMMP 11
0.47	0.22	1	0.31	***Vault 2***	*r* = 16; EPs = 8 CCP = 1; δ = 4.3 t = Ter+Bif SCP 7; RMMP 4
0.24	0.275	0.31	1	***Vault 3***	r=27 NCP=8 CCP=1 δ =5.1 t=Ter+Bif SCP 2; RMMP 7

V. Conclusion

The aim of this research is to propose a method for generating secured template from fingerprint images. The fingerprint minutia features and their related parameters have been analyzed and implement a method that can hide fingerprint features by adding chaffs (fake minutiae) and changing real minutia information to generate secured template in this research. The template pattern is different for every new fingerprint, even if for the same fingerprint with different parameters; and hence, there is no way to get fingerprint features or vault information. From the experiment, it is conclude that the proposed method didn't degrade the performance compared to existing method. It has achieved highest accuracy for verification and has diversity that creates dissimilar vaults to resist correlation attacks. The future research will be employed to extend this method to cancellable approach that will resist all types of attacks on the fingerprint template database in a smart application.

Acknowledgment

We thank the ICT Division (Ministry of Posts, Telecommunications and Information Technology, Bangladesh) for their ICT fellowship program. We also grateful to university research cell for their supports and cooperation.

REFERENCES

[1] Jain, Anil K., Arun Ross and Salil Prabhakar, "An introduction to biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology, vol. 14(1), pp. 4-20, 2004.

[2] Jain, Anil K., Arun Ross and UmutUludag, "Biometric template security: Challenges and solutions", Signal Processing Conference, 2005 13th European, IEEE, 2005.

[3] Uludag, Umut and Anil Jain, "Securing fingerprint template: Fuzzy vault with helper data", Computer Vision and Pattern Recognition Workshop, CVPRW'06 Conference on, IEEE, 2006.

[4] Hao, Feng, Ross Anderson and John Daugman, "Combining crypto with biometrics effectively", IEEE Transactions on Computers, vol. 55(9), pp. 1081-1088, 2006.

[5] Sutcu, Yagiz, Qiming Li and Nasir Memon, "Protecting biometric templates with sketch: theory and practice", IEEE Transactions on Information Forensics and Security, vol. 2(3), pp. 503-512, 2007.

[6] Jain, Anil K., Karthik Nandakumar and Abhishek Nagar, "Biometric template security", EURASIP Journal on Advances in Signal Processing, vol. 2008, article no. 113, 2008.

[7] Tanjarul Islam Mishu and Md. Mijanur Rahman, "Vulnerabilities of fingerprint authentication systems and their securities", International Journal of Computer Science and Information Security (IJCSIS), Vol. 16(3), pp. 99-104, 2018.

[8] N. Ratha, J. H. Connell and R. M. Bolle, "An analysis of minutiae matching strength". In Proc. Audio and Video-based Biometric Person Authentication (AVBPA), Halmstad, Sweden, pp. 223–228, 2001.

[9] Bian Yang and Christoph Busch, "Parameterized geometric alignment for minutiae-based fingerprint template protection", IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, BTAS'09, 2009.

[10] Zhe Jin, Andrew Beng Jin Teoh, Thian Song Ong, Connie Tee, "Fingerprint template protection with minutiae-based bit-string for security and privacy preserving", Expert Systems with Applications, Elsevier, vol. 39(6), pp. 6157-6167, 2012.

[11] Munaga V.N.K. Prasad and C. Santhosh Kumar, "Fingerprint template protection using multiline neighboring relation", Expert Systems with Applications, Elsevier, vol. 41(14), pp. 6114-6122, 2014.

[12] Karthik Nandakumar, Anil K. Jain and Sharath Pankanti, "Fingerprint-based fuzzy vault: implementation and performance", IEEE Transactions on Information Forensics and Security, vol. 2(4), pp. 744-757, 2007.

[13] Alisher Kholmatov and BerrinYanikoglu, "Realization of correlation attack against the fuzzy vault scheme", Security, Forensics, Steganography and Watermarking of Multimedia Contents X, Proceedings of the SPIE - International Society for Optical Engineering, vol. 6819, 2008.

[14] Daesung Moon, Seung-Hoon Chae and Jeong-Nyeo Kim, "A secure fingerprint template generation algorithm for smart card", IEEE International Conference on Consumer Electronics (ICCE), 2011.

[15] T. Charles Clancy, Negar Kiyavash and Dennis J. Lin, "Secure smartcardbased fingerprint authentication", Proceedings of the 2003 ACM SIGMM workshop on Biometrics Methods and Applications – WBMA'03, ACM, 2003.

[16] Umut Uludag, Sharath Pankanti and Anil K. Jain, "Fuzzy vault for fingerprints", AVBPA'05 Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication, Springer, Berlin, Heidelberg, pp. 310-319, 2005.

[17] Ki Young Moon, Daesung Moon, Jang-Hee Yoo and Hyun-Suk Cho, "Biometrics information protection using fuzzy vault scheme", Eighth International Conference on Signal Image Technology and Internet Based Systems, IEEE, Naples, Italy, 2012.

[18] Dario Maio, Davide Maltoni, Raffaele Cappelli, James Wayman and Anil K. Jain, "FVC2002: Second fingerprint verification competition", Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), vol. 3, 2002. DB Site: http://bias.csr.unibo.it/fvc2002/databases.asp.

Design and Fabrication of an Affordable SCARA 4-DOF Robotic Manipulator for Pick and Place Objects

[1]Sara Fatima Noshahi, [2]Adil Farooq, [3]Muhammad Irfan and [4]Narumol Chumuang

[1]Dept. of Mechatronics Engineering, University of Engineering and Technology, Taxila, Chakwal Campus, Pakistan
[2]The BioRobotics Institute, Sant'Anna school of advanced studies, Italy
[3]Department of Electrical Engineering, International Islamic University Islamabad, Pakistan
[4]Faculty of Industrial Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand
*For correspondence; E-mail: engrsarafatima@gmail.com, adil.farooq@santannapisa.it

***Abstract* – Automation of industrial sector is growing rapidly due to deployment of precision robots. In this article, we present a low cost local manufactured 4 degree of freedom (DOF) pick and place robotic manipulator can be used for industrial assembly line applications such as textile, automobile and agriculture sectors. The main concerns persist in most of these robotic manipulators are precision and control. To emphasis on this we used Selective Compliance Assembly Robot Arm (SCARA) to automate pick and place tasks. Our designed robotic arm can carry maximum payload of 2 kg with an arm length of 300mm. We also discuss in detail the manufacturing process and testing results of our developed SCARA robotic manipulator.**

Index Terms: SCARA, 4-DOF Grasping Manipulator, Pick and Place.

I. INTRODUCTION

The concept of robot machine is developing very quickly in today's scenario and it works better than human man power. The robots are deployed in many of the industrial sectors to improve the accuracy as well as the automation of the work in order to increase the production and efficiency of the systems. Since manual pick and place tasks are extremely tedious, troublesome and costly [1]. Additionally, human supervisory control causes errors and mistakes. Therefore robotic control is of great importance. In the modem situation, Robots have progressed toward becoming an integral part industrial revolution. Automation of robotic has made significant progress in a various fields such as products identification, wood placement, plastics and hardware sorting [2-3]. The main reason for utilizing robotic manipulator is to reduce human efforts. These robotic manipulators are popular for speed processing, control, precision and accuracy for pick and place activity which is required in assembly line operations. The Selective Compliance Assembly Robot Arm (SCARA) normally has 4-DOF in which one is linear motion and three rotational movements.

To control the manipulator robotic, wireless RF 2.4 GHz radio systems can be used [4-5]. Portable android application with bluetooth module HC05 has been used as well [6]. However, these need to be ensured for proper reliability. In industries, the SCARA robot is used for different tasks like palletizing, and de-palletizing operations and robotic manipulator is most often made to do the task that it is required to perform [7]. Today, technology is advancing with same pace with increasing human needs. The work done to meet these demands can make life easier, which are determined by robotic manipulator studies. Robot manipulators can work with an outside human user or perform predetermined tasks.

Presently, robotic manipulators are used in various industry such as automobile and food sectors. We designed and developed robot manipulator with 4-DOF using 5 DC servo motors. The robotic manipulator can be connected to the android application via Bluetooth module as reported in [8]. Currently, there are many manipulator robotic manipulators available in the market. However, most of them are very costly [9]. Our aim is to develop a low cost robotic manipulator for researchers and academicians in educational and research institutions to learn the fundamentals of robotics such as design, control, dynamics, kinematics and sensing [10]. Also robotic manipulators can be deployed on other mobile platforms for performing tasks in unstructured environments [11].

This paper discusses the steps used in design and fabrication of a 4 degree of freedom (DOF) SCARA robot without deploying complex PLC system to reduce cost. We started with initial specification, design concept, product manufacturing, and testing. In initial specification phase, suitable parameters of the SCARA robotic manipulator are first found. After that, the best design concept for our SCARA robot was chosen among all set parameters. In third phase, product manufacturing is done; the chosen design of the SCARA robot is developed. The direct and inverse kinematics, dynamics of the robot are then

modeled. Off shelf parts were selected based on the derived parameters from initial calculations. Electronic parts such as switches and a controller were selected. Finally, the developed SCARA robotic manipulator was tested in lab environment lifting different payloads up to 2 kg weight to verify our initial targeted specifications.

The paper is further organized as; section II explains the design methodology and modeling. Section III shows the CAED design and manufacturing. Robotic arm main features and capabilities presented in section IV, selected parameters and electronic design in section V and VI, simulation and system overview in section VII followed by conclusion and future direction in section VIII.

II. DESIGN METHODOLOGY AND MODELING

The basic aim of our prototype is to build an automated robotic manipulator for pick and place in assembly line manufacturing. The methodology for the whole process is shown in Fig. 1.

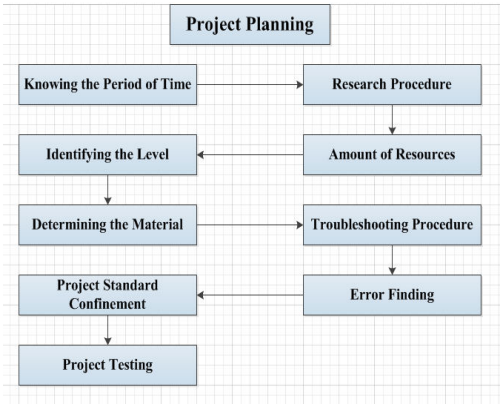


Fig. 1: Project Layout

Robotic manipulator general configuration includes kinematics demonstration, outline structure, electronic and programming working plan [12]. Our proposed robotic manipulator Link diagram is shown in Fig. 2

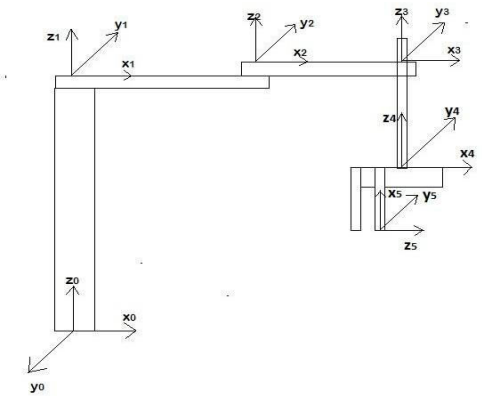


Fig. 2: Link Diagram

Where,
α i-1 = angle from Zi to Zi+1 along Xi
a i-1 = distance from Zi to Zi+1 along Xi
di = distance from Xi-1 to Xi along Zi
Θi = angle from Xi-1 to Xi along Zi
L_1 = 1st Link Length

I	α i-1	a i-1	di	Θi
1	90	0	0	Θ_1
2	0	L_1	0	Θ_2
3	0	L_2	0	Θ_3
4	0	0	0	Θ_4
5	90	L_5	0	Θ_5

L_2 = 2nd Link Length
L_5 = 5th Link Length

The DH calculated parameter values are shown in Table I.

TABLE I: DH PARAMETERS

III. CAED DESIGN AND MANUFACTURING

We designed 3-D robotic manipulator in Solid Edge Design as shown in Fig. 3.

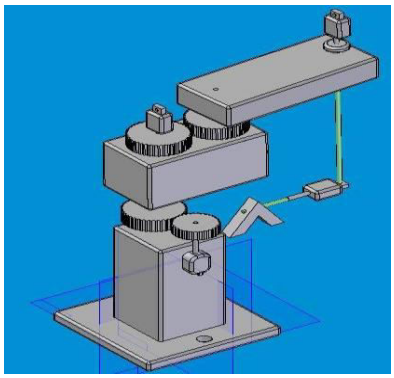


Fig. 3: 3-D Designed Robotic Manipulator

The final SCARA prototype mechanism is shown in Fig. 4 which is powered by 12V DC using Switching Mode Power Supply (SMPS).

Fig. 4: SCARA Robotic Manipulator

The cost of our developed robotic manipulator is under $250 (35000 PKR approx.). It is very less as compared to most of the available SCARA robotic manipulators.

IV. MAIN FEATURES AND CAPABILITIES

The main features of our designed SCARA

robotic manipulator are given in Table II below:

TABLE II: FEATURES AND CAPABILITIES

Features	Capabilities
Control Unit	Robotic manipulator is controlled using PIC18F452 microcontroller and H-bridge circuits
Communication	Manual communication is used for pick and place objects using limit switches.
Robotic Manipulator	It has 4-DOF, 1st Link can rotate complete 360 degrees, 2nd Link by 180 degree, gripper can move only vertically.
Benefits	Minimum human effort, low cost, continuous usage, good accuracy.
Purpose	Our SCARA robot is specifically designed for pick and place operation for industrial applications.

V. SELECTED VALUES AND JUSTIFICATION

We selected best suitable measurements for our mechanism with calculated values in Table III.

TABLE III: DESIGN CALCULATION OF MECHANISM

Component	Type	Dimensions (inch)
1st Link	Iron	Thickness: 1 Width: 2 Length: 12
2nd Link	Iron	Thickness: 1 Width: 2 Length: 10
3rd Link	Aluminum	Thickness: 1 Width: 2 Length: 10
4th Link	Iron	Diameter: 0.3 Length: 9
Gripper Screw	Iron	Diameter: 0.3 Length: 3.6
Gripper	Aluminum	Length: 6.4 Height: 4

Since our aim is for industrial assembly line application we didn't use 5-DOF with more complex design and higher cost. The designed SCARA robot is used for pick and place for a maximum of 2 kg payload with a simpler approach while making it cost effective.

We have used two iron links in SCARA robotic manipulator because iron is rigid. Aluminum is used for 3rd Link showing good weight ratio. We didn't use iron here because aluminum is light and exert minimum load on motors. We used plastic gears due to weight reduction. Screw mechanisms were used because of strong mechanism and currently many industries are using it. Gripper is also made of aluminum due to lower weight. Bearing are used for motion of links and gears.

VI. ELECTRONIC DESIGN

We first simulated our designed electronic board on Proteus software and fabrcated its PCB. The complete circuit is shown in Figure 5.

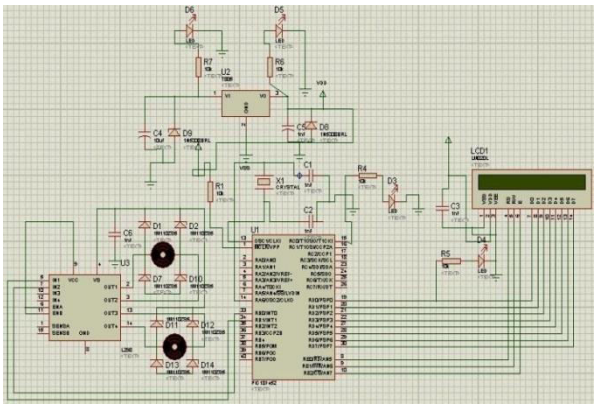


Fig 5: Designed Circuit Simulation

A simple electronics module is designed with push buttons for object grasping. The input signals are processing via controller board to drive the DC motors.

VII. SYSTEM SIMULATION AND OVERALL DESIGN

We system was verified in MATLAB using Transfer Function with pre-defined defined values. The overall block diagram of our designed system is shown in Fig. 6.

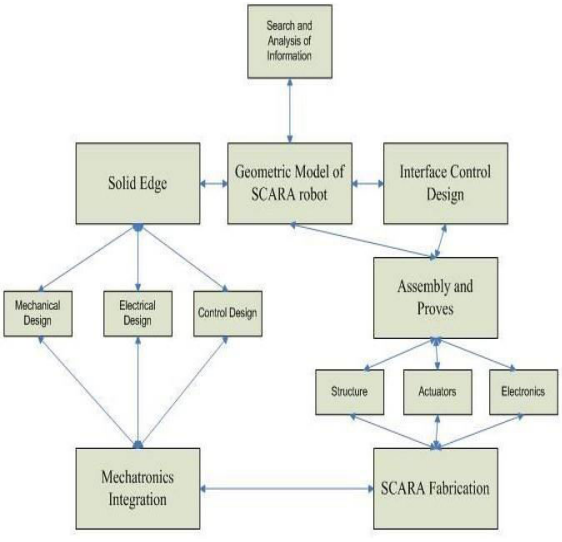


Fig. 6: System Block Diagram

VIII. CONCLUSION AND FUTURE DIRECTION

Robotic Manipulators with less DOF's are more flexible to handle. Our designed 4-DOF robotic manipulator was tested in lab for carrying payloads

up to 2Kg. Modeling and simulations were done to verify our design. We designed a low cost prototype using local materials available in Pakistan. The performance and reliability parameters evaluated for precision and control showed good agreement.

DC gear motors can be replaced with servo motors for more accurate precision. Vision based object detection can be added using CCTV imaging to replace manual control as proposed in [13]. Increase in DOF can provide less error probability in compensation of higher cost.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Mechatronics, University of Engineering and Technology, Taxila, Chakwal Campus for allowing their lab facilities to implement and test our designed 4-DOF SCARA manipulator. We would also like to thank all the authors for their valuable contributions and technical support during the whole process.

REFERENCES

[1] Ashly Baby, Chinnu Augustine, ChinnuThampi, Maria George, Abhilash A P, Philip C Jose “Pick and Place Robotic Arm Implementation Using Arduino” Volume 12, Issue 2 Ver. III (Mar. – Apr. 2017).

[2] AnushaRonanki, M. Kranthi “Design and Fabrication of Pick and Place Robot to be used in Library” Vol. 4, Issue 6, June 2015.

[3] Rahul Gautam, AnkushGedam, AshishZade, Ajay Mahawadiwar, “Review on Development of Industrial Robotic Arm” Volume: 04 Issue: 03 | Mar -2017.

[4] A. Farooq, S. Khaliq and A. Zahid, “A WIRELESS CONTROLLED SEMI-AUTONOMOUS SENSOR BASED UNMANNED GROUND VEHICLE,” Sci-Int. (Lahore), 17(1), pp. 105-108, 2014.

[5] M. Ahsan, K. Abbas, A. Zahid, A. Farooq and S. M. Murtaza, “Modification of a toy helicopter into a highly cost effective, semi-autonomous, reconnaissance unmanned aerial vehicle,” in International Conference on Robotics and Artificial Intelligence (ICRAI), pp. 54–59, 2012.

[6] Ali Medjebouri and Lamine Mehennaoui, "Active Disturbance Rejection Control of a SCARA Robot Arm" Vol.8, No.1 (2015).

[7] Md. AnisurRahman, AlimulHaque Khan, Dr. Tofayel Ahmed, Md. MohsinSajjad, "Design, Analysis and Implementation of a Robotic Arm- The Animator” Volume-02, Issue-10, 2013.

[8] Automatic Material Handling System Using Pick & Place Robotic Arm & Image Processing by Mr. Deepak L Rajnori, A.S Bhide.

[9] https://www.robotshop.com/en/advanced-robotic-manipulators.html

[10] Design and Fabrication of Pick and Place Robot to Be Used in Library AnushaRonanki. Vol. 4, Issue 6, June 2015.

[11] M. Irfan, A. Farooq, Auction-based task allocation scheme for dynamic coalition formations in limited robotic swarms with heterogeneous capabilities, in: IEEE International Conference on Intelligent Engineering Systems, 2016, pp. 210–215.

[12] Manjunath, T. C., "Kinematic Modeling and Maneuvering of A 5-Axis Articulate Robot Arm" world institute of science, building and innovation, pp.363-369, 2007.

[13] N. Chumuang, M. Ketcham and T. Yingthawornsuk, "CCTV based surveillance system for railway station security," 2018 International Conference on Digital Arts, Media and Technology (ICDAMT), Phayao, 2018, pp. 7-12.

Design and Implementation of an Indigenous Solar Powered 4-DOF Robotic Manipulator Controlled Unmanned Ground Vehicle

[1]Adil Farooq, [2]Sundas Arshad, [3]Tayyaba Ansar, [2]Muhammad Irfan and [4]Narumol Chumuang

[1]The BioRobotics Institute, Sant'Anna School of Advance Studies, Italy
[2]Department of Electrical Engineering, International Islamic University Islamabad, Pakistan
[3]Department of Electrical Engineering, University of Engineering and Technology, Taxila, Pakistan
[4]Faculty of Industrial Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand
*For correspondence; E-mail: adil.farooq@iiu.edu.pk

***Abstract*- We have presented in this paper the design, control and implementation of a versatile low cost manipulator with arm gripper configured on an existing unmanned ground vehicle (UGV) for lifting payload (PL) and performing real world tasks. The major development in this work is the robust and efficient stable design of manipulator having four degrees of freedom (DOF) capable of lifting up to 1.5 kg weight for various industrial and non-industrial applications. The communication link is established using two human supervisory controlled wireless four channel 2.4GHz remote controllers, which are separately used for UGV and manipulator for effective maneuvering and control of a 6-DOF overall. The controlling of RC servo motors is made using Arduino Uno controller board. An on board solar panel is used for charging batteries run time during the day. A 50W, 18V standard solar panel is used to enhance the maneuvering time of UGV and manipulator. The unique feature of our UGV is its two rotating head on flippers capable of controlled maneuvering especially in uneven terrain surfaces, stair climbing etc. It can also perform difficult tasks in human unapproachable situations like contaminated or hazardous areas in several industrial and medical applications.**

Index Terms – 4-DOF Grasping Manipulator, UGV, Wireless Controlled, Solar Powered.

I. INTRODUCTION

Nowadays, hazardous and contaminated situations caused either by human beings in wars or by nature calamity, which has forced us to seek new technologies to deal with emergencies more efficiently. It is therefore necessary to introduce capable robotic machines, which cannot only save human life in case of difficult situations but also can perform human capable tasks [1] . In order to increase the efficiency and safety of ground operations, mobile and service robotics is a well-known emerging research area in recent years with addition to multi-functional capabilities [2]. These are generally known as unmanned ground vehicles mostly used in civil and military operations [3], rescue purposes [4], [5] and industrial applications [6]. Such mobile UGVs can also maneuver on rough and uneven terrains [7].

Unmanned ground vehicles can perform rescue operations in hazardous and contaminated areas in war zones and industrial sites. Most of the existing robotic manipulators embedded on the UGVs for carrying and lifting weights for performing different tasks use either PC interface, human supervisory control or Dual tune-multi frequency existing technologies [8], [9].

Currently, the robotic manipulator controlled unmanned ground vehicles are widely used for industrial or military application. However, most of them are either expensive to import or unavailable in developing countries like Pakistan. This issue led us for carrying out this research and to develop an indigenous solution locally. We designed and implemented a low cost robotic manipulator on an existing UGV, which can further be integrated, with modern manipulator for payload lifting and performing real world tasks. To control the robotic manipulator, we used a simple remote control of 2.4GHz radio frequency system. The proposed design of robotic arm has 4 degrees of freedom (DOF) each controlled though a function channel of a wireless RF transducer, where an electrical signal is received at each joint of a DC motor connected. The UGV with mounted robotic manipulator and gripper is also capable of self-charging using embedded standard solar panels for on board rechargeable batteries.

The paper is further organized as; section II explains the design overview. Section III presents the implementation, results and discussion. Future work in IV, conclusion and future work are explained in section V and VI.

II. DESIGN OVERVIEW

In order to improve reliability of modern robotic systems, generally two parallel approaches are often used. First approach is to understand its failure modes [10] and it's modeling while the second approach is to analyse failure of available field data and use models for stimulation and predicting reliability of robotic system as shown in Fig. 1.

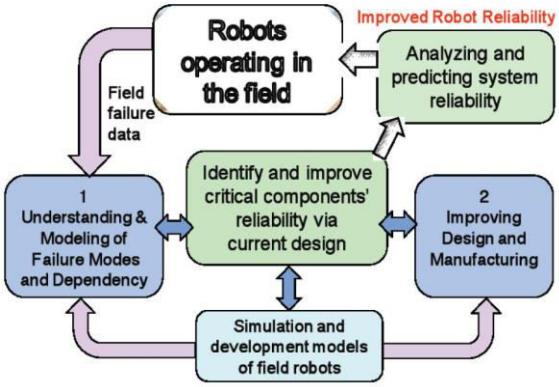


Fig. 1: Robot Reliability

Our designed 4-DOF robotic arm manipulator is controlled from an unmanned ground vehicle with flexible movement using additional two head on flippers. The complete prototype hardware system is divided into three main modules. The first is the mechanical designing and implementation of robotic manipulator and arm gripper using a CAED software. In the second module, we simulated and verified the electronics circuits before fabricated final printed circuit boards for controlling mainly the high voltage motor driving modules. The third and final stage was the integration and testing of solar charger with DC batteries operated on 12 V supply.

A. Simple CAED Design

We designed the robotic mechanical manipulator in Autodesk inventor 2013 CAED software. The main parts consist of base pivot, base plate, base tower, robotic clamp and linear actuator. Each part were first designed individually and assembled in the end as shown in Fig. 2 below.

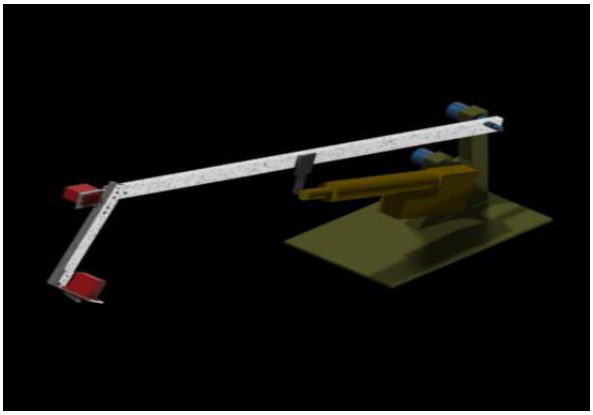


Fig. 2: Manipulator 3D Design

B. Robotic Arm Manipulation

Robotic arm manipulators are used for many applications in industries, medical and natural disaster for grasping and lifting heavy and small objects and to carry out tasks with extreme concentration and accuracy.

We have designed a 4-DOF robotic manipulator. High Torque RC servo motors are connected at each DOF joint link which is controlled through PWM pulse signal generated by Arduino controller board from the RF receiver.

C. Design and Calculations

The two main parts of our designed robotic manipulator are the linear actuator and gripper. Four-bar actuator mechanism is used for to and fro movement while we selected finger type gripper for the pick and place having capability to lift weights up to 1.5 kg.

To determine the required magnitude of the gripper force as a function of these factors, in which weight alone is the force tending to cause the pat to slip out of the gripper.

Torque = Force x Distance

Some other factors while picking up an object are

Torque = Object Weight x Arm Length + Arm Weight x 1/2 x Arm Length

We summed up all forces applied to half of the arm length.

In a 4-DOF, we estimate the desired motors to be used using Denavit-Hartenberg (DH) parameters as shown in Fig. 3.

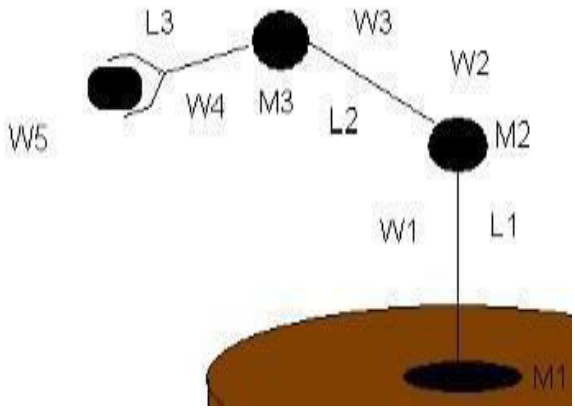


Fig. 3 Weights and Lengths of 4 DOF

The following are the parameters from Fig. 4 described in details in Table I.

Table I: Assigned parameters

Variables	Abbreviation
W1	Weight of the Base Motor
W2	Weight of the Arm Linkage between Base and Upper Arm
W3	Weight of the Motor at Joint of Shoulder and Upper Arm
W4	Weight of the Arm Linkage between Base and Upper Arm
L1	Length of Arm Linked between Base to Upper-Arm
L2	Length of Arm Linked between Shoulder to Fore-Arm
L3	Length of Arm Linked of Fore-Arm Including the Wrist
M1	Base Motor
M2	Upper Arm Servo Motor

The motors torque equation will be:

$M_1 = (L_1/2 \times W_1) + (L_1 \times W_2) + ((L_1+L_2/2) \times W_3) + (L_1 + L_2) \times W4 + (L_1 + L_2 + L_3) \times W_5$ .....(1)

$M_2 = (L_2/2 \times W_2) + (L_2 \times W_3) + (L_2 + L_3/2) \times W_4 + (L_2 + L_3) \times W_5$ .....(2)

$M_3 = (L_3/2 \times W_3) + (L_3 \times W_4) + W_5$ .....(3)

Where M_1 = Base Motor (0 to 90-degree rotation to and fro)
M_2= Arm Motor 1
M_3 = Arm Motor 2 (controlling the movement of gripper)

In order to find out weight our finger gripper could carry we calculated gripping force, contacting fingers, weight of the object carrying, effect of gravity

The required magnitude of the gripper force is calculated using equation below

$$N_f \times F_g = W \times g \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where, N_f is the number of contacting fingers, F_g is the gripper friction, W is the weight of object gripped equal to 1 kg and g is the effect of gravity and acceleration. The required gripper force calculated is

$$F_g = 6N \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

The complete manipulator system weight obtained is 2.4 kg. The calculated data is entered in the following Table II. The robotic arm by calculations is supposed to lift weight up to 5 kg but found practically to lift up to only 1.5 kg of weight due to mechanical constraints.

Table II: Calculated values

Weight Parameters	Weight (kg)	Length Parameters	Length (m)
W1	0.7	L1	1.4
W2	0.4	L2	0.65
W3	0.15	L3	0.25
W4	0.24	-	-

The calculated torques for our design have been found to be 26Nm in the wrist forearm joint, 95Nm at the joint of forearm and upper arm and 180Nm at the joint of upper arm and shoulder. The bottom motor's torque calculated is 190Nm while the output arm accuracy obtained is around 0.95cm.

Fig.4 Manipulator Controlled UGV

Fig. 4 shows our developed robotic manipulator prototype on a UGV for picking an object.

D.Low Cost Design

The robotic arm manipulator and the rotating head on flippers UGV was designed, fabricated and tested using local off the shelf low cost materials such as DC motors used from photocopier machines, RC servo motor from toy helicopters, wheels made from nylon material, and a common industrial used conveyer belt is used for sides on movement. Motor driving circuits were designed using dual mode H-bridge and tested on drive mechanism of DC servomotors [11]. A 12V DC electro mechanical relays and mechanical switches were used to drive the motors carrying high currents up to 10A.

E. Solar Charging Mechanism

We used solar energy to increase the maneuvering time of UGV by charging the DC batteries through a simple designed solar charging circuit shown below in Fig. 5.

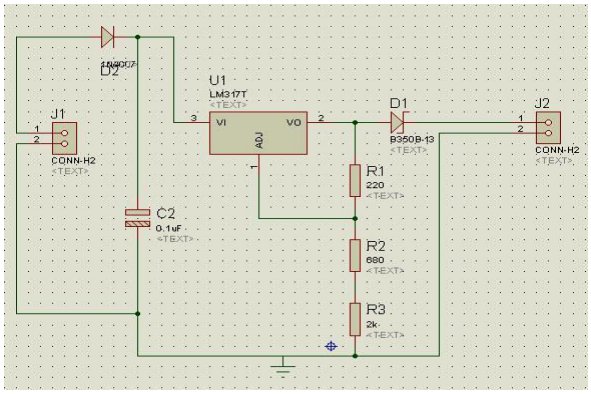


Fig. 5 Solar Charging Schematic

A standard solar panel was used in our final prototype in Fig. 6. The applied input voltage given was between 9-20V, and output voltage obtained varied from 7-14V enough to charge the batteries on board. It was tested and observed an increase in time from 25-30min using solar charging.

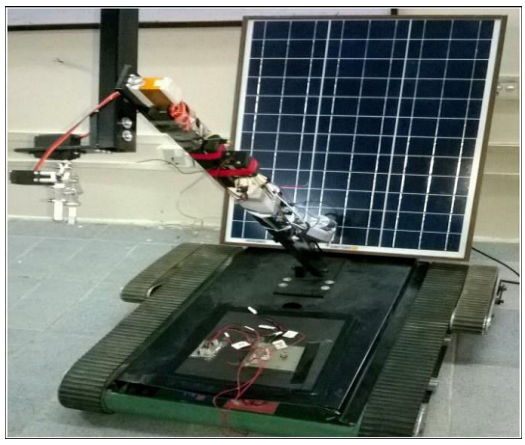


Fig. 6 Implemented Solar Panel

III. OVERALL SUMMARY AND PROBLEMS

The robotic arm manipulator has three main parts. The linear actuator, gripper and RC servo motors. The gripper arm is used for pick and place of objects. A linear actuator is used for the vertical movement of robotic arm. A titanium gear RC servo motor is used for gripper. Three RC servos are used for controlling the robotic manipulator, one for moving and controlling the rod of arm. The second part is the controlling and rotation while the third is opening and

closing of robotic clamp.

We designed the motor drivers H-bridge circuits using replays and switches because they can sustain high currents. The main problem occurred was to overcome the electromotive force. For this purpose, we used power diodes. The UGV is controlled through four channel RC system, which was used bought from a local hobby toyshop.

For designing the solar charger circuit, we used a standard solar panel of 50W and maximum 18V and current is 2.78A. We successfully tested our mounted manipulator and implemented solar panel on the UGV as shown in Fig. 6 for stair climbing, removing hurdles, pick and placed random objects in contaminated areas.

The overall cost of a single robotic arm manipulator was only $400 USD very low as compared to similar existing robotics arm systems. This can further be reduced in mass production. The existing indigenous fabricated UGV cost was under $1000 USD completely manufactured locally in Pakistan.

IV. MANIPULATOR DESIGN AND COST COMPARISON

A list of existing robotic manipulators with multiple DOF embedded on maneuvering platforms is presented in Table III. A cost comparison with payload lifting capability is tabulated with our design. Currently (1$ = 152 PKR Approx).

Table III: Comparison of different Manipulators

Manipulator Design	DOF	PL (Kg)	Cost ($)
Elastic Robotic Arm [12]	7	2	4135
Prosthetic Robotic Arm [13]	6	0.3	1616.24
Robotic Assisted Transfer Device [14]	5	67.5	2416.42
Our Proposed Design	4	1.5	400

Here we can see a tradeoff between cost and performance (DOF and PL) in order to meet the goal for a specific application. Almost all available robotic manipulators are expensive compared to payload carrying capability. However, unlike most robotic manipulators, which are designed for finer manipulation using higher DOF. It is observed that fewer DOFs reduce control complexity and save physical space, which is paramount for any mobile platform.

V. CONCLUSION

We proposed an efficient design and development of a low cost effective robotic manipulator mounted on a UGV having multi function capability. An arduino controller was used for controlling of RC servos and linear actuator using speed control algorithm. All electronic modules were controlled through human supervised remote-control system. Reconfigured UGV maneuvering time was enhanced in daytime using embedded on board solar panel by 20-30min. The designed solar powered 4-DOF robotic manipulator can be used in military, industrial automation and in supervised surgical applications.

VI. FUTURE WORK

To experiment identifying and carrying different payloads with enhanced precision and making the UGV fully autonomous using AI and computer vision tracking capability [15]. Furthermore, multiple of similar swarm robotic manipulator controlled unmanned ground vehicles can be incorporated for performing specific task in a group to increase the multi-functional capabilities as proposed by M. Irfan and A. Farooq [2].

ACKNOWLEDGEMENT

The authors would like to thank the Department of Electronics Engineering, International Islamic University Islamabad for issuing the UGV-II hardware prototype to implement and test our designed 4-DOF robotic manipulator with a gripper arm. We would also like to thank all the authors for their valuable contributions and suggestions for writing this manuscript.

REFERENCES

[1] A. Farooq, S. Khaliq, and A. Zahid, “a Wireless Controlled Semi-Autonomous Sensor Based Unmanned Ground Vehicle.,” *Sci. Int.*, vol. 27, no. 1, pp. 105–108, 2015.

[2] M. Irfan and A. Farooq, “Auction-based task allocation scheme for dynamic coalition formations in limited robotic swarms with heterogeneous capabilities,” *2016 Int. Conf. Intell. Syst. Eng. ICISE 2016*, pp. 210–215, 2016.

[3] B. B. Nair, T. Keerthana, P. R. Barani, A. Kaushik, A. Sathees, and A. S. Nair, “A GSM-based versatile Unmanned Ground Vehicle,” in *International Conference on “Emerging Trends in Robotics and Communication Technologies”, INTERACT-2010*, 2010.

[4] A. W. Y. Ko and H. Y. K. Lau, “Intelligent robot-assisted humanitarian search and rescue system,” *Int. J. Adv. Robot. Syst.*, 2009.

[5] J. Suthakorn *et al.*, “On the design and development of a rough terrain robot for rescue missions,” in *2008 IEEE International Conference on Robotics and Biomimetics, ROBIO 2008*, 2008.

[6] A. G. Bruzzone, M. Massei, R. Di Matteo, and L. Kutej, “Introducing Intelligence and Autonomy into Industrial Robots to Address Operations into Dangerous Area,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019.

[7] G. Bayar, A. B. Koku, and E. I. Konukseven, “CoMoRAT: A Configurable All Terrain Mobile Robot,” *Proc. 11Th Wseas Int. Conf. Autom. Control. Model. Simul.*, 2009.

[8] S. Karmoker, M. M. H. Polash, and K. M. Z. Hossan, “Design of a low cost PC interface Six DOF robotic arm utilizing recycled materials,” in *1st International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2014*, 2014.

[9] R. Kannan Megalingam, T. Pathmakumar, T. Venugopal, G. Maruthiyodan, and A. Philip, “DTMF based robotic arm design and control for robotic coconut tree climber,” in *IEEE International Conference on Computer Communication and Control, IC4 2015*, 2016.

[10] M. Ahsan, K. Abbas, A. Zahid, A. Farooq, and S. Mashhood Murtaza, “Modification of a toy helicopter into a highly cost effective, semi-autonomous, reconnaissance Unmanned Aerial Vehicle,” in *2012 International Conference on Robotics and*

Artificial Intelligence, ICRAI 2012, 2012.

[11] J. J. Craig, “Introduction to Robotics - Mechanics and Control 3E (John Craig).pdf,” *IEEE Journal on Robotics and Automation*. 1986.

[12] M. Quigley, A. Asbeck, and A. Ng, “A low-cost compliant 7-DOF robotic manipulator,” in *Proceedings - IEEE International Conference on Robotics and Automation*, 2011.

[13] V. S. Kumar, S. Aswath, T. S. Shashidhar, and R. K. Choudhary, “A novel design of a full length prosthetic robotic arm for the disabled,” in *Advances in Intelligent Systems and Computing*, 2017.

[14] G. G. Grindle, H. Wang, H. Jeannis, E. Teodorski, and R. A. Cooper, “Design and User Evaluation of a Wheelchair Mounted Robotic Assisted Transfer Device,” *Biomed Res. Int.*, 2015.

[15] N. Chumuang, M. Ketcham and T. Yingthawornsuk, "CCTV based surveillance system for railway station security," 2018 International Conference on Digital Arts, Media and Technology (ICDAMT), Phayao, 2018, pp. 7-12.

Thai Keyword Extraction using TextRank Algorithm

Rattapoom Kedtiwerasak*, Ekkarat Adsawinnawanawa†, Pimolluck Jirakunkanok†, and Rachada Kongkachandra*
Email: r.kedtiwerasak@gmail.com {ekkarat, pimolluck}@moddang.org rdk@sci.tu.ac.th
*Department of Computer Science, Faculty of Science and Technology, Thammasat University, Phatumthani, Thailand
†Feedback180 Co., Ltd., Bangkok, Thailand

***Abstract*—Information Extraction (IE) refers to the automatic extraction of structured information such as entities, relationships between entities, and attributes describing entities from unstructured sources. Keyword Extraction is the part of information extraction to discovering implicit and potentially important keywords in underlying unstructured natural-language texts. Due to the inherent chracteristic of Thai written language which does not explicitly use any word delimiting characters, identifying individual words. In this paper, an alternative method to word formation for noun phrase recognition is proposed. The word formation is improving keyword extraction using the compound noun pattern. We use the word formation to applying the TextRank algorithm to grouping the noun phrase, thire is selected as candidate to calculate in algorithm. The experiments are 2,727 documents in banking domain from social online such as Facebook, Twitter, online news. The experimental results yield 30.73% of accuracy with significant improvement by word formation.**

***Keywords*—TextRank algorithm, keyword extraction, information extraction, noun phrase recognition, word formation, compound noun pattern**

I. Introduction

Information Retrieval (IR) is a process of discovering useful information or knowledge from unstructured text corpus. One important task in IR is Information Extraction (IE) which is the process of discovering implicit and potentially important keywords in underlying unstructured natural-language texts [1], [2]. The Keywords play important role in several applications such as text summarization, summarised information for searching, and question answering system. The benefit of using keyword list as document representative can reduce reading time, assist the user to arrange the large amount of documents and facilitate users to communicate with machine.

Approaches for keyword extraction can be carried out, such as supervised and unsupervised machine learning, statistical methods and liguistic knowledge [3]. TextRank algorithm [4] is the keyword extraction systems based on the unsupervised machine learning techniques respectively. Although this keyword extraction systems yield the acceptable extraction performances, that will be suitable for Latin-based languages, such as English and Spanish, in which words are delimited by using special charcters such as period (.), comma (,), and space characters. These languages are often referred to as segmented languages. However, the word-level approach can not be directly applied for some languages, which do not explicitly use any word delimiting characters, such as Chinese, Japanese, Korean and Thai. These languages are referred to as non-segmented languages. For Thai written language, many word segmentation algorithms are available, but none of them yields perfect results due to the ambiguity in language usage [5]–[7]. In addition, the segmented words have been combined and grouped using compound noun patterns, their are increasingly and specifically the meaning [8]–[10].

In this paper, we propose an alternative solution to Information Extraction using TextRank algorithm which is Thai. The approach uses the compound noun patterns to grouping as noun phrase [11]–[13]. The word formation to grouping the noun phrese can improves extracted keywords from TextRank algorithm.

The rest of this paper is organized as follows. In next section, we review the background of TextRank algorithm and related works. In Section 3, the proposed method present the overview of system and explain the word formation step. In Section 4, experiments with accuracy the result. And the paper concludes in Section 5.

II. Background

A. TextRank Algorithm

TextRank algorithm is a graph-based ranking algorithms which are essentially a way of deciding the importance of a vertex within a graph, based on global information recursively drawn from the entire graph [4]. There is applied to rank words based on their associations in the graph, and then top ranking words are selected as keywords. The candidate of keyword is achieves its best performance when only nouns and adjectives are selected as potential keywords [4], [14].

Formally, let $G = (V, E)$ be a graph with the set of vertices V and set of edges E, where E is a subset of $V \times V$. For a given vertex V_i, let $In(V_i)$ be the set of vertices that point to it, and let $Out(V_i)$ be the set of vertices that vertex points to. The formula of TextRank algorithm is defined as follows:

$$WS(V_i) = (1-d) + d \times \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

where $WS(V_i)$ is a word score of vertex V_i and d is a damping factor which set to 0.85 [4], [15]. w_{ij} is a weight of connection between two vertices V_i and V_j which added to the corresponding edge that connects the two vertices.

The elements in the graph is consisted of the word and the type of characteristics, the application of graph-based ranking algorithms to natural language texts cosists of the following steps:

1) Identify text units, and add them as vertices in graph.

2) Identify relations that connect text units, and use these relations to draw edges between vertices in the graph where relations is a co-occurence units.
3) Iterate the graph-basedranking algorithm until convergence.
4) Sort vertices based on their final score and use the values attached to each vertex for decision to keywords.

B. Ralated Works

Previously TextRank algorithms for extracting keywords was proposed in Thai or Asian languages. In [8], a study of chinese and thai bilingual topic detection extract the information of news with keywords of each chinese and thai documents through the TextRank algorithm. Similar to our work, thier study considers the compound words formed by a combination of simple words in thai. However, their work uses the vocabulary contained Chinese-Thai dictionaries to compoud phrases, that is makes over fit phrases from dictionaries. In [9], a study of language-indepent string pattern analysis was presented. Their study considers the generation of substrings from text corpora of non-segmented languages. However, their work focuses on the NLP issues such as morphology and Path-of-Speech tagging.

In [10], a study of extracting keyphrases from chinese news articles uses query logs as the phrase generation training corpus. Thier study considers length and position of phrase involved to extracted keyphreses output. Their keyphrases output is more informative and readable than keyphrases output without using length and position of phrase. Their compounding, length, and position of phrase, convey to increasing the informative of keywords. Thus, our approach imposes the compound noun patterns to order in substring level.

III. Proposed Method

A. Overview

This paper aims to improve keyword extraction process in TextRank algorithm for Thai language using Word Formation. There are four main steps in our system. An overview of the proposed method is following Figure 1. We expect this method can extact thai keywords correctly. Our corpus was automatically segmented and part-of-speech tagged by segmenter. The corpus was specified banking domain, that is comming from social online (i.e. Facebook, Twitter, online news).

This proposed method starts with morphological analysis step, lexical tokens are segmented and tagged. The result of this step is the stream of word with POS tagged. In the second step, the compound word is grouped into one word. The third is candidate selection and the last step is keyword extraction using TextRank algorithm.

B. Modules

1) Morphological Analysis: The first step in this process is Thai Character Clusters (TCCs), based on the Thai laguage spelling features. A TCC is an unambiguous unit that is smaller than a word but lagger than a character and cannot be further devided. The composition of TCC is unambiguous

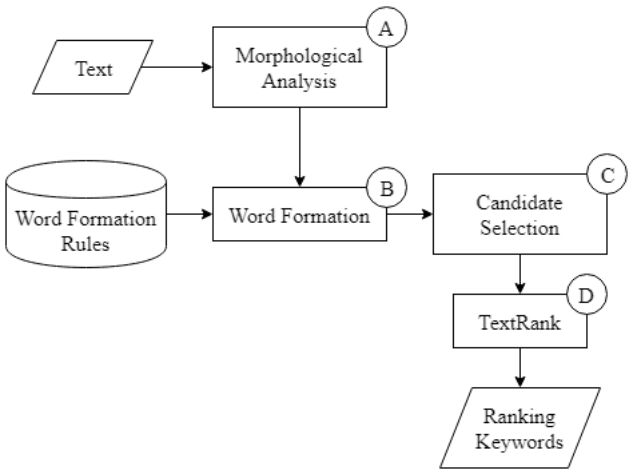


Figure 1: An overview of proposed method

and can be defined by a set of rules. For example, a front vowel and the next character have to be grouped into a same unit. A tonal mark is always located above a consonant and cannot be separated from the consonant. A rear vowel and the previous character have to be grouped into a same unit. The TCC can be applied to improve the accuracy of word segmentation and spell checking [16].

The second step is Word Segmentation (WS). Thai text tokenization require specialized algorithms to find word boundaries prior to tokenization. For Thai word segmentation, Choochart et. al. (2008) used machine learning-based (MLB) approaches to identify and categorize characters surroundding the boundaries. The best of MLB approach for Thai word segmentation is Conditional Random Field (CRF) algorithm, with precision and recall of 95.79% and 94.98% [17]. The result in this step is the stream of segmented word.

The third step is Part-of-Speech Tagging (POS). This step is used a Thai part-of-speech tagging builded corpus named ORCHID [18]. The tag of POS have 30 unique tag. The tagger determines a part-of-speech tag for each word of current sentence. Then it determines a sematic concept. Both concepts rely on the tag probabilities. The output from this process is the word tokens with POS tagged.

The final step in morphological analysis is Named Entity Recognition (NER), a task of locating and identifying named entities into pre-defined categories. Thai name entities can be classified into five groups including person, organization, place, abbreviation and ambiguous/crosstypes to using word together with POS embedded [19]. The process output is the stream of word tokens with POS tagged which identifify the named entity.

2) Word Formation: In the second step of proposed method, the word formation is providing to find the compound boundary, this advantage is to ba able to analyze new words that constructed from the existent morpheme by the application that need not to add those words in the lexicon. These cause reduction of lexicon size as well as word segmentation

ambiguity. The word formation relies on compound noun pattern (see Table I) and noun-phrase analysis system. The output of this step is the stream of word with POS tagged where the position of POS tags match the patterns instead of noun-phrase.

3) Candidate Selection: In the third step, the candidate selection is a brute-force method might consider all words in a document as candidate keywords. Common heuristics include removing stop words and punctuation; filtering for words with certain parts of speech. Candidate keywords are selecting in parts of speech (Noun, Verb, or Adjective) whice are not stop words.

4) TextRank: The TextRank algorithm uses a set of words that could convay the topical content of document are identified, then these candidates are scored/ranked and the best are selected as a document's keywords. This method assumes that more important candidates are related to a greater number of other candidates, and that more of those related candidates are also considered important.

Essentially, a document is represented as a network whose nodes are candidate key words and whose edges connect related candidates. The highest-scoring terms are taken to be the document's key words. The candidate keywords is selected to be keywords of document where that candidate keyword score more than average score of all candidate keywords score.

C. New Word Generation

In Thai language, word formataion is a way of creating new words or compound noun from the existing words to reduce the lexicon size. There is based on word formation rules [11]–[13]. In this paper uses the simple techique of matching Part-of-speech tag sequences, with the intention of capturing the simplicity of the corresponding syntactic structure.

Table I: Compound Noun Pattern

Pattern	Example
n + n	แม่ (mother) + น้ำ (water) แม่น้ำ (river)
n + v	ห้อง (room) + นอน (sleep) ห้องนอน (bedroom)
v + n	พัด (fan) + ลม (wind) พัดลม (a fan)
v + v	กัน (prevent) + ชน (crash) กันชน (bumper)
n + adj	น้ำ (water) + แข็ง (solid) น้ำแข็ง (ice)
adj + n	หวาน (sweet) + ใจ (heart) หวานใจ (sweetheart)
n + n + v	สาย (line) + ไฟ (fire) + พ่วง (tow) สายไฟพ่วง (power cable)
n + v + n	คน (human) + ขับ (drive) + รถ (car) คนขับรถ (driver)
n + n + n	ข้าว (rice) + ขา (leg) + หมู (pig) ข้าวขาหมู (stewed pork leg on rice)
n + v + v	ใบ (leaf) + ขับ (drive) + ขี่ (ride) ใบขับขี่ (driving license)

From Table I show the patterns of compound noun which each patterns compose of the attached words by part-of-speech tag ("n" is a noun, "v" is a verb, and "adj" is a adjective), that patterns based on thai linguistic rules [13]. The process of new word construction is shown in Figure 2.

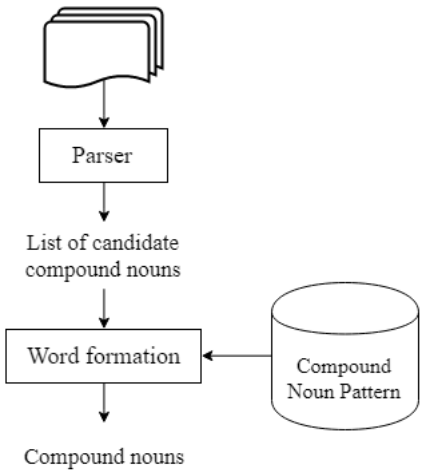


Figure 2: New Word Construction process

Word formation as one of the Noun Phrase (NP) analysis solutions, that occur successively often compounded between primitive units. Moreover, new constructed words can be grouped by Noun Phrase analysis system. Formally, let term in each compound noun pattern is a T set, that compose of noun, verb, and adjective. The rules set of compound noun is inteaded of R. And, the starting symbol for the selection pattern is NP. All this is a structure as 3-tuple:

$$\{T, R, NP\}$$

where:

T is the set of term ($T \in$ {n, v, adj})
R is the set of rules in the grammar (see Table I)
NP is the starting symbol

In the process of noun phrase analysis system, irrelevant words were applied in the boundary identification by keeping in stop list words. In case of having ambiguity about the candidate tree choosing, this process will choose the longest noun phrase.

IV. Result and Discussion

A. Experiments Setting

We set experiment texsting for ability of the proposed method, we use 2,727 datas from social online banking domain by crawling and scratching. We compare the result from proposed method with result from TextRank using only the word segmentation and part-of-speech tagging as a baseline. All dataset and extraction process are same setting in both runs. We measure accuracy by compared with a goal annotation of linguistic experts. The goal annotation has been seperate to

Table II: Accuracy results between baseline and proposed method

	Precision		Recall		F-measure	
	Baseline	Proposed Method	Baseline	Proposed Method	Baseline	Proposed Method
Test set 1	25.57%	**26.18%**	36.74%	**48.30%**	30.15%	**33.95%**
Test set 2	15.40%	**16.74%**	35.41%	**52.50%**	21.46%	**25.39%**
Test set 3	23.55%	**25.61%**	43.00%	**63.13%**	30.43%	**36.43%**
Test set 4	16.13%	**18.21%**	36.75%	**53.22%**	22.42%	**27.13%**

Table III: Average accuracy results between baseline and proposed method

	Precision	Recall	F-measure
Baseline	20.16%	37.98%	26.12%
Proposed Method	**21.69%**	**54.29%**	**30.73%**

4 test set from 4 linguistic experts. We calcuate accuracy of keyword extraction result by correct extracted keyword and partial matching with goal.

B. Experimental Results

From experiment setting, the accuracy results on extracted keywords compare by the goal annotation of linguistic experts. When we compare the results between baseline and the proposed method, which keyword extraction based approaches, it shows the higher performance with 26.12% and 30.73%, respectively. The proposed method has higher accuracy than baseline (4.61%) as follows Table II and Table III.

C. Discussion

According to the experimental results, the proposed method for efficient Thai keyword extraction. The process of word formation can improve keyword extraction result and increase accuracy of extracted keyword correctly. On the other hand, the limitation of this work is the dataset whose from social online. There are having the noisy text such as insertion, tranformation, transliteration and onomatapoeia. The noisy text has impact be domino effect in system. For an example case in morphological analysis is affected with result of segmented word and part-of-speech incorrect, that are comimg from unknown words. According to the recall score, it can be seen that the percentage of proposed method is higher than baseline by 16.31. It show that the coverage of the proposed result is better than the baseline.

V. Concludsion

This paper presents a method to get the information retrieval of Thai laguage in the keyword extraction process of TextRank algorithm. This proposed method works as a pre-process along with a segmenter. The proposed method was developed by using the word formation. We use part-of-speech information and compound noun pattern for combine them to noun phrase. There increase accuracy of keyword extraction using TextRank algorithm. The experimental results show that the accuracy of the proposed method higher than baseline. In future work, we will apply semantic to include in TextRank algorithm for calcuration the score to enhance the accuracy of the information extraction system.

Acknowledgment

This work is co-working by the Feedback180 Co., Ltd., Bangkok, Thailand to crawling/scatching and pre-processing data from social online banking domain. And Feedback180's linguistic team had been answer the goal annotation.

References

[1] S. Soderland, "Learning information extraction rules for semi-structured and free text," *Machine Learning*, vol. 34, no. 1, pp. 233–272, Feb 1999. [Online]. Available: https://doi.org/10.1023/A:1007562322031

[2] U. Yong Nahm and R. J. Mooney, "Text mining with information extraction," 09 2002.

[3] S. Siddiqi and A. Sharan, "Keyword and keyphrase extraction techniques: A literature review," *International Journal of Computer Applications*, vol. 109, pp. 18–23, 01 2015.

[4] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *EMNLP*, 2004.

[5] P. Charoenpornsawat, B. Kijsirikul, and S. Meknavin, "Feature-based thai unknown word boundary identification using winnow," in *IEEE. APCCAS 1998. 1998 IEEE Asia-Pacific Conference on Circuits and Systems. Microelectronics and Integrating Systems. Proceedings (Cat. No.98EX242)*, Nov 1998, pp. 547–550.

[6] A. Kawtrakul, C. Thumkanon, Y. Poovorawan, P. Varasrai, and M. Suktarachan, "Automatic thai unknown word recognition," 1997.

[7] V. Sornlertlamvanich, T. Potipiti, and T. Charoenporn, "Automatic corpus-based thai word extraction with the c4.5 learning algorithm," in *Proceedings of the 18th Conference on Computational Linguistics - Volume 2*, ser. COLING '00. Stroudsburg, PA, USA: Association for Computational Linguistics, 2000, pp. 802–807. [Online]. Available: https://doi.org/10.3115/992730.992762

[8] Z. Rang, L. Zhou, J. Zhang, Y. Xian, and Z. Yu, "Chinese and thai bilingual topic detection online," *MATEC Web Conf.*, vol. 100, p. 02055, 2017. [Online]. Available: https://doi.org/10.1051/matecconf/201710002055

[9] T. Yamashita and Y. Matsumoto, "Language independent morphological analysis," pp. 232–238, 01 2000.

[10] W. Liang, C.-N. Huang, M. Li, and B.-L. Lu, "Extracting keyphrases from Chinese news articles using TextRank and query log knowledge," in *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 2*. Hong Kong: City University of Hong Kong, 12 2009, pp. 733–740. [Online]. Available: https://www.aclweb.org/anthology/Y09-2035

[11] N. Pengphon, A. Kawtrakul, and M. Suktarachan, "Word formation approach to noun phrase analysis for thai," in *In the proceeding of SNLP2002*, 2002.

[12] A. Kawtrakul, M. Suktarachan, P. Varasai, and H. Chanlekha, "A state of the art of thai language resources and thai language behavior analysis and modeling," in *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization - Volume 12*, ser. COLING '02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 1–8. [Online]. Available: https://doi.org/10.3115/1118759.1118766

[13] N. Bumrung, "An automatic compound noun extraction system for thai sentences," Master's thesis, Department of Computer Science, Faculty of Science and Technology, Thammasat University, 2015.
[14] S. Rose, D. Engel, N. Cramer, and W. Cowley, *Automatic Keyword Extraction from Individual Documents*, 03 2010, pp. 1 – 20.
[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," in *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, 1998, pp. 161–172. [Online]. Available: citeseer.nj.nec.com/page98pagerank.html
[16] T. Theeramunkong, V. Sornlertlamvanich, T. Tanhermhong, and W. Chinnan, "Character cluster based thai information retrieval," in *Proceedings of the Fifth International Workshop on on Information Retrieval with Asian Languages*, ser. IRAL '00. New York, NY, USA: ACM, 2000, pp. 75–80. [Online]. Available: http://doi.acm.org/10.1145/355214.355225
[17] C. Haruechaiyasak, S. Kongyoung, and M. Dailey, "A comparative study on thai word segmentation approaches," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, vol. 1, May 2008, pp. 125–128.
[18] V. Sornlertlamvanich, N. Takahashi, and H. Isahara, "Building a thai part-of-speech tagged corpus (orchid)," *Journal of the Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 189–198, 1999.
[19] N. Tirasaroj and W. Aroonmanakun, "Thai named entity recognition based on conditional random fields," in *2009 Eighth International Symposium on Natural Language Processing*, Oct 2009, pp. 216–220.

Intelligent Credit Service Risk Predicting System Based on Customer's Behavior By Using Machine Learning

[1]Jittimaporn Chaisuwan and [2]Narumol Chumuang
[1]*Dep. of Industrial Technology Management, Faculty of Industrial Technology,*
Dep. Of Digital Media Technology, Faculty of Industrial Technology,
[1,2] *Muban Chombueng Rajabhat University*, Ratchaburi, Thailand.

[1]jchaisuwan_da@hotmail.com, [2]lecho20@hotmail.com

***Abstract*— This paper present a model for predicting the behavior of customers in an intelligent manner in the form of customer credit risk for use in decision making for executives in business organizations by applying pmanagement's decision to credit corporate customers by machine learning techniques. Machine learning techniques, which is an important technique that is the heart of artificial intelligence (Ai). Decision Tree model that is analyzed for credit business customers in business organizations which consists of six attributes, the number customers of 1,100 records. The risk is classified into three class, namely, low, medium and high. Our model analysis can be done with accuracy of 85.82% of the customer genius in the form of risk to credit customers of the trading organization has been developed to support the decision making of executives more quickly and more effectively.**

***Keywords*— intelligent, credit service, behavior, forecast, risk.**

I. INTRODUCTION

Intelligent customer behavior predicting system as a result, the credit risk of customers is increasing and therefore the business Intelligence system has been developed to help make the decision of organization more systematic and automated [1]-[4]. When the business intelligence system began to play an extensive role in many fields in both public and private careers causing organizations to bear the pressure to plan appropriate strategies to achieve the objectives of the organization and with changing circumstances [5], [6]. Including the methods and steps of the decision- making process in order to respond appropriately and quickly [7]. Although the current technology will change in terms of mining data management [8], [9] and faster database connections creating more complex models [10] or models and analysts need to learn about AI (Artificial Intelligence) system [11],[12], which is artificial intelligence more, is a program that has been written and developed to be intelligent, capable of thinking, analyzing, planning and making decisions by processing from large databases such as amazon Alexa and Siri. Therefore, any type of work that is working as a model can be replaced by all Artificial Intelligence whether driving, accounting or financial analysis, investment and credit, even complex tasks [13]. Must use analytical thinking and can be replaced as well for the financial technology industry. AI is widely used [14],[15], whether it is a business of borrowing money, insurance business, debt collection or credit scoring. If we know how to use AI technology to be useful and always adjust to develop their knowledge and skills will be able to generate many income as well business [16],[17]. Intelligence helps transform the data into information [18], knowledge, and finally users can make intelligent business decisions [19] and take action into the intelligent behavior predicting system to know the risk of release credit will make accurate credit decisions and the least risk of trading of the organization able to make quick decisions in time for commercial lending to customers of the trading organization.

Therefore, this paper is interested for developing a system to predict the behavior of customers in a smart way. In the form of credit risk, the customers of the trading organization to support the predicting and management decisions. The technique used is Decision Tree relies on customer group data of the trading organization from the information system of the trading organization to display both forms, summary tables, graphs, and predicting results of future sales organizations in order to increase the efficiency of data analysis for better quality.

II. RELATED THEORIES AND RESEARCH

Different predictions for each problem are different. In predicting each problem [20], it should be considered important factors. Especially the risk of lending to customers of the trading organization and analyzing customer classification qualification type [21], customer satisfaction along with the classification of consumers according to the consumption motivation.

A. C4.5 Decision Tree

From the research of intelligent behavior predicting systems for customers will see the credit risk characteristics of the trading organization to support the predicting and decision making of executives from the decision tree process will be divided into two types:- regression tree for regression and classification tree for classification or sometime call the

Decision Tree for types: CART [22], [23]. How to do Decision Tree gradually divide the data into two parts (recursive binary split) from the bottom of the tree, called the root node and chase up to the leaf node in the Fig. 1 on the left and make a prediction, target variable with a simple method: use the mean value of the target variable node by split. The data from the root node to the leaf node is done until the specified condition is obtained. Such as the depth of the tree, not more than 10 layers (max dept) or number b data in each of the divided or leaf node a minimum five observations or min sample as shown in Fig. 1.

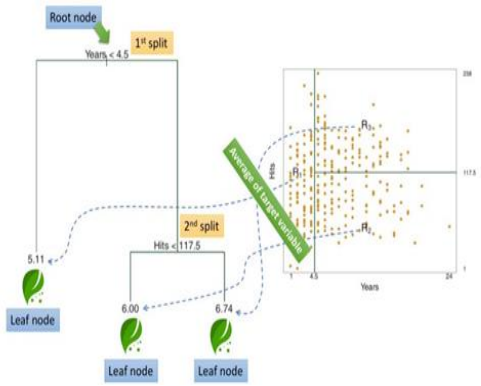


Fig.1. Example of prediction with decision tree from ISLR sixth printing.

The principle of dividing data in each node for information that contains k feature and observation is as follows.

Select 1 feature from *k* feature to sort the data with the value of the selected feature.

Find all possible split points from the data. N observation can find points for dividing possible data *n-1* points (simple idea if min sample = 1)

For dividing each data as possible, calculate the Residual Sum of Squares (RSS) from prediction, target variable with the mean of target variable in each group.

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2, \qquad (1)$$

RSS equation

Eq-1:(1) R_j = each group of observation is divided into all J groups (here is $J = 2$),

(2) y_i = target variable,

(3) = $\hat{y}_{Rj}$ prediction values in each group calculated. Comes from the mean value of the target variable in that group

Gini impurity is a measure of impurity or the purity of the class in each data group divided by each split point for a binary classification problem with a target variable of 0 or 1. A good split should have two groups of data that can separate class 0 with class 1.

Come out clearly in each group. The more you can classify the target variable, the lower the Gini impurity value.

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \qquad (2)$$

Gini impurity equation

Eq-2: Set the class of target variable to have all *K* class (binary classification case is $K = 2$), $\hat{p}_{mk}$ = proportion or % of class *k* within group => If in group or in the node that can be divided, can separate class of target variable can be issued for 1 pure class, which will cause impurity = 0 because the $\hat{p}_{mk}$ value of that class is equal to 1. Puts the values in brackets = 0 and other classes are equal to 0 making the values outside the brackets = 0.

B. Literature Review

A. Gahlaut, Tushar and P. K. Singh, [24] present customer classification in banking system" is a classification of customers in the banking system of Iran according to the form of credit risk. They used TOPSIS to rank the list of applicant credit checks in the Refah Bank after calculating the significant coefficients of each study indicator about real customers according to TOPSIS and the steps of actual customer data that are 103 new customers that have been checked. The results of the ranking show that only 98 suitable predictions and 5 wrongs forecasts are credited with 95% accuracy of all new customers.

S. Kurniawan, R. Kusumaningrum and M. E. Timu [25] studied a classification of customer satisfaction attributes: An application of online hotel review analysis is the classification of customer satisfaction properties: the application of online hotel review analysis. In order to classify the characteristics of customer satisfaction with hotel services, empirical data is collected through Daodao. com, Chinese brand. Online travel review website tripadvisor.com Depending on the message set and content, the analysis found. That The features that create customer satisfaction with hotels are hotels, places, rooms, services, food costs and facilities. The features associated with 7-dimensional hotels are the most important and food availability, value and facilities are not as important as the role of hotel services.

A. L. M. Cruyt, A. A. Ghobbar and R. Curran [26] proposed to analyzed to classify low cost airline customers by using service marketing mix factors (7Ps) by assuming the airline A is the airline of the Thai people and the airline B will be the airline from abroad. To analyze statistical data by allowing 225 airlines each. Samples of Don Mueang airport from 450 questionnaires were processed by using the SPSS statistical program for low-cost airline service behavior. It was found that most of the objectives of traveling for leisure travel were 169 persons to 37. 60% and followed by the purpose of traveling for work and business, 146 persons to 32.40% and traveling frequency is 1 year. -2 times, with the highest number of 356 people, 79. 10% and traveling on weekends and public holidays 268 people to 56.60%.

K. Boonsthonsatit and S. Jungthawan, [27] show a classified online music service customer groups streaming in Thailand, in six groups, and 400 sample groups by using questionnaires by age 5, by submitting questionnaires on

websites related to this subject. The result is a group:- Group (1) of non-primary customers in the range of 36-40 years of age, using only a trial, spending very little. Group (2), IT specialists are 18-36 years old, are students and private employees. Where the income is still low It is easy to press, a few buttons can be used. Group (3), the general user group is in the age range of 26-30 years. It is used during trial and continuous use and the cost is moderate Is a student and a private employee whose income is still low. Group (4), innovation group is the main customer who accepts new technology in the age range of 31-40 years, high quality, not monotonous, original service, reasonable cost. Group (5), current group high yield group Not sensitive at the price range between 36-40 years old, influenced by media sharing per group. The last one, age under 18 years is a potential group. Use the service quite high Is a group of young students / students.

H. Amroun, M. H. Temkit and M. Ammi, [28] proposed the relationship between consumer innovation and behavior watching and accepting the repertoire of drama that is an online by using the relationship test between consumer innovation and the behavior of watching TV drama online and acceptance of the repertoire that is retro a total of 400 study samples were used to watch online TV and were aged between 18-35 years old, mostly female 78.50%, male 21.50%. The results showed that Most of the sample groups have a bachelor's degree education and occupation of students and students, with the majority of sample groups having an average monthly income of less than 10,000 baht.

III. METHODOLOGY

Research is a system for predicting the behavior of customers in an intelligent manner. Will see the credit risk characteristics of the trading organization There is a plan for data preprocessing. The steps are as follows.

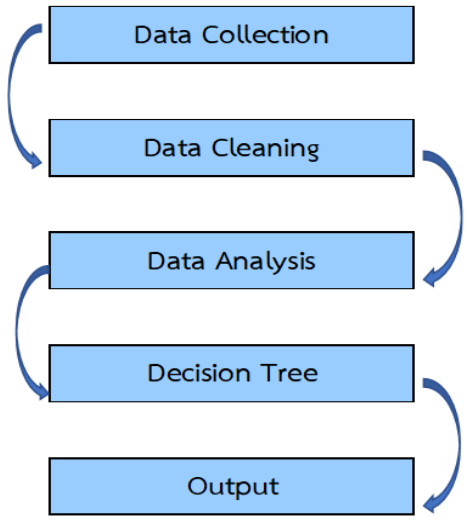


Fig.2. Overview of this system.

Step 1: we start with the data collection in the form of 2 types of storage is:-

Primary data is the original report that is currently in use in the organization which is a basic report that executives can use immediately. Which is in the form of presentation as show in Fig. 2.

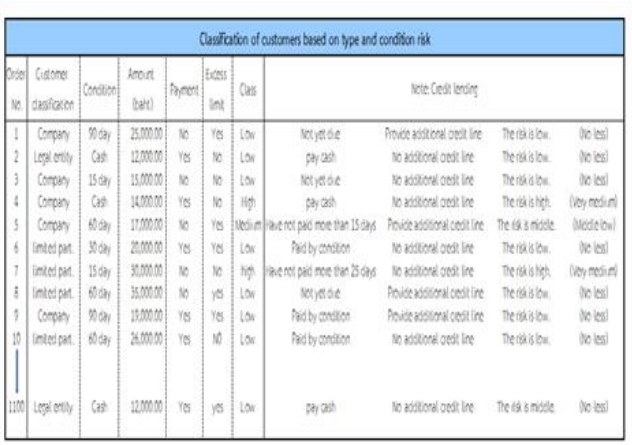


Fig. 2. Sample of the original customer's behavior database.

Secondary data (secondary data) is the collection of additional information from the trading organization's information system, such as individual customer data . Total sales data trading conditions with financial and accounting files, etc. from table 4, collect all customer data collecting data includes 1,100 customer data and six attributes as shown in Fig. 3.

Classification of customers based on type and condition risk

Order No.	Customer classification	Condition	Amount (baht)	Payment	Excess limit	Class
1	Company	90 day	25,000.00	No	Yes	Low
2	Legal entity	Cash	12,000.00	Yes	No	Low
3	Company	15 day	15,000.00	No	No	Low
4	Company	Cash	14,000.00	Yes	No	High
5	Company	60 day	17,000.00	No	Yes	Medium
6	limited part.	30 day	20,000.00	Yes	Yes	Low
7	limited part.	15 day	30,000.00	No	No	high
8	limited part.	60 day	35,000.00	No	yes	Low
9	Company	90 day	19,000.00	Yes	Yes	Low
10	limited part.	60 day	26,000.00	Yes	N0	Low
1100	Legal entity	Cash	12,000.00	Yes	yes	Low

Fig. 3 Sample of customer's behavior database after cleaning step.

Step 2: the data cleaning process are used for cutting irrelevant data such as topic for quality in data analytic process. Then use the typed search filter improve all six attribute.

Step 3: the data analysis is to send data, click on the open File button, select the file to input all of data set 1,100 records by cutting order "*No.*" There are six attributes already mentioned by Decision Tree.

Step 4: the result of each attribute according to rules created from trees help decide from this tree view above. The architecture tree of our system illustrate in Fig. 4.

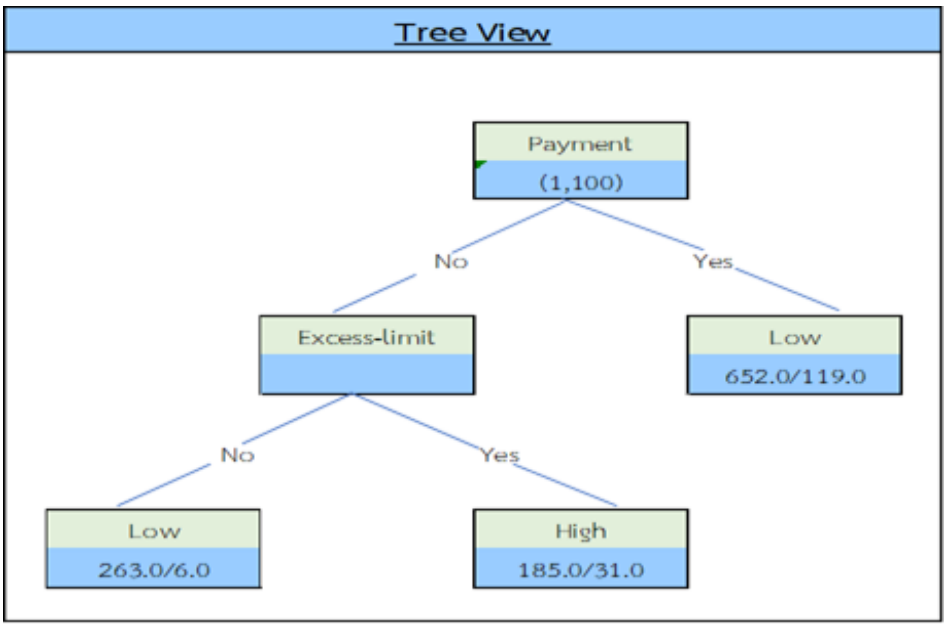


Fig. 4. Show a tree of an intelligent credit service risk predicting system based on customer's behavior.

Rule 1. The number of 1,100 customers has already paid 662 people. Not yet paid 448 because the payment is not yet due, therefore the risk is low or no risk at all.

Rule 2. The 448 unpaid customers and the organization provide additional credit and not yet paid 263 unpaid time, therefore low risk.

Rule 3. Customers who do not pay 448 and the organization does not provide additional credit. Not yet paid185 more. Not yet due to pay a small amount and exceed the amount of payment time: therefore, there is a high risk.

Rule 4. Customers that the organization provides additional credit in the number of 263 customers who do not have 6 conditions, a little more than payment, low risk.

Rule 5. Customers that the organization does not provide additional credit in the number of 185 customers who did not perform the conditions, 31 were in excess of the payment period: high risk.

All of this rule of C4.5 algorithms for prediction the credit risk of customer with six attributes are shown in Fig.6.

IV. EXPERIMENT RESULTS

Intelligent customer behavior predicting system in the form of risk to credit customers of the trading organization for the support and predicting of customers in the trading organization by using sales data and registering each customer as a database use the organization's information system to analyze through the business intelligence system with C 4.5 algorithm. The customer's behavior data set amount of 1,100 records are used for contribute prediction model. We design our experimental with 10-folds cross validation. The results show as in Table I. For accuracy rate of the risk predicting system based on customer's behavior are show in Table II.

TABLE I. THE ACCURACY BY CREDIT'S RISK CLASS

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.992	0.411	0.863	0.922	0.923	Low
0.000	0.000	0.606	0.000	0.160	Medium
0.975	0.033	0.832	0.975	0.898	High

TABLE II. THE ACCURAY RATE

DETAIL	RECORDS	RATE
Correctly Classified Instances	944	85.8182 %
Incorrectly Classified Instances	156	14.1818 %
Kappa statistic	-	0.6207
Root mean squared error	-	0.2818

Confidence in this perspective without having to face very high risk or very little risk in providing credit to customers

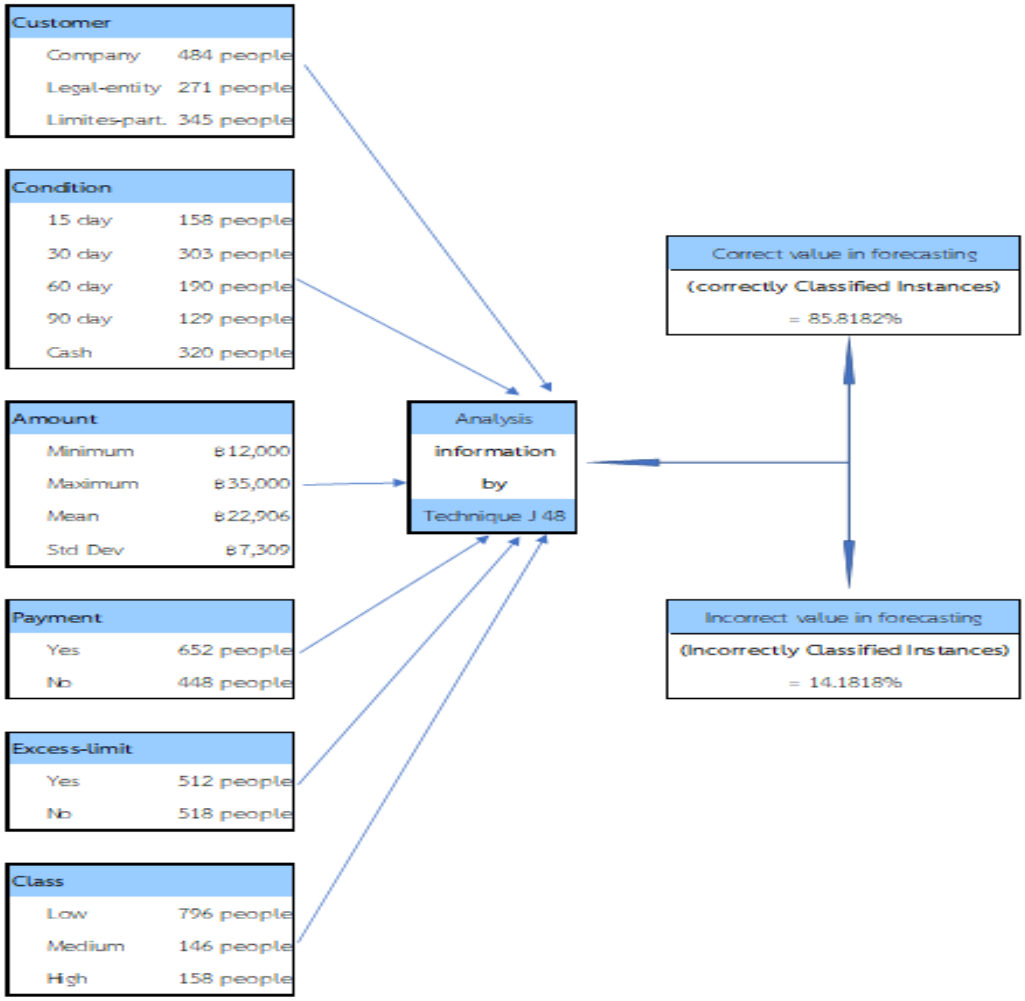


Fig.6. The description of attributes used for our system.

V. CONCLUSIONS

The results for intelligent behavior predicting systems In the form of risk to credit customers of the trading organization From 1,100 corporate customers to forecast and make executive decisions in deciding to release each client's credit In conditions that are not the same from the differences in customer behavior. Therefore can see that the organization can make predictions in advance and the decision to release credit is accurate or up to 85.82% and is not accurate or lacks confidence only 14.18%. Therefore, the intelligent behavior predicting system for customers Is a support for predicting of customers in the trading organization and increase the efficiency in analyzing and planning the management of executives in making loan decisions for customers in the trading organization without risk or risk of having bad debts from customers in the trading organization

Suggestions for intelligent behavior predicting systems for customers In the form of risk to credit the customers of the trading organization in the Software Weka system can used six Attribute to forecast the J48 technique to support the predicting of customers in the trading organization and increase the efficiency of the analysis and planning of the management of the executive in deciding to let customers in the organization continue And is a view to increasing sales without risk or risk Therefore holding the system to predict the behavior of customers in an intelligent manner used to make decisions of corporate executives in lending to customers.

REFERENCE

[1] S. Susan, S. K. Khowal, A. Kumar, A. Kumar and A. S. Yadav, "Fuzzy Min- Max Neural Networks for Business Intelligence," 2013 International Symposium on Computational and Business Intelligence, New Delhi, 2013, pp. 115-118.

[2] A. Sienou, A. P. Karduck, E. Lamine and H. Pingaud, "Business Process and Risk Models Enrichment: Considerations for Business

Intelligence," 2008 IEEE International Conference on e- Business Engineering, Xi'an, 2008, pp. 732-735.
[3] Y. Shi and X. Lu, "The Role of Business Intelligence in Business Performance Management," 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, Kunming, 2010, pp. 184-186.
[4] N. Chumuang, M. Ketcham and T. Yingthawornsuk, "CCTV based surveillance system for railway station security," 2018 International Conference on Digital Arts, Media and Technology (ICDAMT), Phayao, 2018, pp. 7-12.
[5] V. Zamudio, P. Zheng and V. Callaghan, "Intelligent Business Process Engineering: An Agent Based Model for Understanding and Managing Business Change," 2012 Eighth International Conference on Intelligent Environments, Guanajuato, 2012, pp. 141-148.
[6] A. Karapantelakis and J. Markendahl, "Challenges for ICT business development in intelligent transport systems," 2017 Internet of Things Business Models, Users, and Networks, Copenhagen, 2017, pp. 1-6.
[7] M. Hachem and B. K. Sharma, "Artificial Intelligence in Prediction of PostMortem Interval (PMI) Through Blood Biomarkers in Forensic Examination–A Concept," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 255-258.
[8] N. Chumuang, "Comparative Algorithm for Predicting the Protein Localization Sites with Yeast Dataset," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 369-374.
[9] J. Ming, L. Zhang, J. Sun and Y. Zhang, "Analysis models of technical and economic data of mining enterprises based on big data analysis," 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, 2018, pp. 224-227.
[10] T. Jin and Z. Fu, "Data Mining for Complex Thermal System Modeling," 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Shandong, 2008, pp. 560-564.
[11] W. Cui, Z. Xue and K. Thai, "Performance Comparison of an AI-Based Adaptive Learning System in China," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 3170-3175.
[12] R. L. L. Sie et al., "Artificial Intelligence to Enhance Learning Design in UOW Online, a Unified Approach to Fully Online Learning," 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Wollongong, NSW, 2018, pp. 761-767.
[13] M. Hachem and B. K. Sharma, "Artificial Intelligence in Prediction of PostMortem Interval (PMI) Through Blood Biomarkers in Forensic Examination–A Concept," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 255-258.
[14] N. Chumuang and M. Ketcham, "Model for Handwritten Recognition Based on Artificial Intelligence," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-5.
[15] M. Ketcham, T. Ganokratanaa and S. Bansin, "The Forensic Algorithm on Facebook Using Natural Language Processing," 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, 2016, pp. 624-627.
[16] T. S. Adeyelure, B. M. Kalema and K. J. Bwalya, "Development of Mobile Business Intelligence framework for small and medium enterprises in developing countries: Case study of South Africa and Nigeria," 2016 4th International Symposium on Computational and Business Intelligence (ISCBI), Olten, 2016, pp. 11-14.
[17] J. Kiruthika and S. Khaddaj, "Impact and Challenges of Using of Virtual Reality & Artificial Intelligence in Businesses," 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), Anyang, 2017, pp. 165-168.
[18] B. Ramesh and A. Ramakrishna, "Unified Business Intelligence Ecosystem: A Project Management Approach to Address Business Intelligence Challenges," 2018 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, 2018, pp. 1-10.
[19] P. Valter, P. Lindgren and R. Prasad, "Artificial intelligence and deep learning in a world of humans and persuasive business models," 2017 Global Wireless Summit (GWS), Cape Town, 2017, pp. 209-214.
[20] Zhu Xiaoliang, Yan Hongcan, Wang Jian and Wu Shangzhuo, "Research and application of the improved algorithm C4.5 on Decision tree," 2009 International Conference on Test and Measurement, Hong Kong, 2009, pp. 184-187.
[21] B. Nie, J. Luo, J. Du, L. Peng, Z. Wang and A. Chen, "Improved algorithm of C4.5 decision tree on the arithmetic average optimal selection classification attribute," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 1376-1380.
[22] P. Charoenporn, "Reservoir inflow forecasting using ID3 and C4.5 decision tree model," 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), Beijing, 2017, pp. 698-701.
[23] Xuefei Wang and Yan Shi, "Design and implementation of targeting advertising system based on C4.5 algorithm," 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), Harbin, 2015, pp. 669-672.
[24] A. Gahlaut, Tushar and P. K. Singh, "Prediction analysis of risky credit using Data mining classification models," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-7.
[25] S. Kurniawan, R. Kusumaningrum and M. E. Timu, "Hierarchical Sentence Sentiment Analysis Of Hotel Reviews Using The Naïve Bayes Classifier," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-5.
[26] A. L. M. Cruyt, A. A. Ghobbar and R. Curran, "A Value-Based Assessment Method of the Supportability for a New Aircraft Entering Into Service," in IEEE Transactions on Reliability, vol. 63, no. 4, pp. 817-829, Dec. 2014.
[27] K. Boonsthonsatit and S. Jungthawan, "Lean supply chain management-based value stream mapping in a case of Thailand automotive industry," 2015 4th International Conference on Advanced Logistics and Transport (ICALT), Valenciennes, 2015, pp. 65-69.
[28] H. Amroun, M. H. Temkit and M. Ammi, "Study of the viewers' TV-watching behaviors before, during and after watching a TV program using iot network," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 1850-1855.

Thai-English and English-Thai Translation Performance of Transformer Machine Translation

Kanchana Saengthongpattana[1], Kanyanut Kriengket[1], Peerachet Porkaew[1,2], Thepchai Supnithi[1]
Language and Semantic Technology Laboratory [1]
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
Institute of Computing Technology, University of Chinese Academy of Sciences [2]
Beijing, China
{kanchana.sae, kanyanut.kriengket, thepchai.supnithi, peerachet.porkaew[1,2]}@nectec.or.th

***Abstract*—In this paper, the machine translation models were applied to the Thai-English and English-Thai machine translation task. We investigated three models of machine translation on Thai and English sentence pairs. The translation performance of the transformer model is better than that of the recurrent neural network and the traditional statistical machine translation models. We found that the BLEU scores of the transformer model were the highest in both Thai-English (44.22%) and English-Thai (46.48%) translations. Besides, the results were also analysed linguistically. In comparison with the three models, the errors about detailed description and wrong word ordering were mostly found in the SMT model, whereas wrong word choice and missing words were mostly found in the RNNs model. Although the transformer model could perform much better than others, three error categories – under-translation, over-translation, and incorrect lexical choice – were also found.**

Keywords—Transformer Machine Translation, Recurrent Neural Machine Translation, Statistical Machine Translation, Thai-English Machine Translation, English-Thai Machine Translation

I. Introduction

Owing to the different language families, any characteristics of Thai and English languages are not the same. For instance, verb inflection, word segmentation, and sentence boundary occurs in English, whereas those are not found in Thai. These are extremely challenging for the machine translation between Thai and English languages. Machine translation is the usage of the computer system to achieve the automatic translation from one language into another. From the survey, we found that there has been researches on English-Thai translation by using the rule-based method [1]. The rules are stored in a database for retrieving as a guide for translation. T. Chimsuk [2] proposes a Thai to English framework. The framework can be added a pattern of the language to the system. The system produces the corresponding patterns between Thai and English language. The machine translation research for Thai language has been continuously developed.

Presently, Transformer mechanism is the state-of-the-art machine translation. The transformer model has been proposed by Google researchers in 2017 [4], which can improve the deficiency of RNNs based on attention mechanisms. In addition, this new architecture achieves better results in various tasks of NLP than Neural Machine Translation (NMT) [5], the previous method, based on deep neural networks, i.e. Recurrent Neural Networks (RNNs) [6].

In the late 1980s, Statistical machine translation (SMT) was introduced by the researchers at IBM's Thomas J. Watson Research Center [7]. The translations are generated on the basis of statistical models and their parameters are derived from the analysis of bilingual text corpora. However, such an approach has always been mentioned until now, due to its simplicity and effectiveness.

This paper aims to investigate the three models of machine translation on Thai and English sentence pairs. In addition, we provide systematic analyses of each model on different patterns of source texts.

The rest of our paper is organized as follows. In section 1, we give an overview of the research area. In section 2, we describe machine translation models. The performances of three models were investigated and analysed in section 3 and section 4. The last section is the conclusion.

II. Machine Translation Models

A. Statistical machine translation

Statistical machine translation (SMT) is a corpus-based approach for machine translation which utilizes the concept of probability to find the best translation (y) of a given source sentence (x). Generally, this process is defined as follows.

$$y' = \text{argmax}_y p(x|y)p(y) \quad (1)$$

The $p(x|y)$ and $p(y)$ are the translation and language models respectively. Phrase-based translation models are widely used in many machine translation services. More advance systems prefer to operate on hierarchical phrases [8] and even more complex like trees on source side i.e. F. Zhai *et al.* [9] and Y. Bengio *et al.* [10]. For the language modeling, it is possible to model with n-gram or recurrent neural network approaches [11]. In this work, we observed only the conventional SMT approach as described in S. Lin *et al.* [12].

A main shortcoming of conventional SMT systems is the way that words are represented, usually with strings or symbols. This kind of representation limits the model to find hypotheses pairs which are exactly matched with the given phrase. Some attempts have been explored more on finding hypothesis by using string similarity [13].

B. Recurrent neural machine translation

A recurrent neural network (RNNs) we used in this work was proposed by D. Bahdanau *et al.* [13], called RNNSearch. We briefly review the design of the architecture of RNNSearch here.

The encoder is bi-directional gated recurrent neural network. The hidden state sequences of forward direction $(\overrightarrow{h}_1, \overrightarrow{h}_2, \ldots, \overrightarrow{h}_I)$ and of backward direction $(\overleftarrow{h}_1, \overleftarrow{h}_2, \ldots, \overleftarrow{h}_I)$ are stacked to form the annotation vector $[\overrightarrow{h}_i ; \overleftarrow{h}_i]$. The decoder is two steps gated recurrent neural networks. The decoder is another recurrent neural network to produce target

words. More specifically, the decoder generates the next word from the conditional distribution as formulated by

$$p(y_i|s_j, y_{<j}) = \text{softmax}\left(g(s_j, c_j, y_{j-1})\right), \quad (2)$$

where g is linear transformation of the current decoder state s_j, source representation for time *j-th* c_j and the representation of previous target y_{i-1}. Note that source representation c_j is weighted sum over source annotations which is computed by

$$c_j = \sum_{i=1}^{I} \alpha_{ij}\, h_i, \quad (3)$$

where

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{I} \exp(e_{kj})} \quad (4)$$

and

$$e_{ij} = v_a^T tan\, h\left(U_a s_j^{'} + W_a h_i\right) \quad (5)$$

Note that $s_j^{'}$ is the intermediate hidden state at time *j-th* as formulated by

$$s_j^{'} = \text{GRU}_1\left(y_{j-1}, s_{j-1}\right) \quad (6)$$

Then, decoder hidden state is calculated by the second GRU.

$$s_j = \text{GRU}_2\left(c_j, s_j^{'}\right) \quad (7)$$

To train RNNs, a training set consists of S sentence pairs $\{< x^{(s)}, y^{(s)} >\}_{s=1}^{S}$, and the training objective is to maximize log likelihood w.r.t model parameter θ.

$$\mathcal{L}(\theta) = \sum_{s=1}^{S} \log p\left(y^{(s)}|x^{(s)}; \theta\right) \quad (8)$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{J^{(s)}} \log\, p\left(y_j^{(s)}|s_j^{(s)}, y_{j-1}^{(s)}, c_j^{(s)}\right) \quad (9)$$

C. Transformer-based Neural Machine Translation

We briefly give an overview of Transformer machine translation in this subsection. Transformer [4] is a new encoder-decoder architecture entirely relying on the attention mechanism and feed-forward neural network. For the encoder part, the Transformer network encodes the input sequence with self-attention mechanism. The encoded output of self-attention is computed by:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right), \quad (10)$$

where Q, K and V are query, key and value of a specified layer respectively. In the encoder, Q, K and V are the same, which is word embedding of an input sequence packed together. This method allows the encoder to model global dependencies efficiently.

For the decoder part, two types of attention mechanism are applied *i.e.* decoder self-masked attention and decoder-encoder attention. Self-masked attention is very similar to self-attention in the encoder part except that only the left side of the current position is allowed to be attended, because it is necessary to preserve auto-regressive property of the decoder which generates output from left to right. The decoder-encoder attention layer uses the decoder hidden states as Q and uses encoder output matrix as K and V.

In practice, Q, K and V are divided into smaller chunks, specified by number of heads. Then, outputs of attention in each chunk are calculated independently. Those outputs will be concatenated to form the final matrix. With self-attention, each position can compute its output in parallel. The training process of Transformer is the same as of RNNSearch.

While the recurrent-based encoder allows a position to be seen by other nearby position, which can be seen as local context modeling. Transformer model allows each position to expose freely to other position without long distance limitation.

III. Experimental Setup

In this paper, we did the experiment by using the Thai-English bilingual parallel corpus collected by National Electronics and Computer Technology Center (NECTEC), Thailand. The number of this corpus is 641,533 sentence pairs in various domains. The corpus was preprocessed and then divided into a training set (513,226 sentence pairs), a development set (28,307 sentence pairs), and a test set (100,000 sentence pairs). Preprocessing tasks included Thai corpus word segmentation. This corpus information was shown in the following table.

TABLE I. Experimental Corpus

corpus	sentence	word	training	development	test
Thai	641,533	11,088	513,226	28,307	100,000
English	641,533	12,080	513,226	28,307	100,000

We adapted byte pair encoding (BPE) [14] to pre-process Thai and English corpora. The neural machine translation models used in this paper were consistent in the basic parameter settings. Because each model had its own structure, it was difficult to achieve consistency in terms of performance of parameters. In addition, the parameters tuning in each model were adjusted to achieve of the highest performance. As the aforementioned reason, we used the transformer based on the fairseq modeling toolkit [15]. We also used the Bilingual evaluation understudy (BLEU) score [16] for performance evaluation.

IV. Experimental Results

To verify the performances of machine translation models on Thai-English and English-Thai translations, based on the same corpus, three machine translation models – SMT, RNNs and Transformer models – were used in this experiment. The experimental results are shown in 2 issues that are the overview of the results and the performance of the transformer model.

A. The overview of the results

The overview of the results consists of the BLEU scores of all tested models, as shown in Table 2, and the samples of Thai-English and English-Thai translations for each model, as shown in Table 3 and 4.

TABLE II. EXPERIMENTAL RESULTS

Model	Thai-English	English-Thai
	BLEU	**BLEU**
SMT	30.71	31.96
RNNs	36.65	25.89
Transformer	**44.22**	**46.48**

As the above table, it shows the BLEU scores of three machine translation models used in this experiment. It was found that the RNNs model did well in Thai-English translation whereas the SMT model did well in English-Thai translation. However, the model performing best was the transformer model scoring with nearly 50%, the highest BLEU scores, for both Thai-English and English-Thai translations.

According to the aforementioned performance scores, the error analysis had been done. Mostly, the translation results of the transformer model were satisfying and consistent with the reference source. For the SMT model, its results were quite close to the reference source, but the errors about detailed description and wrong word ordering were found most. In contrast, although some translation results were rather similar to the reference source, it was very surprising that wrong word choice and missing word translation were mostly found in the RNNs model.

To clarify the above translation results and any found errors, it could be seen in Table 3 and 4. For Thai-English translation results in Table 3, it was found that the results of the transformer model were similar to the reference translations, while the ones of the SMT model were slightly different. It translated by using detailed description, which was "a school for the maintenance of orphans", in Example 1; however, it could translate like the reference translation in Example 2. Contrarily, the translation results of the RNNs model were wrong by using the incorrect word "infant" instead of "orphan", in Example 1, and missing many words in Example 2.

TABLE III. SAMPLES OF THAI-ENGLISH TRANSLATIONS

Method	Example 1	Example 2
Original text	โรงเรียน สำหรับ เด็ก กำพร้า	คอมพิวเตอร์ ส่วน ตัว ชนิด นี้ พกพา ได้ ง่าย .
Reference translation	an orphan school	this kind of personal computer is easy to carry.
SMT	a school for the maintenance of orphans	this kind of personal computer is easy to you.
RNNs	an infant school	this kind
Transformer	an orphan school	this kind of personal computer is easy to carry

For English-Thai translation results in Table 4, it was found that the results of the transformer model were also similar to the reference translations. The SMT model translated incorrectly by ordering wrong word positions in Example 1, which was that, in fact, the word "เคลื็ด (sprained)" had to be after the noun "นิ้วมือ (finger)"; nevertheless, it could translate like the reference translation in Example 2. Contrarily, the translation results of the RNNs model were wrong in both Example 1 and 2. The word "ความเร็ว (speed)" were wrongly used instead of "นิ้วที่เคลื็ด (sprained finger)" in Example 1, and the predicate of the sentence, consisting of a verb and an object, was missed in Example 2.

TABLE IV. SAMPLES OF ENGLISH-THAI TRANSLATIONS

Method	Example 1	Example 2
Original text	Check my sprained finger.	My teacher praised me today
Reference translation	ตรวจ นิ้ว ที่ เคลื็ด ของ ฉัน .	ครู ของ ฉัน ชื่นชม ฉัน วัน นี้
SMT	ตรวจสอบ นิ้วมือ ของ ฉัน เคลื็ด	ครู ของ ฉัน ชื่นชม ฉัน วัน นี้
RNNs	ตรวจ ด้วย ความ เร็ว ของ ฉัน	วัน นี้ อาจารย์ ของ ฉัน
Transformer	ตรวจ นิ้ว ที่ เคลื็ด ของ ฉัน .	ครู ของ ฉัน ชื่นชม ฉัน วัน นี้

All in all, the transformer model had the highest BLEU scores and gave the best satisfying results in both Thai-English and English-Thai translations.

B. The performance of the transformer model

As the above part, it was indicated that the transformer model's performances were much better than others. Now, we will describe the performance of this model in detail from its BLEU scores and the error analysis.

Firstly, the transformer model's BLEU scores of Thai-English and English-Thai translations on the data test (100,000 sentence pairs) were quite close. According to Figure 1 and 2, it was found that the BLEU scores with more than 90% occurred most in both Thai-English and English-Thai translations, which emphasized that this model could automatically translate Thai-English and English-Thai text quite well. Surprisingly, the second level of BLEU scores was between 10-20% showing poor translated Thai-English and English-Thai sentences. Also, the third one was less than 10%, giving poor Thai-English and English-Thai translations. Other levels were quite close. However, the second level of BLEU scores remarkably showed numbers of sentences which were more than half of all sentences in the first level, so to analyse any errors found in the translated sentences of the second level of BLEU scores could be logical.

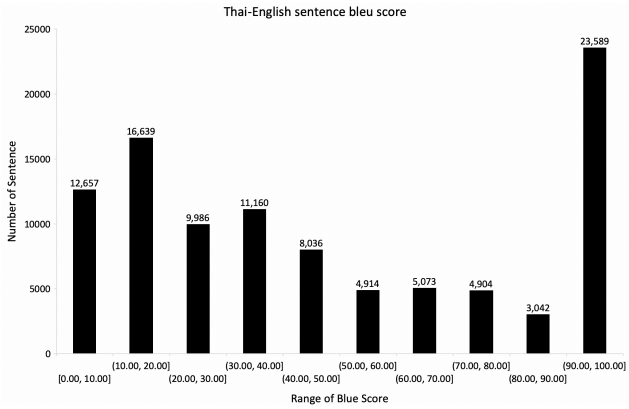


Fig. 1. The BLEU scores range of the transformer Thai-English translation

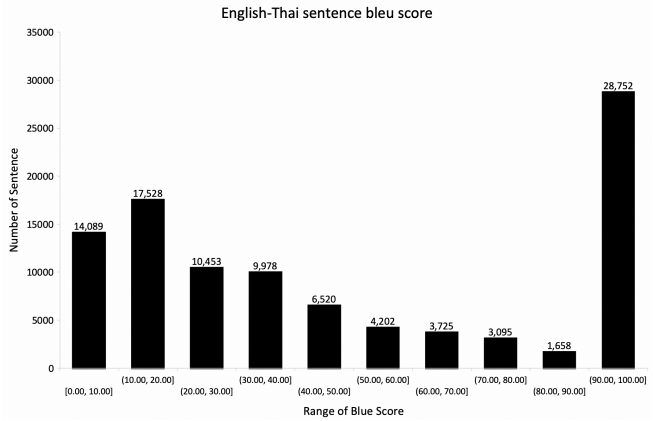


Fig. 2. The BLEU scores range of the transformer English-Thai translation

Secondly, after analysing the error of Thai-English and English-Thai translations, we found that there were 3 error categories: under-translation, over-translation, and incorrect lexical choice, which occurred in both Thai-English and English-Thai translations. The first error type was under-translation which was that some words missed in the target language. The examples of under-translation were shown in the below table.

TABLE V. SAMPLES OF UNDER-TRANSLATION

Source	Target	Missing words
ทิศ ตะวันออก ของ ชายฝั่ง ทาง ตอน ใต้	south coast	ทิศตะวันออก (east)
saddle sore	การ เจ็บ อานม้า	จากการนั่งนาน ๆ (from sitting a long time)

As in Table 5, there were some missing words in both Thai-English and English-Thai translations. For Thai-English translations, only the phrase "ชายฝั่ง ทาง ตอน ใต้ (southern coast)" were translated, whereas the word "ทิศตะวันออก (east)" missed. For English-Thai translations, each meaning of the words "saddle (อานม้า)" and "sore (การเจ็บ)" were translated. In fact, a prepositional phrase "จากการนั่งนาน ๆ (from sitting a long time)" should be used for combining the words "saddle (อานม้า)" and "sore (การเจ็บ)" for clear description, so this phrase "saddle sore" should be translated as "a sore from sitting on a saddle for a long time". The second error type was over-translation which was that unnecessary words, phrases, or clauses were added in the translation result. The examples of over-translation were shown in the below table.

TABLE VI. SAMPLES OF OVER-TRANSLATION

Source	Target	Unnecessary words
เขา เป็น คน มี รสนิยม	he 's tasteful dresser	dresser
a japanned table	โต๊ะ ญี่ปุ่น ที่ มี ราคา สูง	ที่ มี ราคา สูง (which are expensive)

As the above table, unnecessary words or clauses were added in both Thai-English and English-Thai translations. For Thai-English translations, the word "dresser" was not needed because only the word "tasteful" was enough to translate meaningfully, and the original text did not specify the type or domain of being tasteful. For English-Thai translations, the clause "ที่ มี ราคา สูง (which are expensive)" was not needed because it was not mentioned in the source. Contrarily, the original text was just "a japanned table" without any content about the word "expensive". Therefore, to add the clause "ที่ มี ราคา สูง (which are expensive)" was over-translation. Lastly, it was incorrect lexical choice which was that the chosen words for the translation were wrong, as shown in Table 7.

TABLE VII. SAMPLES OF INCORRECT LEXICAL CHOICE

Source	Target	Incorrect words
ฉัน เป็น คอ หนัง	i 've got a leather neck	've got, leather neck
knuckleduster	นัก เขียน เรื่อง เหลวไหล	นัก เขียน เรื่อง เหลวไหล (nonsense writer)

As in Table 7, wrong word choice occurred in both Thai-English and English-Thai translations. For Thai-English translations, the word "เป็น (is)" in the original text showed the quality or state, so to translate "เป็น (is)" into "'ve got" showing possession was incorrect. Moreover, the word "คอ หนัง" in the original text was ambiguous because it has two meanings, a movie fan or a leatherneck which is a soldier in the US Marine Corps. So, the translated word "leatherneck" in the target text could be wrong. For English-Thai translations, the word "knuckleduster" in the original text was translated incorrectly because this word has only one sense with no ambiguity. Actually, it means "a metal weapon worn over the knuckles to increase the injuries when hitting a person" or "สนับมือ" in Thai. The translation result of the word "knuckleduster" that was "นัก เขียน เรื่อง เหลวไหล (nonsense writer)" was obviously wrong. Owing to automatic machine translation, incorrect word choice could occur.

To sum up, the performance of the transformer model could be described from its BLEU scores and error analysis. In general, the model was quite good, proved by its BLEU scores with more than 90%; however, the second level of BLEU scores showed poor translation as its scores were between 10-20%. After the error analysis of Thai-English and English-Thai translations, three error categories which were under-translation, over-translation, and incorrect lexical choice were found. Achieving the errors as stated, the transformer model should give better translation results.

V. CONCLUSION

In this paper, statistical machine translation (SMT), neural machine translation (RNNs), and transformer models were used in Thai and English translation tasks. Through comparison, findings are as following:

1) Although the translation from RNNs gave very impressive output, there were still some limitations. Most of RNNs systems could produce more fluent translations than SMT. But, in some case, RNNs lost the original meanings of its sources [17][18].

2) Translating long sentences is one of the main problems in RNNs because the decoder only generates output from left to right. If an incorrect word is generated in the first half of the output, the second half cannot recover easily when long-term dependencies exist.

3) For some low-resource languages, the translation quality of RNNs does not yield better than SMT [19]. So, training RNNs is a challenging topic.

4) We explored transformer to be compatible with SMT and RNNs models. We found that the transformer yielded the best results based on our dataset.

5) The errors about detailed description and wrong word ordering were mostly found in the SMT model.

6) Surprisingly, the errors about wrong word choice and missing word were mostly found in the RNNs model.

7) Sentence structure types or sentence complexity did not affect any translations errors of each model.

8) The performances of the transformer model were much better than others.

9) Three error categories found in Thai-English and English-Thai translations were under-translation, over-translation, and incorrect lexical choice.

ACKNOWLEDGMENT

This work is support by collaboration between Language and Semantic Technology (LST) laboratory, National Electronics and Computer Technology Center (NECTEC) and Institute of Computing Technology (ICT) and Chinese Academy of Sciences (UCAS).

REFERENCES

[1] K. Chancharoen, N. Tannin and B. Sirinaovakul, "Sentence-based machine translation for English-Thai," IEEE Asia-Pacific Conference on Circuits and Systems, Microelectronics and Integrating Systems. Proceedings, Chiangmai, Thailand, 1998, pp. 141-144.

[2] T. Chimsuk, "A Framework for Implementation of Thai to English Machine Translation Systems," Doctor of Philosophy (Computer Science), School of Applied Statistics, 2010.

[3] K. Phodong and R. Kongkachandra, "Improving Thai-English word alignment for interrogative sentences in SMT by grammatical knowledge," 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, 2017, pp. 226-231.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need", Advances in neural information processing systems, 2017, pp. 5998-6008.

[5] G. E. Hinton, S. Osindero, and Y. W. The, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18.7, 2006, pp. 1527-1554.

[6] P. Brown; S. Della Pietra, V. Della Pietra, and R. Mercer, "The mathematics of statistical machine translation: parameter estimation," Computational Linguistics, MIT Press, vol. 19 (2), 1993, pp. 263–311.

[7] D. Chiang, "A hierarchical phrase-based model for statistical machine translation," Proceedings of ACL2005, 2005, pp. 263–270.

[8] Y. Liu, Q. Liu, and A. Lin, "Tree-to-string alignment template for statistical machine translation," In Proceesings of ACL2006, 2006, pp. 609-616.

[9] F. Zhai, J. Zhang, Y. Zhou, and C. Zong, "Tree-based translation without using parse trees." In Proceedings of COLING 2012, 2012, pp. 3037–3054.

[10] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," Journal OF Machine Learning Research, vol. 3, 2003, pp. 1137–1155.

[11] P. Koehn, O. F. Josef, and D. Marcu, "Statistical phrase-based translation," In Proceedings NAACL2003, 2003, pp. 48–54.

[12] S. Lin, Z. He, Q. Liu, "Partial matching strategy for phrase-based statistical machine translation," The Annual Meeting of the Association for Computational Linguistics (ACL), 2008, pp. 161–164.

[13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate." In Proceedings of ICLR2015, 2015.

[14] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units." Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2015.

[15] Faieseq documentation, https://fairseq.readthedocs.io/en/latest/ [Accessed: Jul. 19, 2019].

[16] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a method for automatic evaluation of machine translation." 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002, pp. 311-318.

[17] J. Niehues, E. Cho, T.-L. Ha, A. Waibel, "Pre-Translation for Neural Machine Translation," "The 26th International Conference on Computational Linguistics (COLING)", Osaka, Japan, 2016,.

[18] Z. Tu, Z. Lu, Y. Liu, X. Liu, H. Li, "Modeling Coverage for Neural Machine Translation," The 54th Annual Meeting of the Association for Com- putational Linguistics (ACL), 2016, pp. 76 – 85.

[19] O. Firat, K. Cho, Y. Bengio, "Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism," The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), San Diego, USA, 2016 .

Development of Fun Hint Game Applications for Special Children on Smart Devices

Burin Narin
Computer Education Department
Muban Chom Bueng Rajabhat University
Ratchaburi, Thailand
burinnarin@hotmail.com

Titichaya Sribun
Computer Education Department
Muban Chom Bueng Rajabhat University
Ratchaburi, Thailand
cookietitichaya@gmail.com

Tanniga Yoongrum
Computer Education Department
Muban Chom Bueng Rajabhat University
Ratchaburi, Thailand
may584144020@gmail.com

Abstract—This paper illustrates the use of Fun Hint Game applications for special children on smart devices. The sample of this experiment were 13 special children from Child Welfare Place with Brain Disabilities Ratchaburi Province, Thailand. The propose of the research were to 1) to design and develop Fun Hint Game application for special children 2) to find the performance of Fun Hint Game applications for special children 3) to compare achievement of students after learning through Fun Hint Game applications for special children on Smart Devices 4) study the level of satisfaction of students towards the Fun Hint Game applications for special children. The results of the research were 1) applications for the special children game on mobile devices were effective for use 2) achievement of students after learning through Fun Hint Game applications for special children higher than before, with statistical significance at the level of .05 3) the students' satisfaction with the fun hint game for special children on mobile devices was at the highest level ($\bar{x}$ = 4.54, S.D. = 0.67)

Keywords—special child, smart device, applications

I. INTRODUCTION

Special children were children that were different from other children of the same age, physical, developmental, intellectual and cognitive ability, and behavior [15]. Special children were divided into 3 main groups which were, 1) children with special abilities, 2) children with disabilities, 3) Poor and underprivileged children [17]. The Ministry of Education has defined the categories and criteria of 9 types of people with disabilities : 1) persons with visual disabilities, 2) persons with hearing impairments, 3) persons with intellectual disabilities, 4) persons with physical disabilities, 5) persons with learning disabilities, 6) persons with speech and language impairments, 7) persons with behavioral or emotional disabilities, 8) autistic person and, 9) persons with disabilities [10]. In this research, we have chosen to develop media for people with learning disabilities. Which were considered as one of the nine defects in children that require special care. The development of learning media to be used with a group with learning disabilities (Dyslexia) is an abnormal learning language that results in the ability to read, spell or write books. There are problems with the separation of the letter sounds and the letter sounds to the different words, which sufferers have this level of intelligence and normal vision. Able to study like a normal person It just requires a lot of effort and takes a long time to read the book [13]. Therefore, developing applications on mobile devices such as mobile phones or tablets That can promote learning and help children develop some of their basic skills by being an application that has the ability to interact And draws students' attention [14]. When the students have fun, it will increase the efficiency of learning.

II. RELATED LITERATURE

Special Children comes from the word " children with special needs" is for children who will need extra support, additional services and also need added guidance in daily routine, social, and academic for improving their potential. Helping program is basis on their individualized or the children's specific need. Special Children were divided into 3 main groups, 1) Children with special abilities , there are less care of helping or seriously developing because most people think they've got abilities or able to survive by themselves but in fact, this is more added pressure to them by expectation how they could have more abilities or learning method. Normal learning method is not satisfy their needs, and may causes boredom also restricted special abilities to be limited. 2) Poor and Underprivileged Children are poor children in poor families those lack of growth factors and learning abilities including disadvantaged in education children due to other reasons such as street children, being used for labor or alien child. 3) Children with disabilities [17]. The Ministry of Education has defined the categories and criteria of 9 types of people with disabilities : 1) persons with visual disabilities, 2) persons with hearing impairments, 3) persons with intellectual disabilities 4) persons with physical disabilities, 5) persons with learning disabilities, 6) persons with speech and language impairments, 7) persons with behavioral or emotional disabilities, 8) autistic person and 9) persons with disabilities [10]. Each group of those mentions are special children with special needed. They should have been received more additional care with essential methods which is different from usual in case of helping them to develop their potential, in order to meet healthy physical, great mental, equal in education and social acceptance [17]. "Children with special needs" means to special children who need specialized education method in different way from the usual methods, both in contents, learning methods, evaluation and also teaching materials. However, Physical defects, Cognitive impairment, Emotional and Society need to stimulate and pushing them to suitable teaching methods on the basis of children individualized need. "Intellectual disability children" are slower development than normal children. Intelligence measurement is lower than General. "Intellectual disability children" may have personality problems or lack of self-confidence due to failure at work and must rely on other, can't learn abstract or complicated things. So the general teaching methods doesn't suit them.

This research is purpose for developing teaching materials by designing into games and trial only with Learning Disabilities. The importance of designing learning through games was considered to be the use of educational innovation, with the use of interactive and entertaining forms of games to be combined with content within subjects that require students to learn. Designed in a combination style that will enable students to gain both knowledge and enjoyment at the same time [5]. Teaching using games Is a teaching-learning process that is student-centered Is a game that is played for learning. The objective is for the learners to learn while or after playing the game. Causing learners to have meaningful learning according to the specified objectives by allowing students to play the game according to the rules and use the results of the students' games to use in summarizing the learning [12]. The game structure consists of 3 main components, which are 1) challenges 2) responding and 3) have the feedback to players [8]. The used of games to help increase learning of students. Will enable the learners to learn as much as possible as in figure 1.

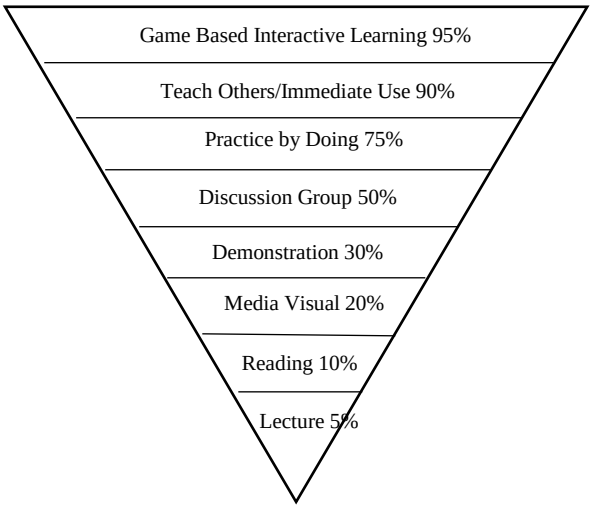


Fig.1. The importance of learning through games compared to other forms of learning [5].

Most of Educational handle games are emphasize about game participating and leading to learning achieve. The highlight of game is student will play over and over again. Including 3 main parts ; Input, Process and Outcome. There are 2 parts of Input process, 1) contents 2) characteristics of game (the game processing is follow by the game system such as 1) User Judgment 2) User Behavior and 3) System Feedback. The ultimate result is learning outcomes [7] according to figure 2.

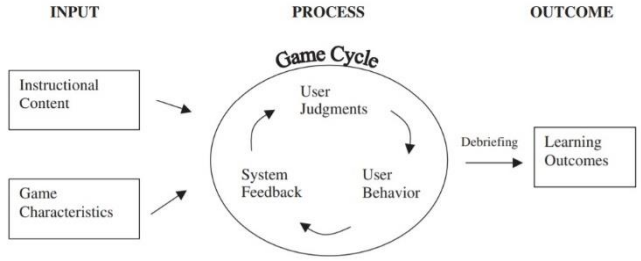


Fig. 2. INPUT-PROCESS-OUTCOME Game Model

III. METHODOLOGY

Mobile devices such as smart phone and tablet that widely used recently. Many students bring theirs to school, it's the age of applications generation [6]. In general, people used these devices in messaging, searching, socializing or playing games. Due to many applications about education, mobile devices can enhance learning for example game applications, game developing also game design should define objective, improving model and test all the game parts. Those mentioned process can apply for a good learning process design and define sub purpose to create model for object test development from prototype [16]. Vocabulary is the first basic language unit for student because it is the main element for listening, speaking, reading and writing. In addition, knowing vocabularies and meaning will lead students create new phrases and sentences by learning from small units to large. It can be seen that learning vocabulary is important for language learning [11]. This research is an Experimental Research and the target group is 13 children with learning disabilities from the mental and intellectual disability home, Ratchaburi, selected by Purposive Selection. In this research were to developed Fun Hint Game for special children on mobile devices has the following process

1) Study information of special children in the research. the research was chosen with special children in the group of people with learning disability. Syndrome of children with problems or lack of learning skills which consists of reading and writing and math impairments and Reading Disability group with reading skills slower than children of the same age who have similar experiences and learning. The impaired skills of this group of children are disability in reading words correctly lack of fluency in reading. There is a particular term for this group of people with disabilities, which is dyslexia. reading disability also includes disorders in reading comprehension [18].

2) Game design and composition procedures. The researcher therefore designed a game of hints for special children on mobile devices. By creating challenges through a hint game developed with Action Script 3.0, responding to players via the touch screen of mobile devices such as tablets and smartphones. That uses the Android operating system. In the vocabulary design the researcher has analyzed the ability of the sample. Vocabulary selection criteria should be simple word and easy to learn , often use in daily life [11]. Studying from the research of Yu-Ju Lan, Hsiao, I. Y. T., & Mei-Feng Shih. (2018) the research is about children with communication disabilities. They created 3D virtual learning media about virtual life (Second Life) by simulate vary places and surrounding that concern children's experience such as kitchen, dining room, convenience store, play ground, school health center, zoo, supermarket, local market, night market, public area and [20]. Therefore choose vocabulary in everyday life and are quite simple words. Since researcher realize those vocabularies are in children's daily life, multimedia has been brought to motivate and encourage them such as sound and picture or comics [3][9] as figure 3 from application below.

Fig. 3. First page of game application.

In the design of Fun Hint Game for special children on mobile devices has defined 3 categories as follows : 1) item categories including appliance in regularly used for example school bag, fan, etc, 2) animal categories are animals as Monkey, Lion, Elephant, etc. and 3) fruit categories consist various types of fruit such as guava, watermelon, durian and etc. Real pictures are being used together with vector graphics to make the game more interesting.

Fig. 4. Menu of game application.

Each category has 2 levels. Allowing gamers to enter the game in each level. In which level 1 will be a hint game with pictures provided and the gamers will drag the consonants into the correct words as shown in the picture. When the required words have been completed, will continue to the next word. After that, will pass to level 2. Gamers must choose the picture that matches the vocabulary shown.

Fig. 5. Fun Hint Game level 1

Identify applicable funding agency here. If none, delete this text box.

Fig. 6. Fun Hint Game level 2

3) System Feedback. When players answer wrong questions or correct answer the game will respond to players, if the word bounce back to the beginning as telling learner the answer was wrong. And collecting points for playing at each level in order to process the data. Take the application to test with special children. Coordinated with the mental and intellectual care home of Ratchaburi province. To bring the game for students to try. In which the samples are tested before using the application once the data has been collected therefore began the process of trying to study the application. And keep the test results again after the study by the application is finished.

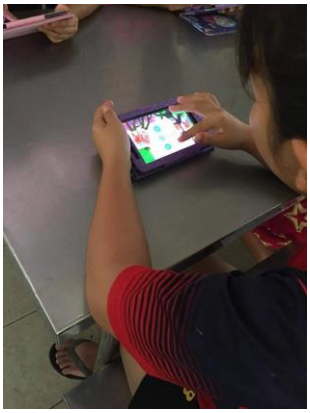


Fig. 7. Special children play Fun Hint Game.

4) Learning outcome. The last steps were data analysis and the conclusion of the experiment by using the tests in the game to find out the effectiveness and achievement of students.

IV. EXPERIMENTAL

From the experimentation of the application of fun hint game for special children on mobile devices with a sample group of 13 students with learning disabilities of the mentally and mentally disabled children in Ratchaburi province in The result were found 1) the Fun Hint Game Application performance is 75/100, which is higher than the set 75/75 threshold. 2) the comparing the academic achievement of students studying with the Fun Hint Game application for special children on mobile devices, it was found that student achievement after school was significantly higher than before studying at level .05 as shown in Table 1.

TABLE I. LEARNING ACHIEVEMENT

Evaluation list	N	mean	S.D.	t-test	Sig
Pre-Test	13	7.54	2.40	3.628	.00
Post-Test	13	10	0		

3) Bring the questionnaire to inquire about student satisfaction by scale model (Rating Scale) and find the average ($\bar{x}$) and standard deviation (S.D.) Level determination divided into Level 5 is maximum score, Level 4 is well , Level 3 is moderate, Level 2 is low and Level 1 is the lowest. Criteria for interpretation will use an average of class interval to analysis by concept of Best [1] as details below

average 4.50-5.00 means the satisfaction level = maximum

average 3.50-4.49 means the satisfaction level = well

average 2.50-3.49 means the satisfaction level = moderate

average 1.50-2.49 means the satisfaction level = Low

average 1.00-1.49 means the satisfaction level = the lowest

The results of the study of the satisfaction level of learners on the Fun Hint Game for special children on mobile devices the results show that it is in the highest level as in Table 2.

TABLE II. SATISFACTION LEVEL

Evaluation list	Satisfaction		
	$\bar{x}$	*S.D.*	*Satisfaction level*
Learning applications help students understand the lesson.	4.30	1.25	more
Learning applications help students achieve their learning goals.	4.61	0.65	most
Learning applications help promote self-learning skills.	4.84	0.37	most
Learning applications allow students to participate in comments.	4.30	0.75	more
Learning applications help to apply knowledge in everyday life.	4.30	0.85	more
Learning applications help to save time in learning.	4.53	0.77	most
Learning applications can create interesting atmosphere in the classroom.	4.00	1.15	more
The use of learning applications is not complicated, easy to understand.	4.69	0.63	most
The content structure covers the purpose of the application.	4.84	0.37	most
The application is suitable for students in the age range.	4.46	0.52	more
The content length of each topic is appropriate.	4.61	0.77	most
Examples consistent with the application.	4.31	0.75	more
The language used is appropriate.	4.77	0.44	most
Application communication is clear in both picture and sound.	4.92	0.28	most
There is a quiz before and after the lesson.	4.61	0.50	most
The overall satisfaction level with the application.	4.54	0.67	most

V. CONCLUSION

The result of this study makes it possible to find a game for children with Fun Hint Game application on mobile devices. It is reliable in which 1 3 students from the sample group from the mental and intellectual disability home Ratchaburi Derived from a specific selection by selecting only students with learning disabilities. Found that students have better academic achievement and were satisfied with the application Fun Hint Game for special children on mobile devices at the highest level. The satisfaction result of user found that the overview is at the maximum level ($\bar{x}$ = 4.54, S.D. = 0.67) the most maximum satisfaction field is Communication, clear both video and ($\bar{x}$ = 4.92, S.D. = 0.28) and the lowest satisfaction field in this study is Learning, applications can create interesting atmosphere in the classroom ($\bar{x}$ = 4.00, S.D. = 1.15). according to the research of Phimchanok [19] .With the introduction of some game design formats such as Word finder and Choose it, which are to find the missing letters of the EasyLexia level 1 and level 2 applications, according to research from [5] to solve children with learning disabilities. And also concern with observations from many research found that game can support learning of children in the age between 6 - 12 years. Game can stimulant children to be more interested in learning even in difficult contents, using game motivation can encourage children pay more attention. [2][4]

VI. ACKNOWLEDGMENT

Thank you the mental and intellectual disability home, Ratchaburi, Parents and all staff for assistance in accessing information about special involve providing advice on game design and development for trial. Thank you Department of Computer Education, Faculty of Science and Technology, Muban Chom Bueng Rajabhat University for supporting in develop and improve this application (Fun Hint Game)

VII. REFERENCES

[1] Best, John W. (1977). Research in Education. .ed., Englewood Cliffs, New Jersey. Prentice-Hell, lnc.

[2] Burguillo, Juan. (2010). Using game theory and Competition-based Learning to stimulate student motivation and performance. Computers & Education. 55. 566-575. 10.1016/j.compedu.2010.02.018.

[3] Chen, P.C. (2010). New trend of electronic textbooks. Educ. Mon., 36–40.

[4] Dickey, Michele. (2010). Murder on Grimm Isle: The impact of game narrative design in an educational game‐based learning environment. British Journal of Educational Technology. 42. 456 - 469. 10.1111/j.1467-8535.2009.01032.x.

[5] El-Said, M., & Mansour, S.(2008). Game Based Learning Creating a Triangle Of Success: Play, Interact and Learb. Interational Journal Intelligent Games & Simulation, 5(2).

[6] Gardner, H. & Davis, K. (2013).The app generation. London: Yale University Press.

[7] Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. Simulation & gaming, 33(4), 441-467.

[8] Jan L. Plass, Bruce D. Homer & Charles K. Kinzer. (2015). Foundations of Game-Based Learning. Educational Psychologist, 50(4),258-283.

[9] Lai, C.F.; Chang, C.S. Design and development of online course using scenario-based strategy in teacher education program for classroom management course. J. Educ. Media Libr. Sci. 2005, 42, 433–449.

[10] Ministry of Education. (2009). Determination of Types and Criteria for people with educational disabilities.

[11] Nguyen Thi Nhu y.(2013). Using Vocabulary Games to Develop Thai Vocabulary Learning of The First-Year Students at The University of Social Sciences and Humanities, Ho Chi Minh City, Vietnam. Master of Arts Degree in Teaching Thai as a Foreign Language. Srinakharinwirot University.

[12] Porntip Wongsinudom. (2015). The Development of Tutorial Application on Tablet With Peer to Peer Learning Affected Learning Together of The Third Grade Students in Petchaburi. Master of Education Program in Educational Technology, Silpakorn University.

[13] Rasheed, T., Siddiqua, A., & Naureen, S. (2016). Exploring Behavioral Problems and Psychological Adjustment of Special Children with Learning Disabilities. Journal of University Medical & Dental College, 7(3), 60-63. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=119270948&site=ehost-live

[14] Roxani Skiada, Eva Soroniati, Anna Gardeli & Dimitrios Zissis. (2014). EasyLexia: A Mobile Application for Children with Learning Difficulties. Procedia Computer Science 27(2014), 218 – 228.

[15] Sirirat Ularntinon. (n.d.). Special Child. Retrieved from http://humaneco.stou.ac.th/UploadedFile/72202-8.pdf

[16] Supamit, & Chanseawrassamee (2012). Teaching Adult Learners English Through a Variety of Activities： Perception on Games and Rewards.

[17] Thaweesak Sirirutraykha. (2019). Special Child. Retrieved from http://www.happyhomeclinic.com/academy/sp01-specialchild.pdf

[18] Thurdphong Thongsriratch.(n.d.). Learning Disorder. Ramathibodi Hospital, Thailand.

[19] Udomphon, Phimchanok. (2015). Development Application on Tablet for the Mathematics Learning Disability. 10.13140/RG.2.1.1203.4403.

[20] Yu-Ju Lan, Hsiao, I. Y. T., & Mei-Feng Shih. (2018). Effective Learning Design of Game-Based 3D Virtual Language Learning Environments for Special Education Students. Journal of Educational Technology & Society, 21(3), 213–227. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=130867686&site=ehost-live

Query-by-Example Word Spotting with Fuzzy Word Sizes

Amornchai Wiwatcharee
Department of Computer Engineering
Mahidol University
Nakhon pathom, Thailand
amornchai.wiw@student.mahidol.ac.th

Tanasanee Phienthrakul
Department of Computer Engineering
Mahidol University
Nakhon pathom, Thailand
tanasanee.phi@mahidol.ac.th

***Abstract*—This paper proposed the query-by-example word spotting model for handwritten documents with image fuzzification. Fuzzy size of word images was used to size of problem. The number of classes in each set were decreased, which made it is easy to choose parameters. The Pyramid of Histogram of Oriented Gradients (PHOG) feature and Support Vector Machine (SVM) were employed to use in the model. IAM handwritten database was used for evaluating the model. The result demonstrates that the micro precision of model with image fuzzification and without image fuzzification were 35.11% and 23.54% respectively. However, the accuracies of the models were 35.11% and 40.14% respectively. Thus, the image fuzzification can be used for reducing of type one error with slightly accuracy loss.**

Keywords—word spotting, query-by-example, fuzzy logic, image fuzzification, support vector machine, handwritten recognition, pyramid of histogram of oriented gradients.

I. INTRODUCTION

The increasingly of the digitalized handwritten documents makes the researchers made the numerous of research in field of computer vision on handwritten documents in the last decade. However, in the computer vision on handwritten documents tasks can be divided into two tasks that are handwritten recognition i.e. pattern recognition on handwritten documents and word spotting [1]. In handwritten recognition, the task is to answer that what is the word of the queried word image. In word spotting, the goal is return all the location of the queried word image in the document. Thus, word spotting method is like searching tool which base on the handwritten recognition model.

The word spotting method can be separated into two methods which are Query By Example (QBE) and Query By String (QBS). In QBE, the samples of query word image are used for training and searching. On the other hand, QBS is used the word string to query and searching in the document. Normally, the word spotting method is composed of these three steps which are pre-processing step, feature extraction step and word matching step. In generally, the pre-processing step is like image enhancement step to make the feature extraction step has more solid feature. Mainly image enhancement is noise removal, image sharpening, gamma normalizing.

In feature extraction step, the word represent features are extracted in various of image features such as Histogram of Orientation Gradients (HOG) [2] feature, Scale-invariant Feature Transform (SIFT) [3], Local Binary Pattern (LBP) [4] features, and so on. In word matching step, the word represent features are matched with the documents word images using similarity measurement techniques or the classifier to consider the word image to return the word occurrences of the document.

The features vector in image recognition has been improving throughout the past decade. Pyramid histogram of orientation gradients (PHOG) [5] feature which was used in smile detection. The cropped mouth images are edge detected by Candy edge detector. Then, the images are divided into the grid of pyramid level and extract the HOG feature in each grid to produce the PHOG features vector. Elastic histogram of orientation gradients (EHOG) [6] is used the elastic meshing technique to divide the cell of the HOG to provide the EHOG feature. Gradient Discrete Cosine Transform (G-DCT) [7] is the feature that perform the DCT on each cell of HOG and the DCT coefficients are composed to the G-DCT features vector.

Characteristic Loci feature [8] is the feature vector that collect the number of intersections in eight directions. HOG and LBP descriptors are used to produce the matrix of Gradient Local Binary Patterns (GLBP) [9]. Furthermore, the feature vector can be encoded or pooled to make the fix size feature vector for arbitrary images size. Mhiri et al. [10] proposed the multiscale feature using spherical k-means and soft-threshold operator to encode the Principal Component Analysis (PCA) image patch. He et al. [11] proposed the spatial pyramid pooling to pool the feature of the convolutional neural network into fix size.

However, most of the researches were aimed to improve the feature extraction and similarity measure techniques. Rarely to have the research that focus on preprocessing step. Thus, this paper proposed the preprocessing step using image fuzzification. The idea is using the size of the images to filter out some words that obviously have different size. In general, size of the same word class should not be widely distributed. However, the various of handwritten style can make the distribution wider. In this paper, we aim to present the QBE word spotting model with image fuzzification. Fuzzy set [12] is the set with real number membership value between 0 and 1 instead of binary membership value in ordinary set.

In this paper, sizes of the sharpened images are used to separate the images into fuzzy sets. Then, the PHOG features vector are extracted on each image in each fuzzy set. Finally, the support vector machine (SVM) [13] is used as the classifier of the model and used as a retriever of word images as well. The rest of the paper is constructed as follow. In Section II, the fuzzy word spotting model is described. The experimental result is presented in Section III. Finally, section IV is the conclusion of this paper.

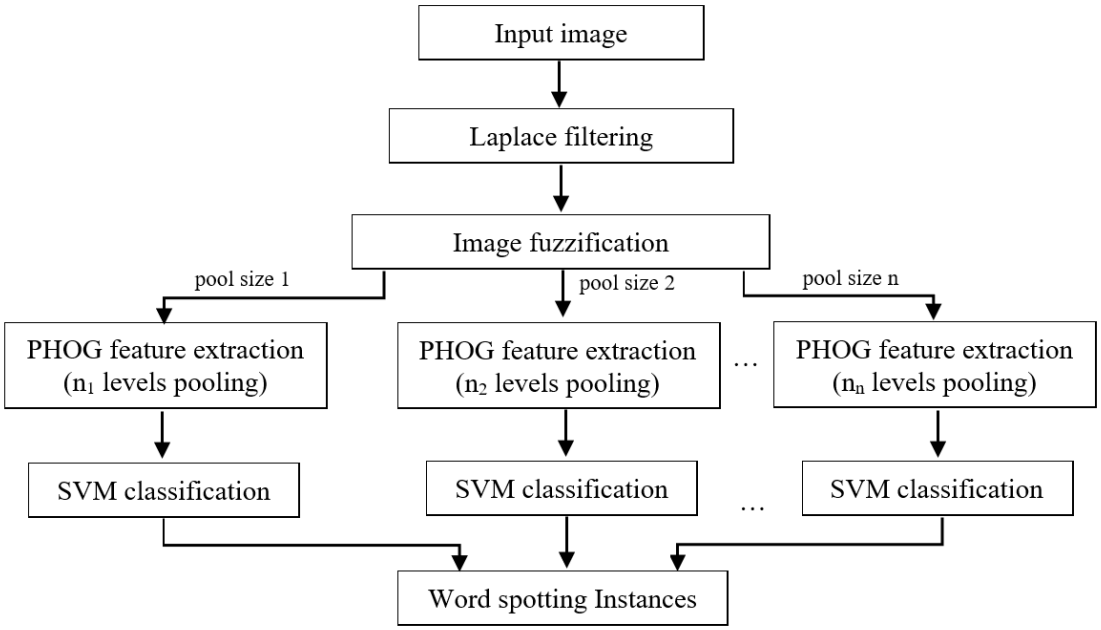


Fig. 1. The proposed QBE word spotting model.

II. Fuzzy Word Spotting

The proposed model comprises four steps that are pre-processing, image fuzzification, feature extraction, and word spotting step. First, the input word images are sharpened by 2D Laplace filter. Then, the sharpened images will be categorized into fuzzy sizes by image fuzzification. In each size fuzzy sets, the PHOG features are extracted with different pyramid levels. Finally, the SVM classifier is trained by represented word image PHOG features vector. In querying, the query word image is do the same procedure as training and the classifier will classify the class of the query word image and then return the word instances that have same class as the query word. The proposed model can be illustrated as Fig. 1.

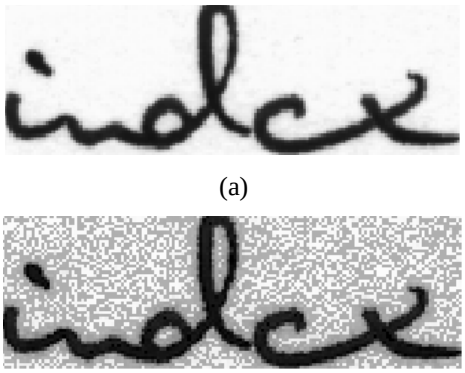
A. Pre-processing

Document word images could have the tiny word size like "I" word images which can make the feature extraction step faces the extraction issue. Thus, the word images that are tiny size (smaller than 16 pixels width or height) are resized to either 16 pixels width or 16 pixels height regarding the aspect ratio of the image using interpolation. Then, all word images are sharpened by using 2D Laplace filter to detect the edge of the images and subtract that edge images to the original ones to obtain the sharpened images, and the edge detected images can be achieved by convolution the images with the kernel in (1).

$$\Delta_{i,j}=\begin{bmatrix}0 & 1 & 0\\ 1 & -4 & 1\\ 0 & 1 & 0\end{bmatrix}*I\,, \qquad (1)$$

where I is original image and
$\Delta_{i,j}$ is output pixel in coordinate (i,j).

The different of the pre-sharpened image and post-sharpened can be visualized in Fig. 2.



(a)

(b)

Fig. 2. The visualization of difference of the (a) original image and (b) sharpened image.

B. Image fuzzification

The sharpened images are categorized using membership function in fuzzy logic. Fuzzy logic is the additional theory of the classical set. In classical set, the membership value of the data is binary. On the other hand, the membership value of the data in fuzzy set is real number in range 0 to 1 (membership value also can be greater than 1 or less than 0 but rarely used.). The general membership function in fuzzy logic are such as triangle function, trapezoidal function, Gaussian function, sigmoid function, and so on.

In this model, we used the trapezoidal function to partition the size pool. Size of the images (width pixels × height pixels) is used as the input in the membership function to add the images into each size pool. As it is fuzzy logic, the images that have size between two size pools are added into both pools. For each size pool, the membership function is tuned by using grid search on the membership function parameter.

The classes of word images in validation data are used for evaluating the size pools. The best partition of the size pool is the partition that achieve the highest number of trained samples in validation data. The membership function of images size of our model is presented in Fig.3.

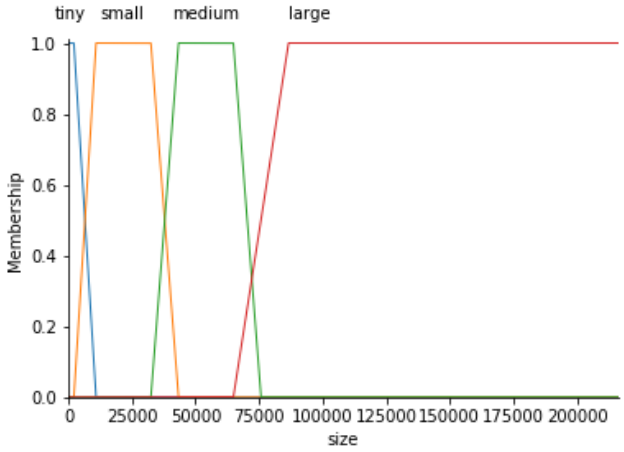


Fig. 3. The membership function of images size of the model. The pool size can be divided into 4 pools which are tiny, small, medium, and large. This partition was obtained by using grid search to tune the membership function.

C. *Feature extraction*

For image on each pool, the PHOG feature is extracted. PHOG feature is like the combination of pyramid representation and HOG feature. In pyramid representation, let n is the number of levels of pyramid. On the base (first level), the representation of this level is original image. On upper level (p level), the image is represented by divide the image into 2^p sub-images which is separated by vertical and horizontal line that make the output sub-images have the same portion. If the image size is $h \times w$, the window size (l pixels) and strides size (s pixels) on vertical and horizontal to construct the sub-images are defined as (2) and (3).

$$l_{vertical} = s_{vertical} = \lceil h/p \rceil , \quad (2)$$

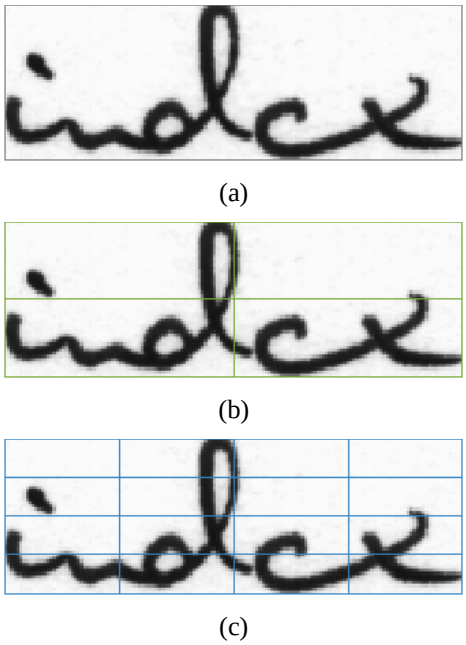
$$l_{horizontal} = s_{horizontal} = \lceil w/p \rceil \quad (3)$$

However, width and height of the image can be indivisible and make the window size and length are exceed the original size. Therefore, the image padding is employed to pad the image to the divisible size. The number of pixels to pad in vertical and horizontal are defined as (4) and (5).

$$pad_{vertical} = (p \times l_{vertical}) - h , \quad (4)$$

$$pad_{horizontal} = (p \times l_{horizontal}) - w \quad (5)$$

The padding pixels are divided for top and bottom in vertical by $\lfloor pad_{vertical}/2 \rfloor$ and $\lceil pad_{vertical}/2 \rceil$, and vice versa for horizontal. This scheme is repeated until $p = n$. Fig. 4. is the visualization of pyramid representation on pyramid level $n = 3$.



(a)

(b)

(c)

Fig. 4. The visualization of pyramid representation when pyramid level $n = 3$. (a) Sub-image in pyramid level $p = 1$ is a original image. (b) and (c) are the representation that when pyramid level p is greater than 1, the image are divided into 2^{p-1} in vertical and horizontal to produce the sub-images.

The HOG feature is gradient feature that collect the gradient of the image in the orientation bins. The HOG feature extraction is as follows. The image is divided into block and cell which block size is 2×2 cells. The horizontal and vertical gradients are computed to obtain the gradient magnitude and gradient orientation. The gradient computing is computed on each cell and then collect the gradients in the orientation bin(s) by interpolation. Gradient magnitude and gradient orientation can be obtained by (6) and (7) respectively, and horizontal and vertical gradients can be achieved by convolve the image with the following kernels in (8) and (9).

$$g = \sqrt{g_x^2 + g_y^2} , \quad (6)$$

$$\theta = \arctan\left(\frac{g_y}{g_x}\right) , \quad (7)$$

$$g_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I , \quad (8)$$

$$g_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I , \quad (9)$$

where g is gradient magnitude,
θ is gradient orientation,
I is input image,
* define as convolution operation, and
g_x and g_y are vertical and horizontal gradients, respectively.

After gradient computation, the feature is normalized by L1 or L2 normalization. The block is slid by 1 cell size and the process is repeated until the end of the image. The visualization on the HOG word image is illustrated on Fig.5.

Fig. 5. The visualization HOG feature on word image.

The PHOG feature is using the HOG scheme. Instead of dividing the image into block and cell, the PHOG use the sub-images from pyramid representation as cell and compute the gradients magnitude and gradients orientation. The final features vector is the concatenated of all HOG features in every sub-image.

D. Classifier and Word spotting scheme

In the word spotting scheme, the training word images are used for training in SVM with radial basis function (RBF) kernel. SVM is the classifier that creates the optimal hyperplane in feature space to separate the data, which maximum the distance between each class. The normal hyperplane is defined in (10).

$$g(\vec{x}) = \vec{w}^{T}\vec{x}+b\,, \tag{10}$$

when $g(\vec{x}_i) \geq 1$, if $y_i = 1$ and $g(\vec{x}_i) \leq -1$, if $y_i = -1$, and the optimal hyperplane require maximizing the margin z which can be defined as (11).

$$z = \frac{1}{\|\vec{w}\|} \tag{11}$$

Margin can be maximized by minimizing $\|\vec{w}\|$ using Lagrangian multiplier method and solved by quadratic programing. The example of optimal hyperplane is presented in Fig.6.

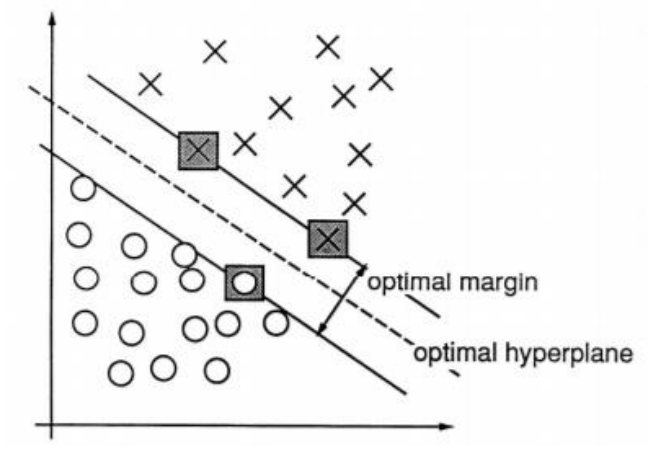


Fig. 6. The example of optimal hyperplane [13].

The classifiers for each size pool was created. Pairwise coupling [14] is applied to transform the output into probability of each class. The classifier is used for recognition the query word image. However, the query word can be not trained in the classifier. Hence, probability of the word image is used for validation and rejection the word image.

The rejection probability (*P(reject)*) can be achieved by tuning the probability in validation data. If the query word's probability is less than $P(reject)$, the model will answer as not found. On the other hand, if query word's probability is greater than or equal to *P(reject)*, the model will classify the trained word images to return the same class word instances. In case of the same image is split into 2 size pools, the membership value is multiplied by SVM probability to obtain the final probability. the model will classify the class that has the higher final probability.

III. Experimental Results

A. Data

The data that used for evaluation is IAM Handwriting Database 3.0 [15]. IAM Handwriting Database 3.0 are English handwritten manuscripts which contain 115,320 isolated word images with 657 writers. Before using the data, we filtered out the punctuation marks (e.g. period, comma, colon, semi-colon, and so on), since there are hardly ever to spotting the punctuation marks. After that, data were split into 3 set corresponding to the IAM dataset defined, which were training set, validation set, and test set that comprised 41,830 images, 12,192 images, and 11,603 images, respectively. Sample images from IAM Handwriting Database 3.0 are presented in Fig. 7.

Fig. 7. Examples of word images from IAM Handwriting Database [15].

B. Model parameters

In Section II, the word spotting method and required parameters of the model was presented. The setting of the parameters can be described as follows.

- The images sharpening that used 2D Laplace filter are relatively too sharp. To soften the sharpened images, the edge images are multiplied by 0.7 before subtract to the edge images are multiplied by 0.7 before adding to the original images.
- In image fuzzification, we defined the size pools into 4 sizes which were tiny, small, medium, large. the membership function that we used in each size pool was trapezoidal function which defined by 4 parts that

are lower limit, upper limit, lower support limit, and upper support limit. Note that, the lower bound of the image size is 0 and infinity is the upper bound. Therefore, the lower limit of tiny size pool is 0 and upper limit of large size pool is infinity. The detail of the parameters in each size pool are shown in Table I.

- Parameters on PHOG feature vectors can be divided into 2 parts, which are pyramid representation parameters and HOG feature parameters. In pyramid representation, the pyramid levels are grown by image sizes. In tiny pool, the pyramid level is 2 levels, and for each size pool increasing, the pyramid levels for that size is the previous pyramid levels added with one more level. In HOG feature, the gradient orientation is divided into 8 bins from 0 to 180 degree and normalized by L2-norm for every size pool.

TABLE I. TRAPEZOIDAL PARAMETERS FOR EACH SIZE POOL

Size Pools	Number of pixels (width × height)			
	Lower limit	***Lower support limit***	***Upper support limit***	***Upper limit***
tiny	0	0	2,163	10,815
small	2,163	10,815	32,446	43,262
medium	32,446	43,262	64,893	75,709
large	64,893	86,525	∞	∞

C. Model performance

We compared the model that used the image fuzzification and without image fuzzification. The parameters of the model without image fuzzification were using the same parameters as model with image fuzzification except the levels of pyramid representation are 3 for all images and only one SVM was used due to the single pool size. The accuracy and micro average precision were employed to measure the performance of the model. Accuracy, and micro average precision can be defined as (12), and (13).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}, \tag{12}$$

$$\text{Micro Average Precision}=\frac{\sum_{i=0}^{c} TP_i}{\sum_{i=0}^{c} TP_i+FP_i}, \tag{13}$$

where *TP* is true positive, *TN* is true negative, *FP* is false positive, *FN* is false negative, and *c* is number of classes.

The micro average precision is employed since the micro average collect the contributions of all classes to compute the precision that make the precision weighted by number of samples in each class. The accuracy in each size pool of the fuzzy model is presented in Table II.

TABLE II. ACCURACY OF THE FUZZY MODEL FOR EACH SIZE POOL

Size Pools	Accuracy
tiny	27.32%
small	38.34%
medium	54.20%
large	57.97%
Overall	35.11%

Table III shows the performance of fuzzy model and non-fuzzy model in term of accuracy and micro average precision. The result shows that the model with image fuzzification has better result in term of micro average precision.

TABLE III. MODEL'S RESULT COMPARISION

Models	Performance	
	Accuracy	***Micro average precision***
Model with image fuzzification	35.11 %	**35.11 %**
Model without image fuzzification	**40.14 %**	23.54 %

However, the accuracy of the model is slightly worse than another. The reason of that is the tested word images can be not trained in that size pool, that make the accuracy loss. On the other hand, if the tested word images were trained in the size pool, this can filter out the other classes in other sizes.

IV. CONCLUSION

The query-by-example word spotting model with image fuzzification was presented. This model used the PHOG feature and SVM with RBF kernel as a classifier. The point of the model is that the image fuzzification can be used for helping the classifier to filter out some words and make the model more stable. The result of the model with image fuzzification in term of micro average was 35.11%, which outperform the model without image fuzzification which was 23.54%.

However, the accuracy was slightly worse the model without image fuzzification. The model can be struggled from huge or tiny handwritten style which make the querying words are not in the suppose size pool, and the features vector of the model is still relatively underfit. If the query word images are in the correct size pool, the image fuzzification can help to filter the irrelevant classes. Thus, the image fuzzification can be used in the word spotting model for reducing the type one error and might trade with some accuracy of the model.

REFERENCES

[1] R. Manmatha, Chengfeng Han and E. M. Riseman, "Word spotting: a new approach to indexing handwriting," Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 1996, pp. 631-637.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.

[3] David G. Lowe, "Distinctive image features from Scale Invariant Keypoint," International Journal of Comupter Vision, pp. 1-28, January 2004.

[4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description ¨with local binary patterns: Application to face recognition," PAMI, pp. 1-15, 2006.

[5] Yang Bai, Lihua Guo, Lianwen Jin and Qinghua Huang, "A novel feature extraction method using Pyramid Histogram of Orientation Gradients for smile recognition," 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, 2009, pp. 3305-3308.

[6] Y. Xia, Z. B. Yang and K. Q. Wang, "Chinese calligraphy word spotting using elastic HOG feature and derivative dynamic time warping," Journal of Harbin Institute of Technology (New Series), 2014, pp. 21-27 vol. 21.

[7] R. Fusek and E. Sojka, "Gradient-DCT (G-DCT) descriptors," 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, 2014, pp. 1-6.

[8] D. Fernández, J. Lladós and A. Fornés, “Handwritten word spotting in old manuscript images using a pseudo-structural descriptor organized in a hash structure,” Pattern recognition and image analysis, Berlin, pp 628–635.

[9] M. L. Bouined, H. Nemmour and Y. Chibani, “New gradient descriptor for keyword spotting in handwritten documents,” *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, 2017, pp. 1-5.

[10] M. Mhiri, M. Cheriet and C. Desrosiers, “Query-by-example word spotting using multiscale features and classification in the space of representation differences,” 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 1112-1116.

[11] K. He, X. Zhang, S. Ren and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 1 Sept. 2015.

[12] F. Dernoncourt, Introduction to fuzzy logic, 2014, pp.5-17.

[13] C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning. vol. 20, no. 3, pp. 273-297, September 1995.

[14] T. F. Wu, C. J. Lin, and Ruby C. Weng. “Probability Estimates for Multi-class Classification by Pairwise Coupling,” Machine Learning. vol.5, pp. 975-1005. December 2004.

[15] U. Marti and H. Bunke. The IAM-database: An English Sentence Database for Off-line Handwriting Recognition. Int. Journal on Document Analysis and Recognition, vol. 5, pp. 39 - 46, 2002.

PLATOOL: Annotation-tool for creating Thai plagiarism corpus

Supon Klaithin, Pornpimon Palingoon, Kanokorn Trakultaweekoon, Santipong Thaiprayoon
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
{supon.kla, pornpimon.pal, kanokorn.tra, santipong.tha}@nectec.or.th

***Abstract*—In 2018, we created TPLAC-2018, a Thai plagiarism corpus, which was manually developed by seven human annotators. The annotators were told to simulate a plagiarized text by using four plagiarism cases, namely, copying and pasting, inserting, replacing and removing. This paper presents PLATOOL, an annotation-tool on a web-based platform, which helps all annotators to easily annotate plagiarism cases in 1052 suspicious documents chosen from 100,000 source documents on Thai Wikipedia website. The tool contains two components i.e., a web interface and a database. The first component is to interact between a user and PLATOOL web server. The second is designed to store, manage, facilitate access to plagiarism corpus. The result of annotation and discussion was concluded in the last section of the paper.**

Keywords—PLATOOL, annotation-tool, Thai plagiarism corpus

I. INTRODUCTION

Being similar to several languages in the field of plagiarism detection, the obstacle in developing an automatic Thai-plagiarism detection is that we cannot discover any real examples of plagiarized texts for training and evaluating our plagiarism detection algorithms [2] [3] [4]. Moreover, there is no any benchmark plagiarism-corpus for Thai language. Therefore, to increase the efficiency of our plagiarism detection, we decided to create the first benchmark Thai-plagiarism corpus by manually imitating the process and method of real text-plagiarizing.

However, the manual annotation from human annotators are very time-consuming and expensive, and lacks of a good tool to facilitate annotators. In addition, Thai language contains more complexities e.g., complex and long noun phrases and unsegmented sentences, that make it difficult to identify word and sentence boundaries. Hence, it is hard to determine the plagiarism cases in Thai, which are important linguistic factors to separate plagiarized texts from non-plagiarized texts. Before starting the full project of *TPLAC-2018* corpus, we did the pilot study in order to search for types and numbers of plagiarism cases being appropriate for the project. On the study, the three linguists were assigned to create their own plagiarized texts in suspicious documents selected from the same fifty source articles on the Thai Wikipedia website. After that, we finally chose four plagiarism cases namely, copying and pasting, inserting, replacing and removing, which were mostly used to change the non-plagiarized into plagiarized texts. These four plagiarism cases were used as an annotation guideline to annotate plagiarized texts in *TPLAC-2018* corpus.

To help speed up the process of manual annotation, we developed a special annotation-tool called *PLATOOL*: *Plagiarism-corpus Annotation Tool*, for easily creating our plagiarized texts in given suspicious documents. From our reviews on annotation tools, there are no systems specially designed for annotating plagiarism cases in plagiarized texts. Accordingly, the tool is designed to help annotators by applying web development technologies to support the annotator's operations. For example, we used string matching, a searching technique, to find a position of fragment in both source and suspicious document. Moreover, we developed user-friendly GUI interface to reduce the complexity of annotation such as time-consuming tasks, a lot of procedures in searching for relevant documents. Our interface design provided annotators with the overview of a selected fragment during the process of plagiarism-case annotation.

The remainder of this paper is organized as follows. Section II presents some related works. Section III shows our system design. Section IV presents fragment searching in the documents. Section V presents the results of data annotation. Some discussion and conclusion are proposed in the last section.

II. RELATED WORK

This section presents two parts underlying the concept and process of our *PLATOOL* development. See more details in A. and B. as follows.

A. Plagiarism corpus and plagiarism cases

A lot of research on plagiarism detection pay more attention to simulate plagiarism corpora by using both manual annotation and machine generation. However, most of them focused on developing plagiarism corpora by using automatic models. For example, [4] and [5] used simple algorithms for separating plagiarized from non-plagiarized texts. It consisted of a small number of annotated cases and was not directly

designed to support plagiarism detection evaluation[1]. The PAN series (2009-2014), plagiarism corpora and plagiarism cases were mostly developed by using artificial plagiarism cases from machine learning algorithms. [2] and [10] concluded that the corpora used to train and evaluate the plagiarism detectors require further improvement because of many annotation errors occurring from artificial corpus-developing. The artificial process could not be enough for generating real plagiarized texts [8]. It was accepted that manual annotation in simulating plagiarized corpora is the realistic need to improve the automatic detectors. In the early stage of simulated corpus, METER (MEASuring TExt Reuse) corpus was created by humans in the field of English newspaper [9]. However, the texts from this corpus was not the real plagiarized texts. Some research tried to use a hybrid method (a manual annotation cooperating with an automatic model), to simulate various types of plagiarism cases. For example, [4] proposed 4 types of plagiarism cases i.e. exact copy, near copy, modified copy and text manipulation (paraphrasing). However, their plagiarism cases were more appropriate for the Persian language than other languages. Three types of copying were explained in the same way and rather difficult to be separated from each other. Moreover, the description about paraphrasing and modified copying was not clear in the details of structure modifications.

The P4P was the Paraphrase for Plagiarism corpus that was developed as a portion of PAN-PC-10 corpus. According to the concept of paraphrase typology, [1] proposed some significant linguistic phenomena related to the process of changing non-plagiarized text sections into plagiarized texts. They called these linguistic mechanisms as plagiarism cases and classified them into four classes i.e. morpholexicon-based, structure-based, semantic-based, and miscellaneous changes. According to their concept of classifying plagiarism cases, we also classify our Thai plagiarism cases in our *TPLAC-2018* into three classes , namely, lexicon-based, structure-based, and semantic-based classes. Finally, we adopted some of their plagiarism cases in each class, and reorganized as four plagiarism cases for our project. See more details of four plagiarism cases as follows.

- Copying and pasting: the way to copy a fragment (a fragment means a word/a phrase/a sentence) selected from a source document and paste it into a suspicious document without any modifications of the selected text from source document.
- Inserting: the way to add a word/a phrase/a sentence in a fragment of suspicious document.
- Replacing: the way to replace a word/a phrase/a sentence in a fragment of suspicious document with a word/a phrase/a sentence that has the same meaning.
- Removing: the way to delete a word/a phrase/a sentence from a fragment of suspicious document.

1-The annual PAN (Plagiarism, Authorship, and Social Software Misuse) workshop and competition series on textual plagiarism detection and authorship analysis has been held since 2009

All the plagiarism cases mentioned above were used as a guideline in the *PLATOOL*.

B. Annotation Tools

TagTick [11] is an annotation tagging tools which a fully functional annotation tagging environment over full-text index Apache Solr. TagTick consists of two main modules i.e., the Virtualizer and User Interface. TagTick Virtualizer implements functionalities for real-time bulk-(un)tagging in the context. User Interface implements user interface for user to produce tagged information spaces.

Surfing Notes[12] is a cloud-based system which allows users to annotate and archive the webpages for personal uses. The system is an integrated web annotation and archiving tool in which registered users can access and manage their private webpage libraries. The Surfing Notes interface are developed following the Model View Controller(MVC) architecture for organizing libraries, managing and viewing webpage archives and annotations. The longest common subsequence (LCS) finding the difference between two text. are used as the base of change detection in the Surfing Notes sytem.

PAS (Proceural Annotation System)[13] is a tools for student and teacher. The tools helped the students in making procedural annotations on a computer software training course and helped the teacher to giving tips and comments to students. Text annotation were also enabled to help student to describe the meaning of the screenshot and the text annotation. Moreover, the teacher can evaluate student's task by inspecting the screenshot and the text annotation.

The GATE (General Architecture for Text Engineering) is the famous and powerful annotation tool, developed at the University of Seffield since 1995. Rang et al.,[14] proposed a method for sematic annotation of semi-structured document (Microsoft Word and Excel) using GATE. GATE can be used to find tokens, space tokens and setences in the document. The main purpose of this tool is to annotate the information in Word and Excel documents with either sematic information or to highlight some necessary information.

A web-based tool for visualization and collaborative annotation of physiological database by using Java applet front-end for visualization and back-end CGI (Common Gateway Interface) script to facilitate user interactive and annotation [15]. The client-side processes requests to the server by sending a CGI request to add or delete the annotation. The front-end interface is designed as a web-based system for visualization using SVG, an XML-based language.

Arabic web application which annotate [16], automatically, Hadiths texts by associating tags with the POS of word in Hadiths corpus. The Arabic annotation web application consists of three spaces namely, an annotator space, an administrator space and a public space. In annotator space, the annotator can annotate Hadiths including the segmentation of word, the POS tag, the root and the pattern. The administrator dispatches the Hadiths to the annotators and assigns each theme to annotator. Furthermore, the administrator can download and generate an annotated corpus in XML format. Finally, the

public space is a page for downloading the corpus by the public user.

From our reviews on various text annotation tools, we found that we can apply some of those techniques to our tool design. For example, the visualization, which implements user interface for an annotator, was designed the interface using MVC framework based on Java programming language. Moreover, some techniques are used to annotate text: highlight, strikethrough and finding string algorithms. Therefore, this paper proposes a system which applies the techniques from the above literature. See details of our tool system in Section 3.

III. System Design

In this section, we describe the system architecture, annotator, and administrator operation. See details in the following sub-sections (Figure. 1).

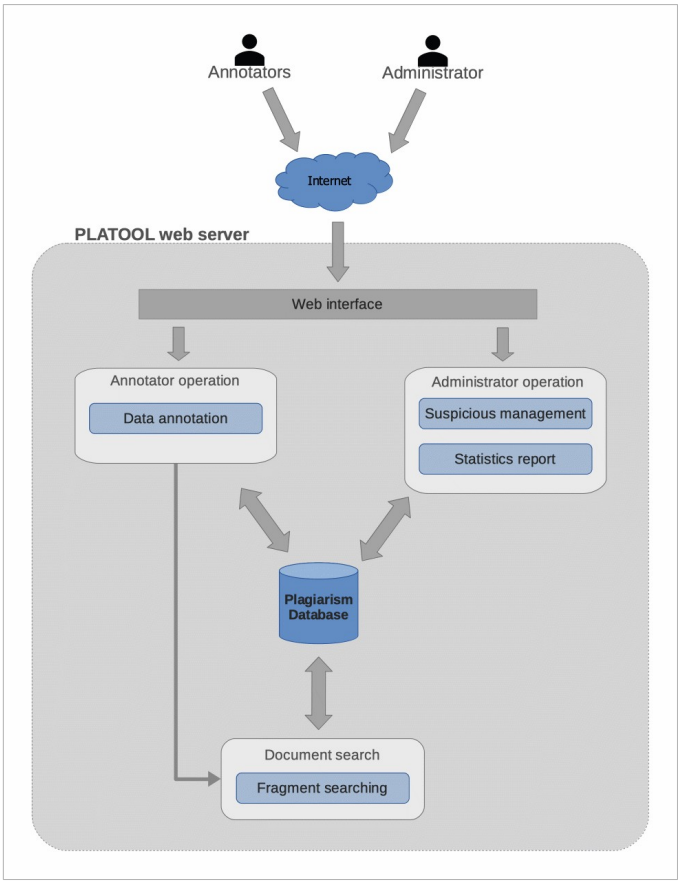


Figure 1. PLATOOL system architecture

A. System architecture

PLATOOL is a web-based annotation tool which developed on web application technologies. The architecture of our tool consists of two components i.e., web interface and database (Figure. 1). The details is given in the following sub-sections.

1) Web interface

For the web interface (Figure. 1), it runs on Apache web server in order to help users access through the internet network. We used AngularJS for developing the front-end application. The back-end was implemented with PHP script for connect to database.

2) Database

The plagiarism database stored in MySQL, is an essential component of PLATOOL website. It contains two types of information, namely, annotators' personal data and plagiarism-corpus information. See more details as follows

a) Annotators' Personal data: This is any personal information relating to identifying an annotator e.g., username, password, email.

b) Plagiarism-corpus information: This is any information relating to data collection and annotated plagiarized-texts in *TPLC2018* development. They were collected in three kinds of table i.e., suspicious table, fragment table, and source table. See the details as follows.

- *Suspicious table:* This table contains two main elements of suspicious table i.e., contents and statistics. The first part "contents" contains both plain text and HTML. The plain text was used for training, testing, and evaluating the plagiarism detection. HTML was used to display a web page in any web browser. Therefore, it helps the annotators to easily annotate data in PLATOOL web page. The second part "data reports" shows statistics of fragments and plagiarism cases such as total fragments and cases, case types (e.g., copying and pasting, inserting, replacing and removing).
- *Fragment table*: This table contains any information of plagiarized texts (plagiarized texts are called fragments in this paper.) such as suspicious document ID, modified text, fragment length, and fragment position in suspicious document.
- *Source table*: This table stores the information of source documents i.e., selected text, start and end position of selected text or fragment in the source document.

B. Annotator operation

To simulate the plagiarized text, this section presents the process of annotation from human annotators. The process consists of five steps i.e. marking the position of fragment in the suspicious document, searching and selecting a source document, selection a text from the selected source in source document, modifying the selected text from the source document, inserting the selected text (called fragment) into the suspicious document. More details are described as follows (See Figure 3).

- Position marking: this process is marking. (1)
- Searching source: searching and selecting a source document. (2)

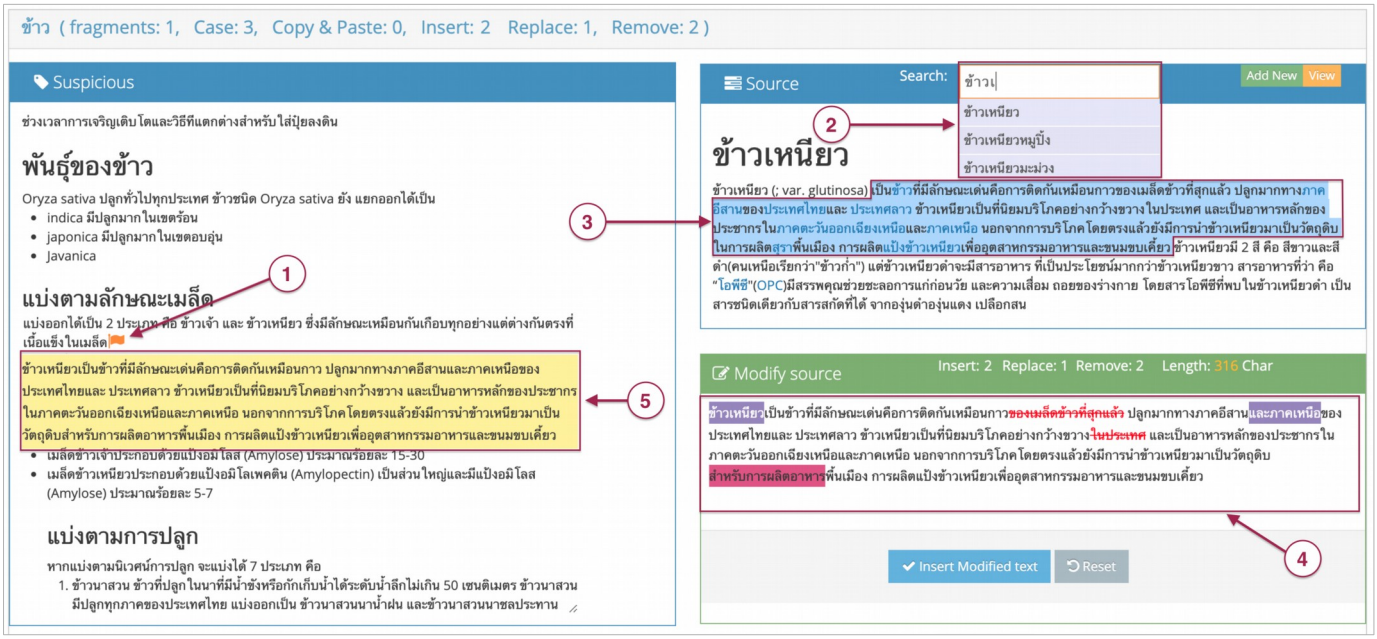


Figure 2. PLATOOL user interface

- Fragment selection: selecting a text from the selected source in source document (called fragment). (3)
- Fragment modification: modifying fragment from the source document. (4)
- Fragment Insertion: inserting fragment into the suspicious document. (5)

C. *Administrator operations*

Administrative operation is a last component that helps manage the system of PLATOOL. The administrator can perform the following operations:

- *Suspicious management:* to assign a suspicious document to users.
- *Statistic report*: to report the result of annotation corpus such as plagiarism statistics, plagiarism type in a pie chart and personal progress.
- *User management*: to manage the user accounts including password resets, creating, changing the status and deleting user.

IV. Fragment Searching in Document

Fragment searching is essential to the system of automatic plagiarism detection, especially to trace forward and backward between source and suspicious document. For example, If we know the position of fragment in suspicious document, we can know the position of plagiarized texts in the suspicious document. Moreover, we can go backwards into the position of those plagiarized texts in the source document. This section proposed our fragment searching based on string matching algorithms. It helps identify the position of fragment in the source and suspicious document. The results from searching, such as fragment length, beginning and end point of fragment, are immediately updated to the plagiarism database (see Figure 1).

A. *Fragment searching in source document*

The contents of source document was tokenized into characters and wer stored in array of character. After deriving string into array, the system will be compared between source and fragment by using string matching technique. Finally, the system returned important data of fragment such as start and end position in source document, fragment's length, as results from matching process and compare. In the final step, the system will be updated the annotated message and new words in the database.

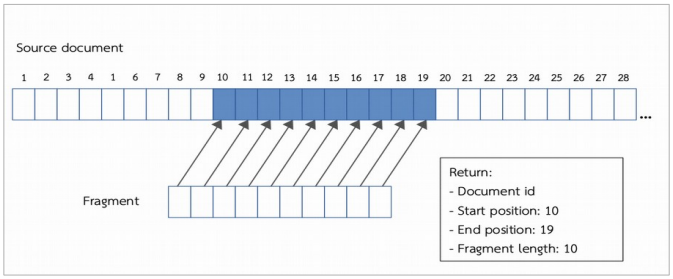


Figure 3. Fragment searching in source document

B. *Fragment searching in suspicious document*

The final process of PLATOOL is to find all fragments (plagiarized sections) in the suspicious documents by using string matching techniques. For example, annotators created plagiarized text in the suspicious document (see Figure. 4) and the corresponding fragments in source document. When the document is submitted to the system. The four fragments, A, B,C and D, can be found in the document by using string matching techniques. All search results, including length, start, and end position, will be updated the fragment table in the database. Moreover, the document's statistic, such total fragment, total case, and case type, update in document table.

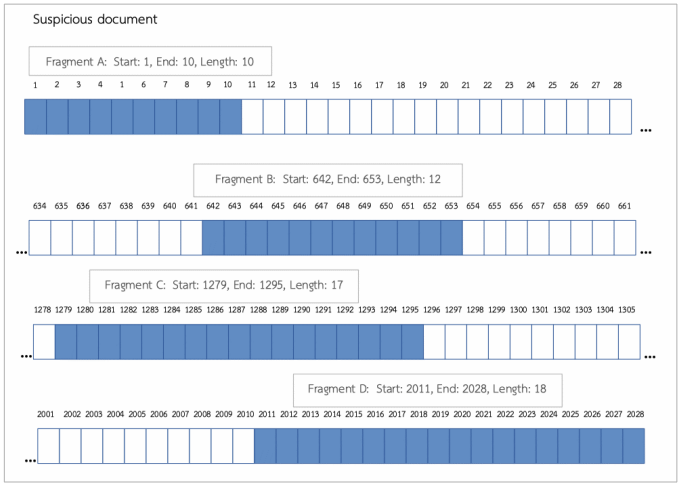


Figure 4. Fragment searching in suspicious document

V. The Result of Data Annotation

This section shows the statistics of Thai plagiarism corpus. Finally, we get 1052 suspicious documents, comprising of 4983 fragments. All fragments contain 4 plagiarism cases i.e., 2137 copying and pasting, 2042 inserting, 2123 removing, 1961 replacing. The length of each suspicious document starts from 790 to 202,375 characters.

TABLE I. Statistics of Plagiarism case

Plagiarism cases	Frequency
Copying and pasting	2137
Inserting	2042
Removing	2123
Replacing	1961

VI. Conclusion and Future Work

In this paper we presents PLATOOL, a web-based annotation tool for creating Thai plagiarism corpus. This web application allowed the annotators to insert a plagiarized text in the suspicious document by using four plagiarism cases, namely, copying and pasting, inserting, replacing and removing. According to the results of the annotation, we concluded that PLATOOL helped annotators to easily annotate suspicious document. Moreover, PLATOOL helped collect statistical of plagiarism corpus annotation. Furthermore, we found that PLATOOL was used to bridge the gap between the annotator operations and programming requirements.

In the future, we plan to increase more features in our existing tool, PLATOOL. For example, this tool will help investigate fragment modifications e.g., how to modify the fragment and to collect the history of modifications. Additionally, annotators can modify the text before and after a fragment in a suspicious document in order to create better natural plagiarized texts.

Acknowledgment

We would like to thank Alisa Kongthon, Choochart Haruechaiyasak and Sawit Kasuriya for their supports and suggestions in this work, and are also thankful to National Electronics and Computer Technology Center in Thailand, for funding us one and a half year of developing TPLAC-2018 corpus.

References

[1] A. Barrón-Cedeño, M. Vila, M.A. Marti and P. Rosso, Plagiarism Meets Paraphrasing: Insights for the Next Generation in Automatic Plagiarism Detection, Association for Computational Linguistics, 2013.

[2] A. Barrón-Cedeño, M. Potthast, P. Rosso, B. Stein and A. Eiselt, "Corpus and Evaluation Measures for Automatic Plagiarism Detection", In proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta, 17-23 May 2016.

[3] A. Zaid, S. Tiun and M. Abdulameer, "Cross-language Plagiarism of Arabic-english Documents Using Linear Logistic Regression", Journal

[4] of Theoretical and Applied Information Technolog, vol 1, pp20-33, January10, 2016.

[5] F. Mashhadirajab, M. Shamsfard, R. Adelkhah, F. Shafiee and C. Saedi. A Text Alignment Corpus for Persian Plagiarism Detection, FIRE, 2016

[6] M. Mansoorizadeh, and T. Rahgooy, Persian Plagiarism Detection Using Sentence Correlations, FIRE, 2016.

[7] P. Clough, Plagiarism in Natural and Programming Languages: An Overview of Current Tools and Technologies, Department of Computer Science, University of Sheffield, UK, Technical Report CS-00-05, 2000.

[8] P. Clough, Measuring Text Reuse, PhD thesis, University of Sheffield 2003.

[9] P. Clough and M. Stevenson, Developing A Corpus of Plagiarised Short Answers, Language Resources and Evaluation, 45 (1), pp. 5-24, 2011.

[10] R. Gaizauskas, J. Foster, Y. Wilks, J. Arundel, P. Clough and S. Pia, Scott, The METER Corpus; A corpus for analysing journalistic text reuse, In proceedings of the Corpus Linguistics Conference 2001, pp. 214-223, 2001.

[11] M. Artini, C. Atzori, A. Bardi, S. La Bruzzo and P. Manghi, TagTick: A Tool for Annotation Tagging over Solr indexes, IEEE/ACM Joint Conference on Digital Libraries, 2014.

[12] S. He and E. Chan, Surfing Notes: an Integrated Web Annotation and Archiving Tool, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technoloty, 2012.

[13] W. Hwang, C.Wang, S. Pan and J. Dong, An Annotation Tool to Support Procedural Knowledge Learning, 2016 International Conference on Educational Innovation through Technology (EITT), 2016.

[14] G. R. Ranganathan, Y. Biletskiy and A. Kaltchenko, Semantic annotation of semi-structured documents, Canadian Conference on Electrical and Computer Engineering, 2008.

[15] M.B. Oefinger and R.G. Mark, A web-based tool for visualization and collaborative annotation of physiological databases, Computers in Cardiology, 2005.

[16] R. Ayed, A. Chouigui and B. Elayeb, A New Morphological Annotation Tool for Arabic Texts, 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), 2018.

The Development of Intelligent Models for Health Classification

[1,2]Wattanapong On-num, [3]Narumol Chumuang and Chairit Siladech[4]

[1]Department of Physical Education, Faculty of Education,
[2] Department of Industrial Technology Management, Faculty of Industrial of Technology,
[3]Department of Digital Media Technology, Faculty of Industrial of Technology,
[4]Department of Educational Research and Evaluation, Faculty of Education,
[1,2,3,4]Muban Chombueg Rajabhat University

[1,2]aongdy11@gmail.com.[3]lecho20@hotmail.com and [4]siladech9@gmail.com

***Abstract*—** This paper focuses on development intelligent health classification models for individuals with C4.5 techniques, which collected physical fitness data of those who exercise in Ratchaburi province. A total of dataset were used for creating models 376 records, and 19 attributes of personal information consists of age, sex, weight, height, pulse, upper blood, lower blood, weight bike, hand womenorce, leg stretch, triceps, biceps, suprailiac, subscapular, leg, grip, womenlex, lung capacity, O^2 and two classes. The classification is divided into five classes, including the best health, good, normal, low and very low. In our experiment, the researcher divided the data set into two groups: training and testing and designed the test using 10-fold cross-validation method. The accuracy rate of C4.5 shown 100% .

Keywords— Development, Intelligent Models, Health, Classification

I. INTRODUCTION

The National Health Development Plan (2017-2021) discusses the health situation of people that have a longer life, but lose more happy years sick and die with preventable diseases [1] – [5]. From individual patient data for treatment and sleeping in the hospital and have health insurance from three main funds, namely universal health coverage medical welfare for civil servants and families and social security in the last 10 years [6]. It was found that the trend of non-communicable diseases of Thai people has increased steadily and most patients are suffering from hypertension, heart disease, diabetes, cancer and kidney failure [7], [8], which is a chronic disease that can prevent in the beginning. Which people must take care of their own health by having the correct consumption behavior weight control, exercise regularly [9], avoid drinking alcoholic beverages, relaxation and stress control, relaxation in daily life including regular self-examination of health and health facilities [10]. In addition, people must also be involved in determining social coexistence measures, such as non-smoking areas, selling safe food, exercise campaign comprehensive, reduction of causal factors including consumption behavior, four factors from the environment traffic and daily lifestyle, which will lead to a healthy society together [11]. From such a situation led to the reform of the country in public health. One of the key issues identified in the public health reform plan is the reform of the health service system. In order to have a standardized health service system cover and link at all levels is fair seize the service center focus on health promotion and disease prevention. The development holistic health manpower and health communication [12].

Office of Sports Authority of Thailand is an organization that plays an important role in promoting physical health, recognizing the importance of public health in Ratchaburi therefore the implementation scope in sports science service project by testing promoting physical fitness for athletes and the general public check for treatment, rehabilitation, injury from sports and exercise including collecting health information of people in Ratchaburi Province. The amount of dataset surveyed is increasing every year but there is still lack of data management which can be predicted for planning activities or budgets or utilizing various benefits. This paper presents the health classification of people to be used to create intelligent models for the classification of population health. Fig.1 illustrate an example of running for health activity, which there are others activity like as swimming, play football, basketball, tennis etc.

Fig.1 Example of the healthy activity.

Artificial Intelligence (AI) indicates the intellectual ability of a machine [13]. The standard of AI is measured by human intelligence, considering reason, speech and vision but this standard is still very far away from the present. Machine Learning (ML) is the best tool available today to analyze, understand, and find patterns of data [14], [15]. One of the key concepts under ML is that computers can be trained automatically, which can be done purely or impossible for humans to do and there are still clear loopholes from previous analysis that machine learning can make decisions with little human intervention. ML uses data to pass it on to an algorithm that can understand the relationship between incoming and outgoing data [16]. It can predict the value or type of new information. With these reason, the contribution of this paper is to apply ML for creating the intelligent model for health classification.

The organization of our paper is sequential by begin with the related of our works and the literature review in section 2. The section 3, our methodology was described. The experimental and conclusion are in Section 4 and 5 respectively.

II. Related work and litterature review

A. Data Mining

Nowadays, data mining has been applied in many types of business both in business and in helping executives make decisions in science and medicine including economic and social aspects. It is like an evolution of data collection and interpretation from the original that has been stored simple information into storage in a database that can retrieve information to use until data mining. It able to discover knowledge hidden in information, which the purpose of data mining can search for important information that is mixed with other data in the database, not just randomly called Knowledge Discovery in Database (KDD) or search for knowledge with 5 types of data which are 1) find relationship rules 2) classification and forecasting 3) grouping data 4) finding outliers 5) trend analysis Data mining is not about fetching new data. Instead, it is about drawing conclusions from the patterns and new knowledge from data which is already collected or recorded. [15], [16].

B. C 4.5 Algorithm

C 4.5 is the process of creating data management models into groups that are assigned by creating rules to help make decisions based on available data. For predicting the occurrence of unrealized data by presenting rules derived from data classification techniques popularly presented in the form of a tree, which is called a tree for decision-making. The decision tree is a structure that shows the rules derived from data classification techniques. The decision tree will look like a tree structure. Where each node represents an attribute, each branch shows the test conditions and the leaf node shows the defined group. Techniques used to create trees from existing data sets. The nature of building trees will use rules in the form of "if conditions and results" such as "if diligent reading a book and passing the exam" etc. The decision tree is a very popular technique. Because the results are easy to understand decision tree techniques will restrict 1 dependent variable data per model. If you want to predict multiple as with Naïve-Bayes, most of the decision tree techniques will not support continuous data modeling. Therefore, data must be identified to be discontinuous by creating a recursive top-down tree by dividing the big problem into small problems or divide-and-conquer. The pseudo code for creating C 4.5 model as shown in Fig.2.

The pseudo code for creating c 4.5 model.

1) create a node N;
(2) if samples are all of the same class C then
(3) return N as a leaf node labeled with the class C;
(4) if attribute-list is empty then
(5) return N as a leaf node labeled with the most common class in samples; // majority voting
(6) select test-attribute, the attribute among attribute-list with the highest information gain;
(7) label node N with test-attribute;
(8) for each known value ai of test-attribute
(9) grow a branch from node N for the condition test-attribute= ai;
(10) let si be the set of samples in samples for which test-attribute= ai;
(11) if si is empty then
(12) attach a leaf labeled with the most common class in samples;
(13) else attach the node returned by Generate decision tree (si, attribute-list, test-attribute);

Fig.2 The pseudo code for creating c 4.5 model.

C. Related Research

Data mining play a vital role in healthcare industry. It is predominantly use for detection and prediction of diseases. Various researchers acknowledged the fact that there is a demonstrated need for the use of data mining in healthcare.

Soni et al. [17], proposed the associative classification approach for better analyzing the healthcare data. The proposed approach was the combined approach that integrated the association rules as well as classification rules. This integrated approach was useful for discovering rules in the database and then using these rules to construct an efficient classifier. In this research, experiments on the data of heart patients were performed in order to find out that accuracy of associative classifiers was better than accuracy of traditional classifiers. Apart from this, the research also generated the rules using weighted associative classifier.

M. A. Hassan, M. E. Shehab and E. M. R. Hamed [18] proposed a comparative study of classification algorithms in e-health environment. The result found that data Mining can be used in enhancing the quality of the medical services offered through analyzing data and discovering hidden patterns and relationships that can enhance and even change the treatment methods adopted. In this paper ten classification algorithms are applied on a patient's dataset obtained from a public hospital's data base that contains patients both medical and personal information needed for diagnoses and treatment decisions. These algorithms are analyzed using a data mining tool and a comparative study is undertaken to find the classifier that performs the best analysis on the dataset obtained using a set of eight performance metrics to compare the results of each classifier.

S. Rallapalli and T. Suryakanthi [19], proposed a prediction model using a scalable randomized forest classification algorithm that can specify the precise classification rate for the risk of diabetes. By identifying strong indicators for accurate predictions is a challenging task. From the factors considered for predictive analysis, models and algorithms need to be studied. Classification algorithms such as Naive Bayes, Linear Regression General supplementary models, random forestry, logistic regression, hidden Markov models must be considered for the development of risk prediction models for diabetic patients. Which the prediction results with various algorithms are satisfactory.

G. Wang, Z. Deng and K. Choi [20], presented predictive modeling when the input feature of the model contains missing data This method also determines the influence of the characteristics along with the missing values in the classification accuracy at the same time, using a long wait-without validation strategy. Assess the effectiveness of this method using QOL for the elderly by using health data collected in the community. This data set relates to

demographic data, socioeconomic status, health history, and health evaluation results of 444 elderly people in the community, with 5% to 60% of the missing data in some input features. QOL is measured using World Health Organization standardized questionnaire The results show that the proposed method outperforms the four general methods for managing missing data cases, property deletion, average determination and neighboring neighbor determination with predictive accuracy. The average is 0.7418. It may be a promising technique for solving missing information in community health research and other uses.

III. Research Methodology

In this research, the research team has determined the steps into four main steps as shown in Fig. 3.

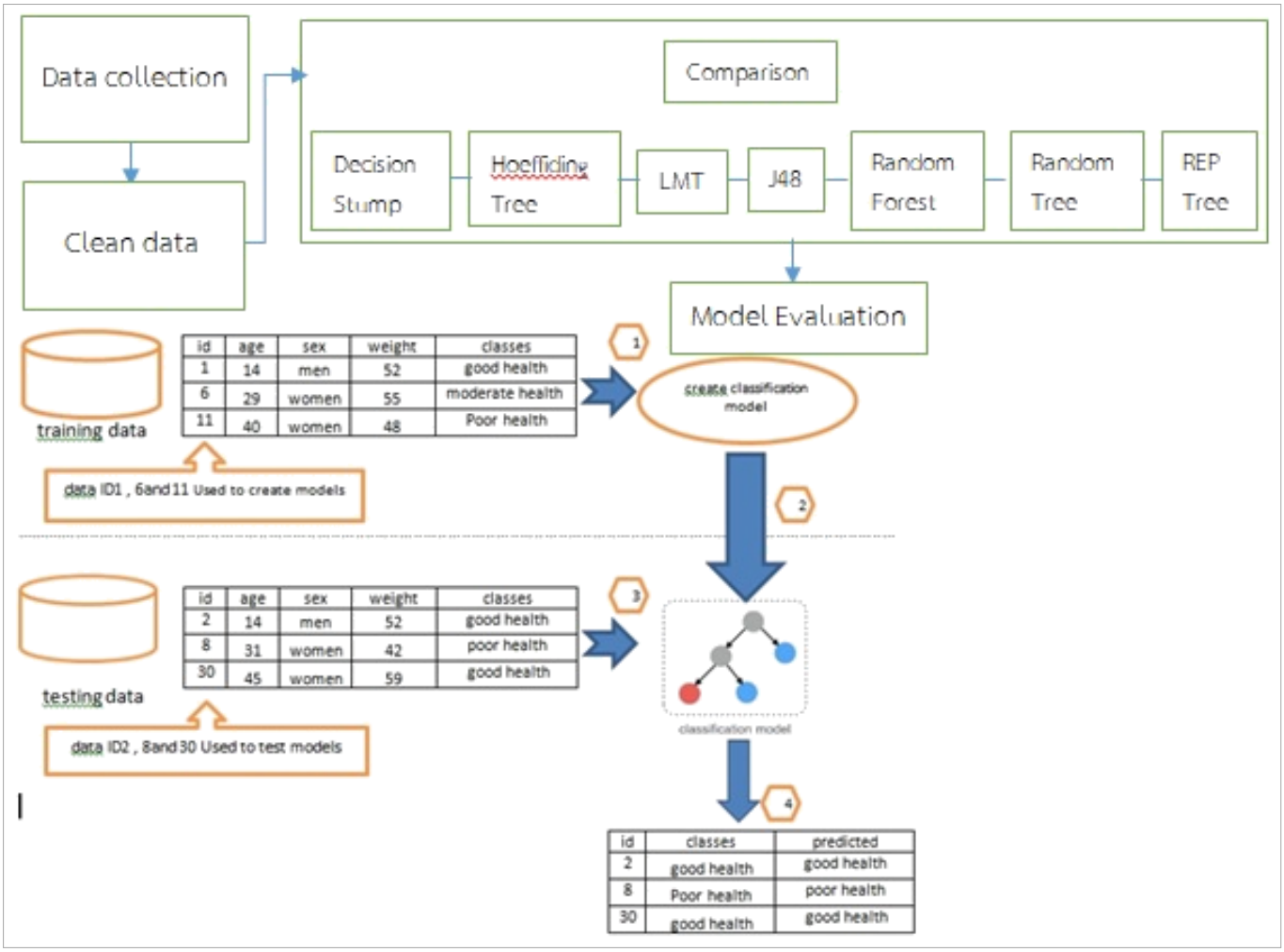


Fig. 3. The overview of development of intelligent models for health classification.

A. Data Collection

The dataset of physical fitness of people in Ratchaburi province. Who exercised at Ratchaburi Stadium Which is the information that the Sports Authority of Thailand Ratchaburi province recorded a total of 376 people and collected all the relevant factors in 19 factors. The last attribute was divided into 5 layers Is the best health, good, normal, low and very low number of 376 people for use in predicting the effects of physical images using data mining with decision tree techniques to compare the forecasting efficiency of the physical performance of the general public in the province Ratchaburi.

B. Data Cleaning Procedures

Data cleaning is the process of checking and editing (or deleting) incorrect data items from a data set tables or databases which is the cornerstone because it means imperfection inaccuracy unrelated to other data, etc. Therefore, must replace, update or delete these incorrect information to provide quality information Data cleaning happened because there is a data inconsistency this may be caused by errors in data recording, data transmission, or the interpretation of data stored differently. More requiring integration with other databases such as data warehouses or multiple databases Therefore has a high chance of being born "unclean information" is up

This step of our system for cleaning the data by bringing a total of 372 records, the number of attributes 19 attributes, to check the integrity of the data to make the data complete and quality. In order to correctly analyze the data the steps to clean the data are as follows.

TABLE I. SAMPLE OF DATA SETS RELATING TO PHYSICAL FITNESS OF THE GENERAL PUBLIC IN RATCHABURI PROVINCE THAT HAS ALREADY BEEN CLEANED

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	age	sex	Weight	height	Pulse	pper_bloo	ower_bloo	eight_bik	d_women	eg_stretc	Triceps	Biceps	Suprailiac	ubscapula	Leg	Grip	womenlex	ng_capaci	O2	criterion
2	14	men	52	167	51	107	68	2	0.69	2.62	6	3	4	7	136	36	10	3600	130	best
3	14	women	33	142.2	85	112	65	1	0.45	2.3	8	8	6	6	76	15	-10	1700	146	best
4	22	women	45	160	78	119	67	1.5	0.67	2.33	9	5	11	13	105	30	4	2100	129	best
5	23	women	50	165	76	103	64	1	0.58	2.4	6	4	7	5	120	29	14	2500	127	best
6	29	women	49	156	65	107	76	1.5	0.49	2.14	15	14	14	9	105	24	22	2600	122	best
7	29	women	55	157	80	135	86	1	0.58	2.95	6	4	13	10	162	32	17	3100	129	best
8	31	women	42	155	89	114	68	1.5	0.71	1.79	14	11	13	15	75	30	15	2100	129	best
9	37	women	44.5	155	85	106	56	1	0.67	2.54	6	4	9	8	113	30	8	2900	136	best
10	40	women	46	156	76	131	82	1	0.61	2.26	13	3	11	7	104	28	14	2100	125	best
11	40	women	48	160	66	115	68	1.5	0.6	1.77	12	4	9	8	85	29	16	2000	122	best
12	40	women	48	162	86	99	61	1	0.58	1.71	5	4	15	10	82	28	8	1900	121	best
13	40	women	48	166	98	108	55	1	0.62	1.77	6	10	15	7	85	30	8	2300	145	best
14	40	women	59	165	97	125	61	1	0.64	2.42	10	6	15	18	143	38	23	2000	129	best
15	40	women	60	155	80	120	62	1.5	0.5	1.77	15	7	22	18	106	30	18	1800	121	best
16	41	women	46	163	111	99	67	1	0.65	2.33	10	3	10	10	107	30	14	2300	140	best

The cleaning data with 372 dataset by consists of attributes shown in Table I. In our work, the fitness by using data mining with decision tree techniques to compare the forecasting efficiency of the physical performance of the general public in Ratchaburi Province.

C. Data Classification

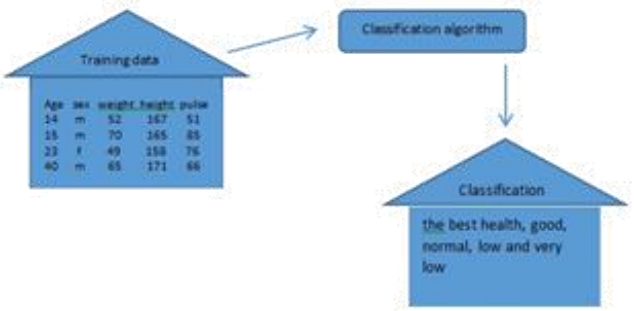


Fig. 4. The sample of the data training process.

Data classification consists of two main processes as shown in the example in Fig.4 that is a test of physical fitness. From Fig.5, it is the process for creating a data separator from an input data set in which each record of data considered consists of a set of attributes that characterize the person doing physical fitness tests and the category of that person's health level by the process of creating data separators by training resulting from the implementation of procedures. The methods for health classification and processing of data. One X record data in the data set considered. It consists of a set of attributes $X = (x_1, x_2, ..., x_n)$, attributes that indicate the attributes of data the record in addition, records also contain more information. One attribute that indicates the category of dataset is discrete. In which the data set is the input for creating the classifier training data set for analyzing physical health data that does not contain the data category that indicates that each person has what level of health. We will be able to analyze the data by linking the records of the physical images that are similar or identical to the same group as follows.

The second step of data classification as shown in Fig. 5 will run the data resolver that built from step one to classify the data by the beginning data classifiers will be tested and valuable accuracy by the accuracy of the classification data the value is satisfactory or acceptable. We will use the data separator in Identifying or indicating unknown data categories using the 10-fold cross validation method that describes in our experimental and the results.

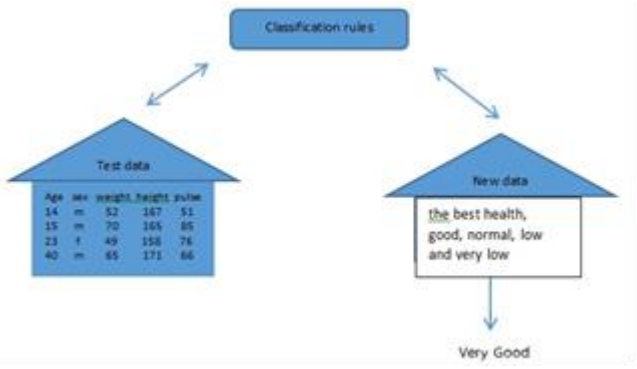
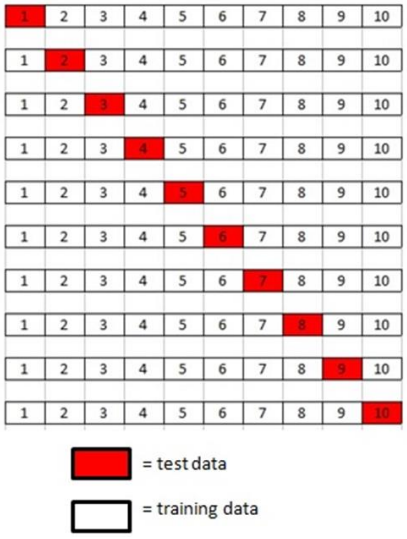


Fig. 5. Learning from information to create a data classification test to measure accuracy.

For our purposed methodology, the development of intelligent models for health classification were designed for highest efficiency by comparison with many machine learning algorithms. The detail about the experimental process describes in the next section.

IV. THE EXPERIMENTAL AND RESULTS

In our system, we design the experimental for health classification with dataset 372 records and 19 attributes and 5 classes with 10-folds cross validation. This process show in Fig. 6. The 10-folds cross validation is a popular way of doing research to be used to test the performance of the model because the results are reliable, the performance measurement by cross-validation. This will divide the information into many parts. Usually shown with the value k. In this paper, we divide the data into 10-fold cross-validation, which is to divide the data into 10 parts, with each part having the same amount of data. After that, one piece of data will be used as a performance tester of model do this until the entire amount is divided.



result

Fig. 6. The implemental of 10-folds cross validation.

From the picture, we divide the training data into 10 equal parts. After that, test the efficiency of the model 10 times as follows:

Round 1 uses data from sections 2,3,4 to 10 to create models and use models to predict data 1 for testing.

Round 2 uses data part 1,3,4 to 10 to create models and use models to predict data part 2 for testing.

Round 3 uses data part 1,2,4 to 10 to create models and use models to predict data 3 for testing.

Round 4: use data from sections 1,2,3 to 10 to create a model and use the model to calculate data 4 for testing.

Round 5 uses data 1,2,3 through 10 to create models and use models to predict data 5 for testing.
Which repeats this cycle until 10 times.

The researcher used the following devices:
1) a notebook computer
2) Software used has the following features
- Windows 10 64 bit operating system
- Intel corei5 - 3337 U. Speed Processor 1.8GHz
- 4.00 GHz main memory speed

We has already taken the information through the research process. Perform data analysis to compare performance with 10-folds method by cross validation using 7 decision tree techniques, including DecisionStump, HoeffidingTree, J48, LMT, RandomForest, RandomTree and REPTree as shown in Table 3

TABLE II. SHOWS THE ACCURACY IN COMPARING THE EFFICIENCY OF ALGORITHMS IN DATA ANALYSIS.

Algorithm	Validity
DecisionStump	66.129
HoeffidingTree	83.0645
C4.5	100.00
LMT	98.1183
RandomForest	97.3118
RandomTree	73.6559
REPTree	99.4624

From the Table II. shows that the C4.5 method provides the highest accuracy with 100 percent. Follow by REPTree , LMT, RandomForest, HoeffidingTree, RandomTree, DecisionStump with accuracy of 99.4624, 98.1183, 97.3118, 83.0645, 73.6559 and 66.129 respectively.

The result show the efficiency for predicting the effects of physical images using data mining with decision tree techniques to compare the predictive effectiveness of physical fitness of general people in Ratchaburi Province By using information from the Sports Authority of Thailand Ratchaburi Province recorded a total of 376 people and collected all relevant factors in 19 factors. By allowing the last attribute to be divided into 5 classes, which are the best class, good, normal, low and very low, with 376 people to use to predict the results of the physical image using data mining with the decision tree technique for comparison Forecasting efficiency of physical fitness of the general public in Ratchaburi Province Perform data analysis based on 10-fold cross validation method. It is found that the J48 technique provides the highest accuracy with 100% accuracy.

Prediction of physical image effects using data mining This research focuses on creating intelligent health classification models for individuals with C4.5 technique, which collects physical fitness data of those exercising in Ratchaburi province. A total of 376 Records, 19 Attributes, including Age, Sex, Weight, Height, Pulse, Upper blood, Lower blood, Weight bike, Hand womenorce, Leg stretch, Triceps, Biceps, Suprailiac, Subscapular, Leg, Grip, womenlex, Lung capacity, O2 Class classification is divided into 5 classes, which are very good health, good health, moderate health. Poor health and poor health. In the experiment, the researcher divided the data set into two groups: training and testing and designed the test using 10-fold cross validation method. The result of C4.5 gave 100% accuracy.

REFERENCES

[1] T. C. Seng et al., "Predicting high cost patients with type 2 diabetes mellitus using hospital databases in a multi-ethnic Asian population," 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, 2016, pp. 240-243.

[2] T. Yaqoob, F. Mir, H. Abbas, W. B. Shahid, N. Shafqat and M. F. Amjad, "Feasibility analysis for deploying national healthcare information system (NHIS) for Pakistan," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Dalian, 2017, pp. 1-6.

[3] S. D. Min et al., "Actual Condition of Korean e-Health: What Do Enterprisers Want for Developing e-Health Industry?," 2007 9th International Conference on e-Health Networking, Application and Services, Taipei, 2007, pp. 304-307.

[4] D. L. Anthony and C. Campos-Castillo, "Do Health Care Users Think Electronic Health Records are Important for Themselves and Their Providers? Exploring Group Differences in a National Survey," 2013 IEEE International Conference on Healthcare Informatics, Philadelphia, PA, 2013, pp. 141-146.

[5] L. Horvath, "Toward smart health care: Building a national health information infrastructure(EESZT) in Hungary," 2017 IEEE 30th Neumann Colloquium (NC), Budapest, 2017, pp. 000011-000012.

[6] A. Fuad, S. S. Mulyono Putri, M. N. Sitaresmi and D. A. Puspandari, "Financial Sources Options for Telemedicine Program within Universal Health Coverage (UHC) Era in Indonesia," 2018 1st International Conference on Bioinformatics, Biotechnology, and Biomedical Engineering - Bioinformatics and Biomedical Engineering, Yogyakarta, 2018, pp. 1-5.

[7] K. Leerojanaprapa, W. Atthirawong, W. Aekplakorn and K. Sirikasemsuk, "Applying Bayesian network for noncommunicable diseases risk analysis: Implementing national health examination survey in Thailand," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 2017, pp. 904-908.

[8] N. Meephak, M. Tepbanchaporn and A. Jarupaibul, "Elder Eat: A smartphone application for recording and monitoring food consumption for Thai elderly," 2018 Seventh ICT International Student Project Conference (ICT-ISPC), Nakhonpathom, 2018, pp. 1-6.

[9] T. R. Khan, K. M. Hossein, K. R. I. Maruf, A. Fukuda and A. Ahmed, "Measurement of illness and wellness score of non-communicable disease patients," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, 2017, pp. 2253-2257.

[10] P. Huang, C. Lin, Y. Wang and H. Hsieh, "Development of Health Care System Based on Wearable Devices," 2019 Prognostics and System Health Management Conference (PHM-Paris), Paris, France, 2019, pp. 249-252.

[11] L. Graham, M. Moshirpour, M. Smith and B. H. Far, "Designing interactive health care systems: Bridging the gap between patients and health care professionals," IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, 2014, pp. 235-239.

[12] F. Prior and T. Dawson, "Development of a Holistic Health Economic Evaluation Tool Leveraging Patient Self-Report," 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, 2016, pp. 56-61.

[13] N. Chumuang and M. Ketcham, "Intelligent handwriting Thai SignatureRecognition System based on artificial neuron network," TENCON 2014 - 2014 IEEE Region 10 Conference, Bangkok, 2014, pp. 1-6.

[14] N. Chumuang and M. Ketcham, "Model for Handwritten Recognition Based on Artificial Intelligence," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-5.

[15] N. Chumuang, "Comparative Algorithm for Predicting the Protein Localization Sites with Yeast Dataset," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 369-374.

[16] S. Thaiparnit, N. Khuadthong, N. Chumuang and M. Ketcham, "Tracking Vehicles System Based on License Plate Recognition," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 220-225.

[17] Y. Zhang, R. Yang, K. Zhang, H. Jiang and J. J. Zhang, "Consumption Behavior Analytics-Aided Energy Forecasting and Dispatch," in IEEE Intelligent Systems, vol. 32, no. 4, pp. 59-63, 2017.

[18] M. A. Hassan, M. E. Shehab and E. M. R. Hamed, "A comparative study of classification algorithms in e-health environment," 2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC), Beirut, 2016, pp. 42-47.

[19] S. Rallapalli and T. Suryakanthi, "Predicting the risk of diabetes in big data electronic health Records by using scalable random forest classification algorithm," 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE), Durban, 2016, pp. 281-284.

[20] G. Wang, Z. Deng and K. Choi, "Tackling Missing Data in Community Health Studies Using Additive LS-SVM Classifier," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 579-587, March 2018.

TVis: A Light-weight Traffic Visualization System for DDoS Detection

Abhishek Kalwar
Dept. of Comp. Sc. & Engg.
Assam Kaziranga University
Jorhat 785006, India
akabhishek554@outlook.com

Monowar H. Bhuyan*
Laboratory for Cyber Resilience, NAIST
Nara 630 0192, Japan &
Dept. of Computing Science
Umeå University, Umeå 901 87, Sweden
monowar@cs.umu.se

Dhruba K. Bhattacharyya
Dept. of Comp. Sc. & Engg.
Tezpur University
Assam 784028, India
dkb@tezu.ernet.in

Jugal K. Kalita
Dept. of Computer Science
University of Colorado, CO 809 18, USA
jkalita@uccs.edu

Youki Kadobayashi
Laboratory for Cyber Resilience, NAIST
Nara 630 0192, Japan
youki-k@is.naist.jp

Erik Elmroth
Department of Computing Science
Umeå University, Umeå 901 87, Sweden
elmroth@cs.umu.se

***Abstract*—With the rapid growth of network size and complexity, network defenders are facing more challenges for protecting their networked computers and other devices from intelligent attacks. Traffic visualization is an important element in the anomaly detection system for visual observations and detection of distributed DoS attacks. This paper presents a visual interactive system called TVis which is proposed to detect DDoS attacks using Heron's triangle-area map estimation. TVis allows network defenders to detect and investigate anomalies in internal and external network traffic at both online and offline mode. We model the network traffic as an undirected graph and compute triangle-area map based on incidences in each vertex for each time period 5 seconds. The system triggers an alarm iff the system found the area beyond the dynamic threshold. TVis performs well in comparison to its competitors.**

***Index Terms*—DDoS attack; visualization; network traffic; online and offline; triangle-area;**

I. Introduction

Network systems are becoming more complex [1] in terms of size, topology, and especially traffic due to the use of the Internet. Simultaneously the number of network attacks against each host has increased exponentially. These attacks are often concealed among the vast amount of legitimate, and seemingly random traffic. Denial-of-Service (DoS) attack attempts to make machines or network resources unavailable to its intended users either temporarily or infinitely interrupting or suspending services of a host connected to the Internet. Normally, DoS attacks are being generated by a single host or a small number of hosts at the same location. Moreover, Distributed DoS (DDoS) attacks are a combination of DoS attacks where attacks are being generated by a large number of hosts. These hosts might be amplifiers or reflectors, or even might be zombies. They usually send the traffic to the target or victim host through the reflectors. Some examples of DDoS attack are: early attacks in 2000 to well-known websites, such as CNN, Amazon, and Yahoo, stopped normal services of these victims for hours [2], [3]. A new form of Mirai botnet based threats is hidden in the Tor network [1].

Most existing network defense techniques and tools still heavily rely on security analyst (SA). These techniques involve a security analyst to analyze and detect network attacks manually. To enhance the human perception and understanding different classes of network attacks, network traffic visualization has become more important in recent years that attempt to speed up the attack detection process through the visual analytics. Malicious activities such as DDoS attacks are relatively easy to implement and rather hard to prevent. Their timely detection at appropriate short time-scales (e.g., in milliseconds) requires processing of vast amounts meta-data (i.e., packet or NetFlow details) distributed throughout the entire network, thus making this a challenging task. DDoS attacks can be classified as low-rate and high-rate attacks based on attack-rate dynamics. A low-rate DDoS attacker attempts to bypass the security system by sending attack packets to the victim at a sufficiently low-rate to elude detection [4]. In a high-rate DDoS attack, the attacker sends the burst of attack packets to the victim within a short interval of time to overwhelm the bandwidth or resources. The task of analyzing both header and payload of network traffic is an NP-complete problem. So, we mostly focus on packet header information and visualization in terms of different parameters to support low-rate DDoS attack detection.

In this paper, we present a visualization system with time-periodic sampled traffic for an analysis period to detect both low-rate and high-rate DDoS attacks. We identify the three consecutive maximal dense vertices, form a triangle and estimate the area of that triangle, which generates using three vertices. Our system is guided by following principles: (a)

*Dr. Bhuyan is on lien from the Department of Computer Science and Engineering, Assam Kaziranga University, Jorhat, India.

[1] https://www.corero.com/blog/932-new-mirai-botnet-threat-hides-in-the-tor-network.html

examine all packets at the monitoring host (in promiscuous mode); (b) use of memory efficient data structures; (c) generate statistical summaries that can be retained for further analysis; (d) generate an undirected graph to visualize the network; and (e) triangle-area mapping of dense incident vertex and estimate area. The contributions of this work are as follows:

- A light-weight traffic visualization system to detect both low-rate and high-rate DDoS attacks using Heron's[1] triangle-area map estimation.
- TVis is cost-effective while it visualizes the network traffic. It can perform visualization in both online and offline modes.
- TVis is validated using testbed and benchmark datasets. Both cases it performs well with its competitors.

The rest of the paper is organized as follows. Section II discusses the related work while Section III introduces the proposed visualization system describing the framework and the model. Section IV reports the performance evaluation using testbed and benchmark datasets and finally concludes with Section V.

II. Related Work

Detection of low-rate DDoS attacks have been gaining more importance since last two decades. Several works have been proposed to detect attacks in large-volume alerts, produced by a detection tool which employed visualization methods. DDoSViewer [5] is a visual interactive system used for detecting DDoS attacks. DDoSViewer is specifically designed for detecting DDoS attacks through the analysis of visual patterns. SeeNet [6] is a visualization technique, which displays network traffic on a colored grid. Each point on the grid represents the level of traffic between a traffic source and a traffic destination. VisFlowConnect [7] uses a simple application of parallel coordinates [8] to display incoming and outgoing network flow data as links between two hosts or domains. It also employs a variety of visual cues to help attack detection.

Girardin [9] propose a visualization technique for static network data through the use of self-organizing maps for attack detection. It works in mapping multi-dimensional data into a 2D using an artificial neural network. Unfortunately, the layout of the hosts varies with each run, forcing users to re-acquaint themselves spatially with the network. Moreover, the algorithm is computationally intensive and designed for static data, making it challenging to use in real-time. The Spinning Cube [10] maps SIP, DIP and Dport to the axes in a 3D plot. The amount of network activity is visualized interactively in the plot using color, displaying certain attacks (eg., port scans) very clearly. Min et al. [11] present a technique to visualize alerts with alert correlation. Each intrusion alert is represented as a dot of various colors according to the attack class. The location and time of an attack are represented as graph's coordinates. The correlation of alerts is represented a line of different colors according to context analysis. This helps rapid intrusion analyses and traces many alerts. The idea overcomes alert flooding and false positive alerts, and provides the context information of intrusions by intuition. Zhou et al. [12] present a low-rate DDoS detection scheme developed based on the distribution of packet size. They estimate the packet size distribution distance between legitimate and low-rate traffic to detect low-rate DDoS attacks. Recently, David and Thomas [13] introduce a dynamic threshold-based DDoS detection scheme for NetFlow traffic and evaluated using real-time datasets.

A. Discussion

Unlike the traditional methods of analyzing textual log data, visualization techniques can increase the efficiency and effectiveness of network attack detection significantly. It can not only help analyst to deal with the large-volume of network data but also help network defenders to detect anomalies through visual analytics. It can even be used for discovering new types of attacks and forecasting the trend of unexpected events. Visualization techniques allow people to see and comprehend large amounts of complex data [19]. Graphics are used to assist IDS investigation and reporting process by helping the analyst to identify significant incidents and reduce false alarms. Complex patterns are clearly displayed over time in an easy way, where each of them can to understand soon [19]. Table II reports a comparison of existing detection mechanisms that detects DDoS attacks.

III. TVis: The Proposed System

This section starts with describing the proposed system architecture followed by the algorithm. It includes the concepts of visualization and detection strategy.

A. TVis: A Framework

We model this system as an undirected graph with each host as vertices $H = \{h_1, h_2, \cdots h_n\}$ and number of incidences on each vertices $I = \{i_1, i_2, \cdots i_n\}$ to form a triangle and estimates area for finding infected period. To get the end-point traffic, we configure our network to redirect all traffic to a particular port. So, TVis can monitor each traffic instance and visualize for detection of both low-rate and high-rate DDoS attacks. The framework of the proposed system is given in Figure 1. TVis uses jNetPcap [20] library for capturing and preprocessing traffic. After capturing network traffic, it filters out the IP packets for subsequent analysis. It uses developed subroutines to extract various relevant features from the IP packets and finally constructs a 5 min traffic feature sample for offline analysis. Then, these samples are formatted to our system for visualization. Due to light-weight in nature of our system, TVis can fast visualize the traffic to support attack detection.

We employ Heron's triangle-area map computation to estimate the infected period based on the incidence in a host. The concept of triangle-area map computation is shown in Figure 2. Let triangle-area map $A = \{a_1, a_2, \cdots a_n\}$, incidence per host,

[1]http://mathworld.wolfram.com/HeronsFormula.html

TABLE I
COMPARISON OF EXISTING DETECTION MECHANISMS

Author and Year	Scheme	Identification	Detection	Real-time Visualization	Real-time Stream	Interactive Zoom-in/out
Y. Zhang et al. [14], 2004	Multi-dimensional hierarchical	Yes	Yes	No	Yes	No
Lakhina et al. [15],2005	Entropy-based	No	Yes	No	No	No
Li et al. [16], 2006	Defeat	Yes	Yes	No	No	No
Van et al. [17], 2015	Metric-based	No	Yes	No	No	No
Zhou et al. [12], 2017	Packet size distribution	No	Yes	No	Yes	No
Behal et al. [18], 2018	D-FACE	No	Yes	No	Yes	No
David and Thomas [13], 2019	Dynamic thresholding	No	Yes	No	Yes	No

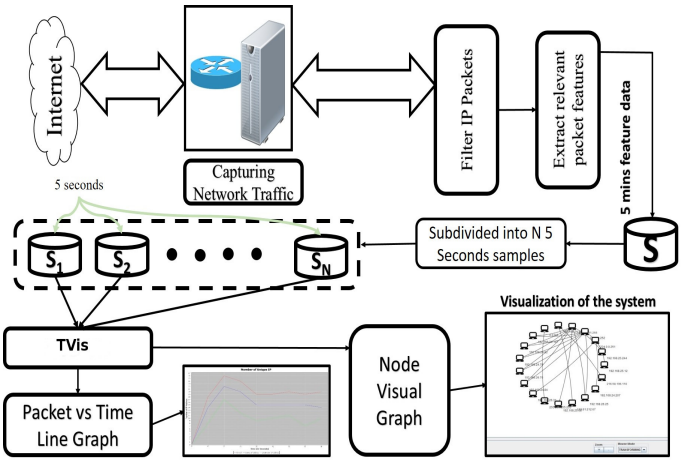


Fig. 1. TVis: a framework of the proposed system

$H_I = \{h_{i_1}, h_{i_2} \cdots h_{i_n}\}$, and time period $T = \{t_1, t_2, \cdots t_n\}$. Hence, the area of a triangle can be defined as:

$$\Delta a_1 = \sqrt{s_k(s_k - h_{i_1})(s_k - h_{i_2})(s_k - h_{i_3})} \qquad (1)$$

where s_k is the semiperimeter with $s_k = (h_{i_1} + h_{i_2} + h_{i_3})/2$, Δa_1 is the area of an infected period.

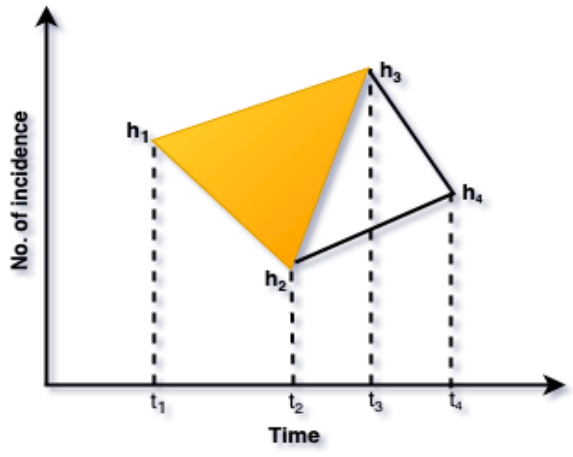


Fig. 2. Triangle-area map computation to visualize infected period

B. TVis: Algorithm

Algorithm 1 shows the major steps to design the TVis system for network traffic visualization and analysis for the detection of both low-rate and high-rate DDoS attacks. It has mainly two modules: online() and offline(). In online() mode, it captures, preprocesses, splits the traffic instances and visualizes them for attack detection. TVis maps the source node with destination node based on connection information. But in offline() mode, it just visualizes the already stored preprocessed traffic instances for detection. TVis works in similar fashion in offline mode except capturing and preprocessing of live packets. So, cost of computation in online() mode is more than offline mode to provide a near real-time performance. This algorithm provides two categories of graphs: sparse graph and dense graph that represented as legitimate and attack traffic, respectively. Further, we list consecutive three vertices with maximal incidents and computes the area of the triangle using Heron's formula. If the area is greater than thresholds $\delta_{A_l}, \delta_{A_h}$ for low-rate and high-rate attacks and also probability of packet loss increases then it generates an alarm. Lower the area of the triangle indicates high-rate attacks and vice-versa.

IV. PERFORMANCE EVALUATION

In this section, we describe the datasets used for performance analysis of TVis and report experimental results in details.

A. Datasets used

We use three different real-world datasets: (i) MIT Lincoln Laboratory [21], (ii) CAIDA DDoS 2007 [22], and (iii) Assam Kaziranga University (AKU) network datasets. The MIT Lincoln Laboratory dataset is real-time and contains pure normal data. The CAIDA DDoS 2007 dataset contains one hour of anonymized traffic traces from a DDoS attack launched on August 4, 2007. This dataset includes mainly two types of attacks: consumption of computing resources and the consumption of network bandwidth. We also used our system to collect and monitor live network traffic of *Assam Kaziranga University* campus. AKU dataset is composed of three categories of traffic such as normal traffic, low-rate attack traffic, and high-rate attack traffic. The network comprised about 500 hosts (both laptop and desktop), 6 L3 switches, and 25 wireless routers inside the University campus. We configure the network in such a way that all traffic passes through the host in which our system is deployed. We used four different

Algorithm 1 TVis (online, offline)

Input: mode ▷ defines the mode of capturing packet
Output: The visualized graph with triangle area

if $mode \neq online$ **then**
  call online( )
else
  call offline( )
end if
function ONLINE( )
  Initialize $storePacket[60]$, $device$, $T[3]$ ▷ It is an array of linked list to store packets as they arrive, and 60 arrays each storing 5 sec data, total 5mins
  Find all the network devices connected to the machine
  $device$ = get the choice of device from the list
  Open the $device$ for capturing in promiscuous mode
  for $i \leftarrow 0$ *to* 59 **do**
    $storePacket[i]$ = CAPTURE(device)
  end for
  ANALYSE(storePacket)
  NODEVISUALGRAPH(storePacket)
  exit
end function
function NODEVISUALGRAPH(storePacket)
  Initialize $graph$ ▷ an undirected graph where vertex are devices on the network and edges represent communication between them, $vertex$ ▷ linked list of devices, i.e., IP addresses, $edges$ ▷ a linked list each having value (v_i, v_j) where v_i, $v_j \in$ vertex
  for $i \leftarrow 0$, to size of $storePacket$ **do**
    for all $packet$ in $storePacket[i]$ **do**
      if $packet$ is an IP packet **then**
        if $packet.sourceIP$ not in $vertex$ **then**
          add $packet.sourceIP$ to $vertex$
        end if
        if $packet.destinationIP$ not in $vertex$ **then**
          add $packet.destinationIP$ to $vertex$
        end if
        if $(packet.sourceIP, packet.destinatio-nIP)$ not in $edges$ **then**
          add $(packet.sourceIP, packet.desti-nationIP)$
        end if
      end if
    end for
  end for
  $graph.addVertex(vertex)$
  $graph.addEdges(edges)$
  TRIANGLEAREAGEN(storePacket, T[])
  return $graph$
end function
function OFFLINE( )
  Initialize $storePacket[60]$ ▷ It is an array of linked list to store packets as they arrive, $time = 5000$ (It is time in milliseconds, 5000 represents 5 sec), $i = 0$ (for accessing the array)
  get the pcap file from user, i.e. $pcapFile$
  open $pcapFile$ to read packets
  for all $packets$ in $pcapFile$ **do**
    if $packet.timestamp > time$ **then**
      $time = time + 5000$
      $i++$
    end if
    add $packet$ to $storePacket[0]$
  end for
  ANALYSE(storePacket)
  NODEVISUALGRAPH(storePacket)
  exit
end function
function TRIANGLEAREAGEN(T[])( )
  Initialize A, k ▷ A indicates the area of a triangle
  for $i \neq k$ **do**
    if T[i] ≥ 1600 and pl ≥ 0.22 **then**
      compute A using Equation 1
    end if
    if $(A_1 \geq \delta_{A_h} || A_1 \leq \delta_{A_h})$ **then**
      Triggers an alarm
    end if
  end for
end function

attacks such as syn flood, smurf, ping flood and fraggle in distributed mode. We attempt to detect both low-rate and high-rate DDoS attacks within a short time interval. So, the network should not go down within a short time span.

B. Results

We evaluate the TVis system in both online and offline modes. In offline mode, we evaluate TVis system using CAIDA DDoS 2007 dataset. Figure 3 shows the visualized network traffic in offline mode. We can see from the figure 3 that how TVis identifies the presence of attack traffic. So, TVis system immediately sends a request to the edge router to drop the packet before entering into the network. It also depends on the period between the attack pulses use to overwhelm the target. We compute triangle area iff the system found consecutive vertices with incidents greater than at least 1600 packets and same time increased packet losses.

We also evaluate TVis system using MIT Lincoln Laboratory dataset to differentiate between normal and attack traffic. Figure 4 shows the visualized network traffic in offline mode using MIT Lincoln Laboratory dataset. From figure 4, we can see that the presence of attacks or not in the traffic. Because it generates the sparse graph that enables to identify the normal traffic for time-periodic sampled traffic.

In online mode, we used the live network traffic of Assam Kaziranga University campus while executing attacks. We capture and visualized traffic for 5 *seconds* interval shown in Figure 7 when executing attacks in the testbed. Also, we observe the unique IP address and packets per protocol within a time period in Figure 5, 6, respectively. As we can see in

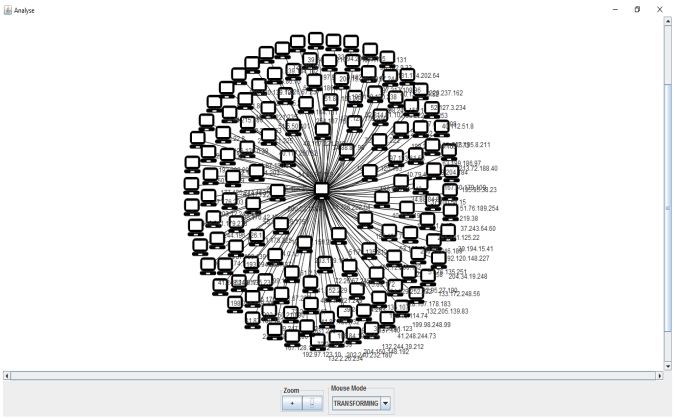


Fig. 3. TVis: visualization of network traffic available at CAIDA DDoS 2007 dataset

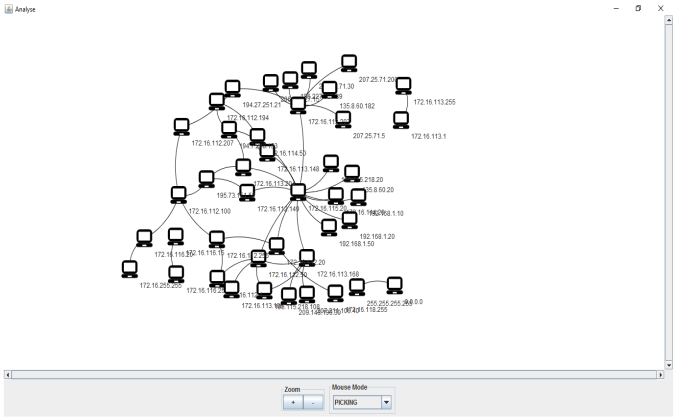


Fig. 4. TVis: visualization of network traffic available at MIT Lincoln Laboratory dataset

the visualization of all hosts in the network are not shown, as it shows only those hosts that are active and either sending or receiving packets including the target. From figure 7 derives the dense graph that enables again to detect attacks with better accuracy.

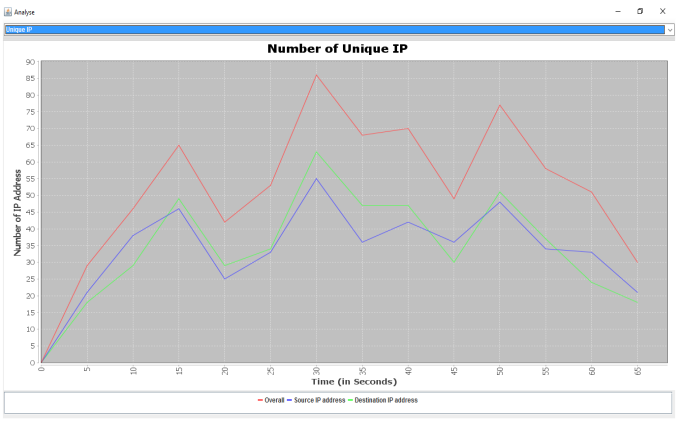


Fig. 5. Number of unique IP addresses

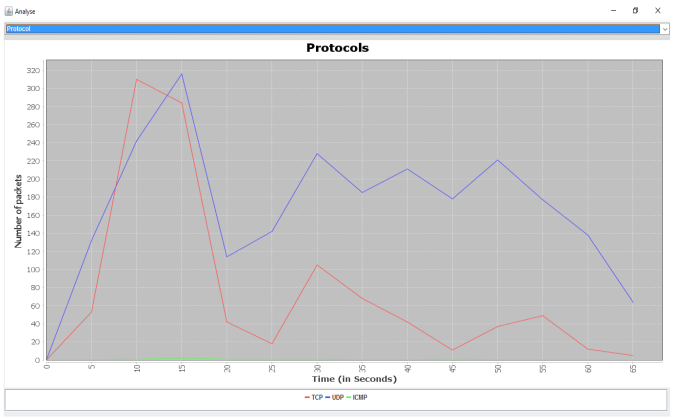


Fig. 6. Number of packets per protocol

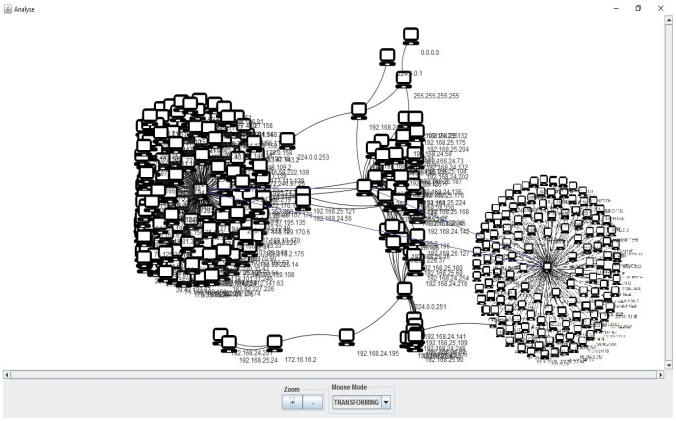


Fig. 7. TVis: visualization of traffic for our network

According to Moore et al. [23], we generate both low-rate and high-rate DDoS attack traffic for evaluation of TVis system. The attack traffic is generated more than 1600 and less than 5000 packets per seconds for low-rate attacks. Similarly, emulate the high-rate attacks in the testbed. However, this number will vary for different datasets and environments. Based on our experiment, we observe that TVis indicates an alarm for attack when normalized areas of the triangle, $\delta_A \geq 0.43$ and $P_k \geq 1600$ packets per seconds with 5 second interval transmits over the network. Our system is significant in view of the following points and also in comparison to recent work [24]. Figure 8 reports ROC curve for the TVis system for detecting DDoS attacks when uses MIT Lincoln Laboratory normal and CAIDA DDoS datasets.

- TVis is cost-effective and can operate in both online and offline mode.
- It is fast, scalable, and able to detect both low-rate and high-rate DDoS attacks effectively.

The triangle area δ_A goes lower for high-rate DDoS attacks and increases for low-rate attacks. Because the high-rate attack is more frequent towards a target with maximum intensity of malicious incidence in short span of time. However, the low-rate attack is less frequent towards target and similar to

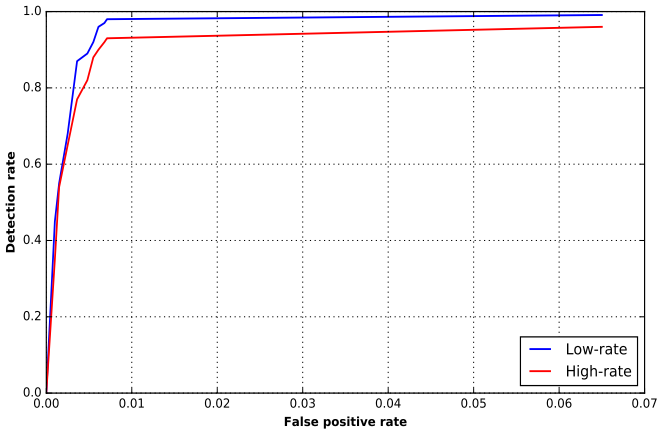


Fig. 8. TVis: ROC curve when compared with existing methods using MIT normal and CAIDA DDoS datasets

legitimate traffic. For our experiment, we found best results when $\delta_{A_h} \geq 0.43$ and $\delta_{A_h} \leq 0.61$, otherwise low-rate attacks. However, TVis shows $\delta_{A_leg} < 0.43$ for legitimate traffic instances.

V. Conclusion and Future Work

The TVis system is presented to visualize the network traffic in real-time for detection of both low-rate and high-rate DDoS attacks. Use of appropriate data structure is helpful to implement the system in near real-time and to support successful detection of both low-rate and high DDoS attacks. The undirected graph is generated and triggers an alarm based on the estimated triangle area of consecutive maximal incident vertices using Heron's formula. Our system performs well in detecting four different classes of DDoS attacks including syn flood, smurf, fraggle, and ping flood. However, we report result for real-time syn flood attacks only in both offline and online mode with extended scalability.

The TVis system is under development to evolve with new attacks. We are also undergoing to add datacenters infrastructures and services visualization features to monitor and prevent incidents in real-time and reduce down time of applications.

Acknowledgment

This work was supported by the Kempe post-doc fellowship via project no. SMK-1644, Sweden. Additional support was provided by the International Exchange Program of the National Institute of Information and Communications (NICT) and JST CREST Grant Number JPMJCR1783, Japan.

References

[1] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju, "VisFlowConnect: Netflow Visualizations of Link Relationships for Security Situational Awareness," in *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. New York, NY, USA: ACM, 2004, pp. 26–34.

[2] L. Garber, "Denial-of-Service Attacks Rip the Internet," *Computer*, vol. 33, no. 4, pp. 12–17, April 2000.

[3] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network Anomaly Detection: Methods, Systems and Tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.

[4] M. H. Bhuyan, A. Kalwar, A. Goswami, D. K. Bhattacharyya, and J. K. Kalita, "Low-Rate and High-Rate Distributed DoS Attack Detection Using Partial Rank Correlation," in *Communication Systems and Network Technologies, 2015 Fifth International Conference on*, April 2015, pp. 706–710.

[5] J. Zhang, G. Yang, L. Lu, M. Huang, and M. Che, *Visual Information Communication*. Boston, MA: Springer US, 2010, ch. A Novel Visualization Method for Detecting DDoS Network Attacks, pp. 185–194.

[6] R. A. Becker, S. G. Eick, and A. R. Wilks, "Visualizing Network Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 1, pp. 16–28, March 1995.

[7] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju, "VisFlowConnect: netflow visualizations of link relationships for security situational awareness," in *Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC 2004), 29 October 2004, Washington DC, USA*, 2004, pp. 26–34.

[8] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multi-dimensional geometry," in *Proceedings of the 1st Conference on Visualization '90*, ser. VIS 90. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990, pp. 361–378.

[9] Girardin and Luc, "An eye on network intruder-administrator shootouts," in *Proceedings of the 1st Conference on Workshop on Intrusion Detection and Network Monitoring - Volume 1*, ser. ID'99. Berkeley, CA, USA: USENIX Association, 1999, pp. 3–3.

[10] Lau and Stephen, "The spinning cube of potential doom," *Commun. ACM*, vol. 47, no. 6, pp. 25–26, June 2004.

[11] B. Min, J. Kim, and S. HongIn, "Visualization of Intrusion Detection Alerts with Alert Correlation," https://hpc.postech.ac.kr/wiki/pds/InternationalConference/bgmin_acns04.pdf, 2004.

[12] L. Zhou, M. Liao, C. Yuan, and H. Zhang, "Low-rate ddos attack detection using expectation of packet size," *Security and Communication Networks*, vol. 2017, 2017.

[13] J. David and C. Thomas, "Efficient ddos flood attack detection using dynamic thresholding on flow-based network traffic," *Computers & Security*, vol. 82, pp. 284–295, 2019.

[14] Y. Zhang, S. Singh, S. Sen, N. Duffield, and C. Lund, "Online Identification of Hierarchical Heavy Hitters: Algorithms, Evaluation, and Applications," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 101–114.

[15] A. Lakhina, M. Crovella, and C. Diot, "Mining Anomalies Using Traffic Feature Distributions," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 217–228, Aug. 2005.

[16] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and Identification of Network Anomalies Using Sketch Subspaces," in *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '06. New York, NY, USA: ACM, 2006, pp. 147–152.

[17] D. van der Steeg, R. Hofstede, A. Sperotto, and A. Pras, "Real-time DDoS attack detection for Cisco IOS using NetFlow," in *Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on*, 2015, pp. 972–977.

[18] S. Behal, K. Kumar, and M. Sachdeva, "D-face: An anomaly based distributed approach for early detection of ddos attacks and flash events," *Journal of Network and Computer Applications*, vol. 111, pp. 49–63, 2018.

[19] W. Wright and P. Clarke, "Visualization Techniques for Intrusion Detection." [Online]. Available: http://handle.dtic.mil/100.2/ADA428197l

[20] jNetPcap, "jNetPcap - what is it?" http://jnetpcap.com/.

[21] MIT Lincoln Laboratory Datasets, "MIT LLS_DDOS_0.2.2," http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/2000data.html, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[22] CAIDA, "The Cooperative Analysis for Internet Data Analysis," http://www.caida.org, 2011.

[23] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, "Inferring Internet Denial-of-service Activity," *ACM Trans. Computer Systems*, vol. 24, no. 2, pp. 115–139, May 2006.

[24] J. Song, T. Itoh, G. Park, and H. Takakura, "An Advanced Security Event Visualization Method for Identifying Real Cyber Attacks," *Appl. Math. Inf. Sci.*, vol. 11, no. 2, pp. 353–361, 2017.

An Effect of Using Deep Learning in Thai-English Machine Translation Processes

***Abstract*— Deep learning has been used in many fields including natural language processing. This paper aims to study the effect of applying deep learning in machine translation processes including word segmentation and translation model generation. We compare the results of the process from traditional statistical method and deep learning and analyze the difference. From experiment, the results indicated that the processes from deep learning obtained higher score in overall. Word segmentation from Bidirectional neural network yielded 0.861 f1 score which was higher than standard n-gram based system for 0.081. The translation results within dataset show that the neural-network-based translation got the best BLEU score in average for 0.43 in which are higher than the traditional statistical approach for 0.16.**

Keywords*— *Deep Learning, Neural Machine Translation

Thai-English translation

I. Introduction

Deep Learning has become well-known for its supremacy in terms of accuracy when comparing to other methods once trained with huge amount of data [1][2][3][4]. In this era of data overloaded, deep learning has been applied to various tasks and become a part of state-of-the-art systems in various disciplines, especially on complex problems such as computer vision, speech recognition and natural language processing. Deep learning has been reported to outperform other traditional approaches with the trade-off of requiring great amount of training data and high-end infrastructure to train in reasonable time [5].

For machine translation task, neural machine translation (NMT) was firstly published in 2014 [2], and several advance versions of it were followingly introduced such as Multilingual NMT [6], Multi-Source NMT [7], Fully Character-NMT [8] and Zero-Shot NMT [9]. Unlike the conventional translation systems, e.g. statistical machine translation (SMT), all parts of the neural translation model are trained jointly to maximise the translation performance [ref] in which results in improving translation performance. The well-known translation services including Google Translate service [9], Microsoft Trans [10], and PROMT[11]have thus used NMT and are recognised as the most practical translation services at the moment.

For automated Thai text translation, most of the past researches were based on SMT approach, especially hierarchical phrase-based translation (HPBT) [12]. A few researches towards NMT of Thai translation has been conducted but yet to be published. In this work, a study on effect of applying deep learning in NLP processes of developing an NMT for Thai-English is conducted in comparison to the traditional SMT approach. The processes include word-segmentation and translation for Thai to English and English to Thai translation service. The result thus can be used as a benchmark for applying advance NMT technique.

II. Related Work

In an automated translation task, a statistical based approach has gained the most favorable attention among all approaches due to its simplicity in development as requiring a sufficient amount of a bilingual corpus. Since the introduction of the neural machine translation (NMT) which exploits the deep learning technique, most of the automated text translation service has been altered to the technique since it is reported in many works [13][14]to achieve significantly higher performance in terms of accuracy and coverage. Not only the deep learning has been used for developing translation model, but other related processes in preparing data for the translation have also been shifted to deep learning including word segmentation and part-of-speech tagging task. In this part, we review the published NMT works to find how they perform and what the works find in their conduct. A brief summary of the works is given in Table 1.

Table1. A summary of existing NMT works by language pair and evaluation result

Paper	Language Pair	Result BLEU (improvement)
[15]	English → French (WMT'14)	37.5(+2.8)
[16]	Arabic → English (NIST OpenMT12) Chinese → English (NIST OpenMT12)	43.9(+3.0) 32.2(+6.3)
[13]	Chinese → English (BTEC and SLDB)	48.70(+1.5)
[17]	English ←→ German English ←→ Russian	25.3(+1.1) 24.1(+1.3)
[6]	English-France English- Germany	39.92(+0.97) 24.60(-0.07)

[18]	Hausa→English Turkish→English Uzbek→English Urdu→English	24.8(+0.8) 21.8(+1.0) 19.5(+1.6) 19.1(+0.3)
[19]	Spain→France Spain→Italy Spain→Portugal Spain→Romania France→Italy France→Portugal France→Romania	32.7(+5.4) 28.0(+4.6) 34.4(+6.1) 28.7(+6.4) 31.0(+5.8) 34.1(+4.7) 31.9(+9.0)
[7]	Romanian→English (WMT2016)	23(+4.0)
[14]	English→French (WMT'14) English→German(WMT'14)	41.67(+0.11) 28.84(+0.06)
[20]	English→German(WMT'14) English→French (WMT'14)	29.8(+0.6) 43.2(-1.1)
[21]	Enlish→Germany(WMT17)	29.5(+0.1)
[22]	Enlish-Germany(WMT'15) Enlish→Finland(WMT'15)	23.74(+2.96) 10.20(+2.37)
[23]	English→German(WMT 2014) English→French(WMT 2014)	28.4(+4.65) 41.8(+2.60)

The review indicates that NMT has been used in many language pairs such as Turkish-English [14] German-English [9][19], and Chinese-English [8][10], and all of them obtains superior result than the traditional SMT approach. With a requirement of high-performance hardware, NMT with sufficient data is undoubtedly preferred. Many works have gone further to improve the method by adding more techniques to handle low language-resources issue [6][14][15] which directly affects the translation performance. For Thai language translation, a little-to-none has been tested and publicly reported. Hence, this work aims to study how NMT performs and sets up the baseline for further improvement.

III. Development of Thai-English NMT

This work applies the basic NMT to develop a Thai to English and

English to Thai automatic translation service. Although the developing method resembles the proposed method from [19], some adjustments to suit Thai language have been done for practical usage. For an overview, there are 3 main processes including data preparation, training translation model and translating as shown in Figure 1.

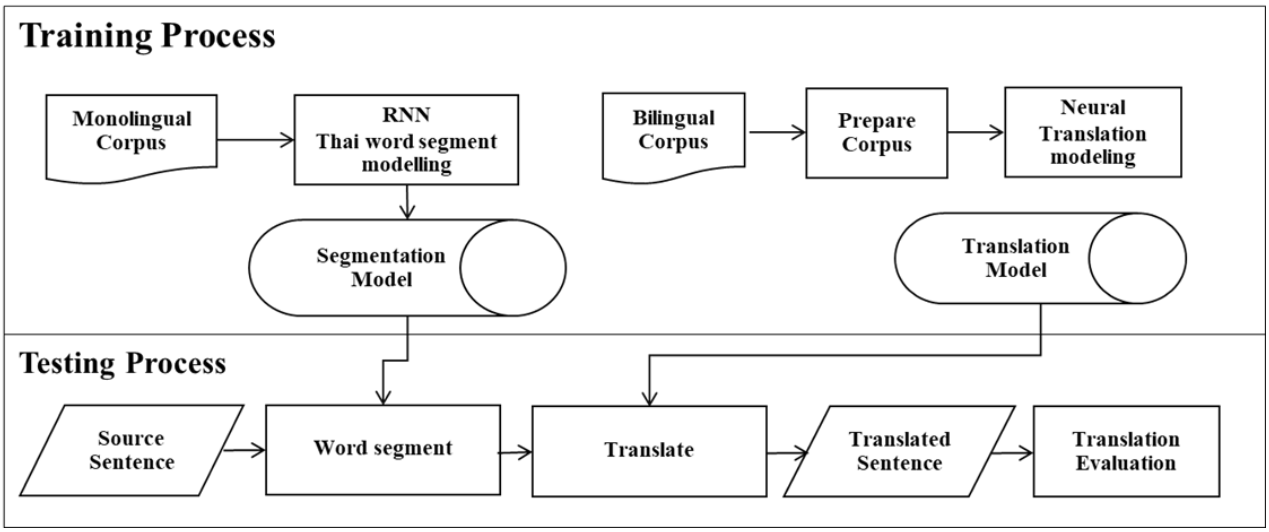


Figure 1. An overview of our NMT processes

A. Thai-English Data Preparation for NMT

The usual process in preparing Thai based natural processing service is word-segmentation. It is well-known that Thai text is expressed in consequence without a clear word boundary [24] and is complex from being semantically ambiguous [25]. Moreover, the concept of words in Thai is rather vague since Thai emerging words in a modern time are from combining existing words together. Thus, the word-segmentation for Thai is a truly specific to the work such as word as a single syntactic element and words by concept level. For translation purpose, the words are preferred as a concept level for reducing complexity in combining elements in a translation model generating phase. To handle the word-segmentation issue, we select bidirectional neural network regarding the method proposed by Boonkwan et al. [26].

In brief, the BRNN method for segmenting Thai words is to consider an N-gram of characters from the learning texts as a feature for using in bidirectional recurrent neural networks (BRNNs) to generate word boundary inference model. For full details, please see [26].

B. Thai-English NMT

The aim of this work is to study the effect of using neural machine translation (NMT) in a comparison to the tradition statistical machine translation (SMT) and its improved version as hierarchical phrase-based machine translation (HPBT). This work applies the conventional OpenNMT [19] in a development. We design the NMT as open service for Thai-English translation opened for user to use it online. The entire translation process is as shown in Figure 2.

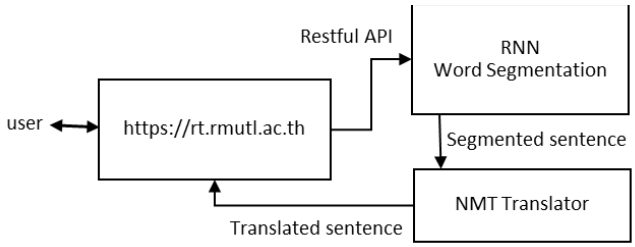


Figure 2. Architecture of our Thai-English NMT service

For details, NMT [19] takes a conditional language modeling view of translation by modeling the probability of a target sentence $w_{1:T}$ given a source sentence $x_{1:S}$ as

$$p(w_{1:T}|x) = \prod_1^T p(w_t|w_{1:T-1}, x; \theta) \quad (3)$$

This distribution is estimated using an attention-based encoder-decoder architecture [3]. A recurrent neural network (RNN) of a source encoder maps each source word to a word vector. Then, it processes these words to a sequence of hidden vectors $h_{1,\dots,}h_s$. The target decoder combines an RNN hidden representation of previously generated words $(w_{1,\dots,}w_{t-1})$ with source hidden vectors to predict scores for each possible next word. A softmax layer is used to produce a next-word distribution $p(w_t|w_{1:T-1}, x; \theta)$. The source hidden vectors affect the distribution by considering on an attention pooling layer that weights each source word to its expected contribution to the target prediction. The complete model is trained end-to-end to minimize the negative log-likelihood of the training corpus. The full details of OpenNMT can be read from [19].

IV. Experiments

The goal of this study is to learn the effect in term of translating results when applying deep learning in automatic translation processes in a comparison to the previous traditional approaches. The experiment was split into 3 parts including word-segmentation, translation within dataset and open translation.

A. *Evaluation of Word Segmentation*

The dataset in this experiment consists of 149,000 sentences with 2,418 terms. 10-fold cross validation was applied. The measurements in this experiment are precision, recall and f1 score. We compare the BRNNs and traditional N-gram. We obtained the result shown in Table 2.

Table 2. Word-segmentation results in comparison between BNN and N-gram

Fold	BRNN			N-gram		
	Precision	Recall	F1	Precision	Recall	F1
1	0.891	0.842	0.866	0.742	0.719	0.730
2	0.837	0.889	0.862	0.771	0.745	0.758
3	0.889	0.829	0.858	0.734	0.699	0.716
4	0.903	0.891	0.897	0.696	0.712	0.704
5	0.887	0.809	0.846	0.702	0.694	0.698
6	0.863	0.891	0.877	0.725	0.713	0.719
7	0.762	0.815	0.787	0.689	0.672	0.680
8	0.874	0.87	0.872	0.711	0.707	0.709
9	0.878	0.865	0.872	0.726	0.744	0.735
10	0.896	0.845	0.87	0.735	0.737	0.736
average	0.868	0.855	0.861	0.7231	0.7142	0.719

B. *Evaluation of Translation within Dataset*

The data set in this experiment was bi-tech. It contains 149,000 Thai-English sentences with 6,524 unique terms. The Thai terms were segmented by BNN and post-edited by linguists. The ratio for training, tuning and testing was 80:10:10 percent. The measurement was BLEU score [22] in which compares the automated translation with a set of good quality reference translations using a modified precision. BLEU's output is calculated to a number between 0 and 1 where 1 is the precise translation same as the reference, and 0 refers to nothing alike. The translation results of SMT (baseline), HPBT and NMT of the same settings were compared as providing in Table 3.

Table 3. BLEU score of Thai-English translation from approaches

Approach	Thai to English BLEU score	English to Thai BLEU score
SMT (PBT)	0.259	0.273
SMT (HPBT)	0.285	0.294
NMT	0.421	0.442

The NMT outperformed both PBT and HPBT in terms of BLEU score. The difference was +0.14 from HPBT which was the commonly used approach in the past years for Thai-English automated translation [27][28]. In a comparison in a sentence level between NMT and HPBY, we compared BLEU score of each sentence from overall 10,000 test sentences and see if there was better, equal (difference is less than 1 score) or worse as shown in Figure 3.

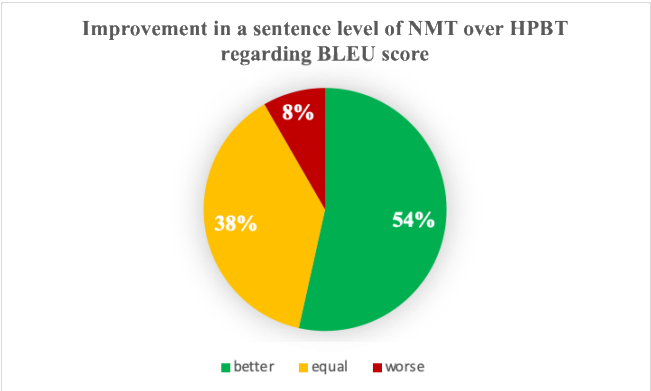


Figure 3. Comparison result of each test sentence for NMT improvement over HPBT

For the worse cases, we looked into them and found that they were the short sentences containing less than 6 six words. The translated words were though in the related

meaning but different form and were thus calculated to be incorrect leading to lower BLEU point. In overall, there were 54% of cases that NMT performed better. With the use of the same setting and data set, the experimental results signify that an NMT could replace the previous approaches regarding translation capability.

C. Evaluation of Translation from Users

To test the translation model with practical usage, we have opened the NMT translation service online at https://rt.rmutl.ac.th as free service available for users since August, 2018. There were about 6,000 Thai sentences and 3,000 English sentences translated in the service. We collected the input and translation results for analysis. Unlike translating within dataset, there were many unknown words in these translations as well as named entities. The inputs were more natural and varied in terms of domain and style.

In this part, we analyse the translation results for the known translation issues including unknown word problem and Thai classifier translation problem. Unknown word problem is a classic problem in automated machine translation since it is difficult to translate the words that were not learned. For Thai classifier problem, it is a difference in grammatical usage between Thai and English in which extra Thai classifier is added in numeral expression such as 'ช้าง 2 เชือก' ←→ 2 elephants. The classifier is thus also translated as another word and cause the translation to be incorrect.

For the unknown word problem, NMT tends to guess the word through the probability via the neural network and returns the translated word with highest probability while the basic SMT approach returns the original word as it is for the words that not existed in the translation model. Thus, we counted the amount of the words that did not exist in the training dataset, and we manually examine that the words can be translated correctly or not. There are 3 types of translation as correct (same sense), partial (semantically related sense to the word) and incorrect (different meaning). For the partial case, we consider the meaning carefully based on types. It will count if 'the translated word and original word are of the same POS' and 'the meaning is more specific or generalized of the same thing' such as dog to mammal (generalized) and literature to document (specific). From analysis, we obtained the results as given in Table 4.

Table 4. Assessment of translation for unknown words from NMT open service

Unknown words	Correct	Partial	Incorrect
4,095 cases (TH → EN)	67 (1.64%)	303 (7.40%)	3,859 (94.24%)
1,788 cases (EN → TH)	49 (2.74%)	171 (9.90%)	1,660 (92.84%)

From the result in Table 4, the NMT could correctly guess the unknown words for about 1-2 percent and partially correct for about 7-9 percent. Those partial correct translations were all acceptable and fit well in a context. However, the incorrect results were misleading and awkward by choosing the word unrelated to the context or selecting opposite meaning term.

With Thai classifier issue, we counted the cases that were inputted to the translation service, and we investigated their translation result. We split the issue into 3 subtypes as unit classifier, collective classifier and measurement classifier. A unit classifier in Thai is an extra word following number with no additional meaning while collective and measuring classifier required to be translated. However, a grammatical expression is different for the two as collective classifier with number should be placed before the core noun in English while measurement classifier is translated and place in same fashion for both Thai and English. Please see the following examples.

- เทียน (candle) 3 เล่ม (candle unit classifier) → 3 candles
- หนังสือ (book) 2 กอง (pile) → 2 piles of book
- วิ่ง (run) 10 กิโลเมตร (kilometer) → run 10 kilometers

The issues have been a great hurdle in traditional statistical approach for lower their translation accuracy, and we want to investigate the performance of NMT for this problem. Hence, we counted the correct translation of these cases and obtained the results as shown in Table 5.

Table 5. Performance of Thai to English NMT on Thai classifier issue

Classifier Type	Cases	Correct amount (percentage)
Unit	153	103 (67.32%)
Collective	92	40 (43.48%)
Measurement	88	65 (73.86%)

The result shows that the Thai to English NMT produced the correct translation for 62% in overall. The 'unit' classifier type which is the often-found issue in the traditional SMT could be solved for about 67% while the 'collective' classifier type obtained the lowest as 43%. We compared these cases with the HPBT and selected some interesting results to be shown in Table 6.

Table 6. Translation results of NMT and HPBT on Thai classifier cases

	Original Thai	NMT	HPBT
1	เขาซื้อหนังสือสาม เล่ม	He bought three book	He buy book three candle
2	เขามีลูกสี่คน	He has four children	He had person four child
3	เบนได้รถใหม่หนึ่งคัน	Ben got a new one	Ben got a rod
4	ที่นี่มีโต๊ะห้าตัว	Here have five tables	Here have table five
5	ฉันมีแก้วหกใบ	I have six glasses	I have six glasses
6	เชือกนี้ยาวสาม เซนติเมตร	The rope is three centimeters long	Three rope is long

7	เสื้อตัวละร้อยห้าสิบบาท	The shirt is fifty baht	The shirt is 150
8	ผมขอเบียร์สองขวด	Two bottles of beer please	Beer ask two bottle

For HPBT, quality of translation was noticeably lower than the NMT in most example cases. However, there were some incorrect translations for NMT such as sentence#1, #3, #6 and #7. There can be classified into two types of incorrection: incorrect numeral agreement and wrong word selection. The former including sentence#1 and #6 was in a grammatical level but was semantically understandable. The latter from sentence#3 and #7 nevertheless was more crucial since the words of core meaning were altered such as the word 'รถ' (car, noun) was missing and replaced with the number as 'one' in sentence#3, and 'ร้อยห้าสิบ' (150, number) was guessed as 'fifty' in which was completely different and undeniably crucial for marketing. The issue thus requires more attention to be solved in further improving since it reduces creditability of Thai to English automated translation in practical use.

V. Conclusion

With the boom of deep learning, neural machine translation (NMT) was invented and become the new standard approach for an automated translation system with noticeably higher accuracy. This paper studies the effect and capability of applying deep learning in machine translation processes in a comparison with traditional statistical approaches for a language pair of Thai and English. The studied processes to be studied include word tokenization and translation. For Thai word tokenization, the chosen deep learning method is bidirectional neural network. The OpenNMT is applied to develop Thai-English translation service. The comparison results indicated that all the processes that apply deep learning technique yield the better results in overall. For word tokenization, the deep learning obtained 0.861 f1 score which was higher than standard n-gram based system for 0.081. The translation results within dataset show that the NMT got the best BLEU score for 0.421 and 0.442 for Thai to English and English to Thai translation respectively, in which are higher than the traditional statistical approach for about 0.16 BLEU score in average. In analysis of practical usage, the NMT could solve the unknown word issue for their ability to guess the word from probability for 9% and 12% for Thai to English and English to Thai translation.

References

[1] M. Popov *et al.*, "A Neural Probabilistic Language Model," *J. Mach. Learn. Res.*, vol. 3, pp. 1137--1155, 2003.

[2] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[3] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv Prepr. arXiv1409.0473*, 2014.

[4] A. Goodfellow, Ian and Bengio, Yoshua and Courville, *Deep Learning*. MIT press, 2016.

[5] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] Y. Wu *et al.*, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *arXiv Prepr. arXiv1609.08144*, 2016.

[7] J. Gu, H. Hassan, J. Devlin, and V. O. K. Li, "Universal Neural Machine Translation for Extremely Low Resource Languages," *arXiv Prepr. arXiv1802.05368*, 2018.

[8] X. Ma and E. Hovy, "End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF," *54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap.*, vol. 2, pp. 1064–1074, 2016.

[9] M. Johnson *et al.*, "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 339–351, 2017.

[10] "Microsoft Translator." [Online]. Available: https://www.microsoft.com/en-us/translator. [Accessed: 09-Sep-2019].

[11] "PROMT Translation Software and Dictionaries." [Online]. Available: https://www.promt.com/. [Accessed: 09-Sep-2019].

[12] P. Luekhong, T. Ruangrajitpakorn, T. Supnithi, and R. Sukhahuta, "A Comparison Study of Thai Translation on Applying Phrase-based and Hierarchical Phrase-based Translation A Comparison Study of Thai Translation on Applying Phrase-based and Hierarchical Phrase-based Translation," in *Advances in Natural Language Processing, Intelligent Informatics and Smart Technology*, no. December, 2016.

[13] S. Liu, N. Yang, M. Li, and M. Zhou, "A Recursive Recurrent Neural Network for Statistical Machine Translation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 1491–1500.

[14] M. X. Chen *et al.*, "The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation," *arXiv Prepr. arXiv1804.09849*, 2018.

[15] M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba, "Addressing the Rare Word Problem in Neural Machine Translation," *arXiv Prepr. arXiv1410.8206*, 2014.

[16] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz, and J. Makhoul, "Fast and Robust Neural Network Joint Models for Statistical Machine Translation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 1370–1380.

[17] R. Sennrich, B. Haddow, and A. Birch, "Neural Machine Translation of Rare Words with Subword Units," *arXiv Prepr. arXiv1508.07909*, 2015.

[18] B. Zoph, D. Yuret, J. May, and K. Knight, "Transfer Learning for Low-Resource Neural Machine Translation," *arXiv Prepr. arXiv1604.02201*, 2016.

[19] G. Klein, Y. Kim, Y. Deng, J. Crego, J. Senellart, and A. M. Rush, "OpenNMT: Open-source Toolkit for Neural Machine Translation," *arXiv Prepr. arXiv1701.02810*, 2017.

[20] M. Ott, S. Edunov, D. Grangier, and M. Auli, "Scaling Neural Machine Translation," *arXiv Prepr. arXiv1806.00187*, 2018.

[21] M. Junczys-Dowmunt *et al.*, "Marian: Fast Neural Machine Translation in C++," *arXiv Prepr. arXiv1804.00344*, 2018.

[22] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

[23] A. Vaswani *et al.*, "Tensor2Tensor for Neural Machine Translation," *arXiv Prepr. arXiv1803.07416*, 2018.

[24] W. A.-P. of the 5th S. & 5th Oriental and undefined 2002, "Collocation and Thai word segmentation," *academia.edu*.

[25] V. Tesprasit, P. Charoenpornsawat, and V. Sornlertlamvanich, "A Context-Sensitive Homograph Disambiguation in Thai Text-to-Speech Synthesis," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL*, 2003.

[26] P. Boonkwan and T. Supnithi, "Bidirectional Deep Learning of Context Representation for Joint Word Segmentation and POS Tagging Conference," in *Advances in Intelligent Systems and Computing*, 2018, no. June.

[27] P. Luekhong, T. Ruangrajitpakorn, R. Sukhahuta, and T.

Supnithi, “A study of a Thai-English translation comparing on applying phrase-based and hierarchical phrase-based translation,” in *nternational Conference on Knowledge, Information and Creativity Support System*, 2017.

[28] K. Kongkachandra and R. Phodong, “Using Linguistic Knowledge for Enhancing Word Alignment in Thai-English SMT,” in *International Conference on Knowledge, Information and Creativity Support System*, 2017.

Thai Vowels Speech Recognition using Convolutional Neural Networks

Niyada Rukwong
Department of Computing, Faculty of Science
Silpakorn University
Nakhon Pathom, Thailand
rukwong_n@su.ac.th

Sunee Pongpinigpinyo*
Department of Computing, Faculty of Science
Silpakorn University
Nakhon Pathom, Thailand
pongpinigpinyo_s@su.ac.th

***Abstract*— The vowel is considered as the core of syllable in each word. This paper aims to present noisy Thai vowels speech recognition by using Convolutional Neural Network (CNN). The noisy Thai vowels dataset is the speech of Thai vowels in real-world situations. The sound is collected in a real environment from several areas which consist of many types of noise at 30 - 40 dB SNR (Signal to Noise Ratio). It constrains 16 kHz speech data is recorded from a mobile phone. The vowel speeches are separated into 2 groups: male's voice and female's voice from 25 male and 25 female speakers. In this research, it constrains 18 classes (9 short vowels and 9 long vowels). Mel Frequency Cepstral Coefficients (MFCCs) are used for feature extraction. The most accuracy rate of the CNN_Thai Simple Vowel (CNN_TSV) model is 90.00% and 88.89% on female and male voices respectively. The comparison results of CNN_TSV model with other models such as Multilayer Perceptron (MLP) and Support Vector Machines (SVM) show that the CNN_TSV model is the most effective for both female and male voices. This research can be used as one of an alternative model to apply for Computer-assisted language learning (CALL) in the future development direction.**

Keywords—Convolutional Neural Networks, CNNs, Vowels, Thai vowels, Classification, Speech Recognition.

I. INTRODUCTION

Vowels always make problems than consonants for second language learners [1] whose native language do not have a difference in the duration of vowels [2]. Moreover, some vowels are difficult to pronounce because the vowel is the type of sound, depending largely on very slight variations of tongue position. Speakers or learners cannot determine for themselves where their tongues are. There is no constriction which we can feel with any precision as we do in consonants. So vowels are most easily described in terms of auditory relationships, or terms of the position of the highest point of the tongue and the position of the lips [3]. To explain or to teach vowel pronunciation effectively, experts or linguists such as phonologists, speech therapist or experienced native speakers with special techniques and extra instruments [4],[5] are required.

Thai language is a tonal language which one syllable combines of Consonant + Vowel (CV) or Consonant + Vowel + Consonant (CVC) [6], [7]. C can be an initial consonant and a final consonant. C(C) is a consonantal cluster. Thai simple vowels are separated into short and long vowels. V is a short vowel. V(V) is a long vowel [8]. Each Thai syllable always starts with a consonant and a tone [9]. Basically, the vowel in each syllable is a nucleus of its syllable [10] and it is the most significant part of the speech event. There are 18 simple monophthongs vowels in Thai vowel system [11] which are divided into 9 short and 9 long. The short and long pairs of vowels are quantitatively different (duration), but they are qualitatively similar (frequency). So, it can be seen that Thai vowels have complex pronunciation. There is no doubt that Thai vowels are much more very difficult for Thai beginner learners or non-native Thai speakers to pronounce Thai vowels properly and correctly.

In this paper, we propose the model of vowels speech recognition for Thai language by using Convolutional Neural Networks (CNNs) as it is one of the most popular deep neural networks. Recently, Convolutional Neural Networks (CNNs) have been applied for Automatic Speech Recognition (ASR) and showed higher performance. CNNs can reduce frequency variant in robustness models. There are a few studies in noise-robust Thai speech recognition, especially for Thai vowels. Since Thai vowel data set are not available for public usage, the data set has recently been collected from 50 Thai native speakers (25 male's voice and 25 female's voice). This paper focuses on apply CNN model in 18 Thai vowels speech recognition with real-world noise.

Remaining part of the paper is prepared as follows: Related work is shown in section II. Section III describes the dataset and the model architecture in our experiments. Section IV reports the results that we obtained. Finally, conclusions are presented.

II. RELATED WORK

There are many research works of vowels have been proposed that can be found in [12], [13], [14], [15], [16]. Recently, Convolutional Neural Networks (CNNs) model has been applied in not only computer vision but also speech recognition. In speech recognition tasks applied the benefits of the CNNs model in various works because CNNs can be used to reduce frequency/spectral variations. CNN for Automatic speech recognition (ASR) [17] used a variety of strategies, such as pooling and weight sharing, which was a technique used in CNN architecture, therefore results were improved. In a small-footprint keyword spotting (KWS) task [18], CNN architecture was used for 14 phrases classification ('answer call', 'decline call', 'email guests', 'fast forward', 'next playlist', 'next song', 'next track', 'pause music', 'pause this', 'play music', 'set clock', 'set time', 'start timer', and 'take note'). It provided 27 to 44% relative improvement in the false reject rate. Large-scale Speech Tasks [19] and Noise Robust Speech Recognition [20], researches proposed the best CNN architecture and strategies. In Large-scale Speech Tasks, achieved a word error rate (WER) of 12% to 14% relative improvement on 3 Large Vocabulary Continuous Speech Recognition (LVCSR) tasks, These tasks

* Corresponding author.

were a 50 and a 400-hour Broadcast News (BN) task and a 300-hour Switchboard (SWB) task respectively. Noise Robust Speech Recognition on Aurora4 reached 8.81% in WER. It also achieved 10.0% relative reduction over the traditional CNN on AMI meeting transcription task. The robustness of CNN acoustic models [21] was used with two techniques to increase performance; autoregressive moving average spectrogram features and channel dropout. Channel dropout method reached 16% in WER with ARMA features and 20% with FBANK features over the baseline CNN. Combination of models was used to increase the efficiency of speech recognition on LVCSR task [22]. Model architecture consisted of CNNs, LSTMs, and DNNs which was called CLDNN. The CLDNN achieved 4 to 6% relative reduction in WER over LSTM. Speech recognition tasks in Thai are shown in [23] using the neuro-fuzzy system with Thai 8 words such as forward, back, left, right recorded in a different noisy environment. It showed that each factor has different effects on recognition accuracy. Double Filter Banks for feature extraction and Euclidian distance for the recognition processes [24] were used with Thai basic voice commanding in various conditions from volunteers (9,000 speech) and achieved accuracy rate is about 96.3 %. Speech Classification was experimented on emotion [25] from 2 corpora which were Interactive Emotional Dyadic Motion Capture (IEMOCAP) and Emotional Tagged Corpus on Lakorn (EMOLA). The emotion was classified into four categories including anger, happiness, neutral, and sadness, It found that each emotion used different features. MFCC with Zero Crossing Rate (ZCR) were good for anger and happiness emotion class with result of 81.95% and 69.86% accuracy respectively. CNN was applied to speech emotion classification [26]. The raw input speech signal was extracted by Convolutional Long Short-Term Memory Neural Network (ConvLSTM-RNN model) and it was classified by Support Vector Machines. It was experimented on IEMOCAP database. The result of SVM with Polynomial Kernel in 192 Phoneme archived accuracy rate at 65.13%. For tasks on vowels, CNN was applied to use with Javanese language (which is a language of Indonesia) [27], [28]. Mel-frequency spectral coefficients (MFSC) was used to extract the feature. Dataset consisted of 250 Javanese middle vowels sound file recorded by only one speaker. The output consisted of 5 classes. The result was 94% accuracy. In [29] applied the reduction of the order of Linear Predictive Coefficients (LPC). The reduced set of Critical Band Intensities (CBI) was selected. The optimization was used in short and long vowels classification and unmixed and mixed vowels recognition in Thai spoken language. The voices were collected from 6 speakers. Result of classifying frames for short and long unmixed vowels for the 3 male model, the 3 female model, and the 2 male-2 female model archived accuracy at 89.39, 89.83, and 87.67 respectively. The 1,134 samples were used for training in the male model and female model. The 1,512 voice samples for the mixed-gender model. Our research is inspired by many speech recognition tasks that used CNNs that we applied to noisy Thai vowels.

III. EXPERIMENTS

We design the experiments to determine the suitable parameters in the CNN architecture for noisy Thai vowels recognition. Our research compares some methods using various strategies such as Padding [20] which experimental results showed that padding in feature maps for very deep CNNs was important. It could save the size of feature maps and made more improvements. Dropout [30] could reduce the over-fitting problem. Dropout, ReLU, and DNN were applied on a 50-hour English Broadcast News task, results over a DNN with sigmoid, and a GMM/HMM system 4.2%, 14.4% relative improvement respectively. Batch Normalize [31] was used to support the convolution neural network for faster convergence in training. And increasing the number of convolution layers and hidden units are used to evaluate the performance of the model. After that compare the CNN model with the Multilayer Perceptron (MLP) model and the Support Vector Machines (SVM) model.

A. Dataset

Thai vowels data set is not available for public usage. Therefore, in this research, noisy Thai vowels dataset is the speech of Thai vowels in real-world situations. The sound is collected in a university environment from several areas which consist of many types of noise at 30 - 40 dB SNR (Signal to Noise Ratio) such as vehicles from the road, people talking in the canteen, music at the college of music, wind in the park, and animals sound like dogs and bird. It constrains 16 kHz speech data is recorded from a mobile phone.

The speech of male and female are separately recorded and collected because of the pitch. Following the principle of linguistics study, the pitch of male and female are different: male's pitch is low, but female's pitch is high. The vowel speeches are separated into 2 groups: male's voice and female's voice from 25 male and 25 female. All of them are 20-25 year-old standard Thai speakers. In this research, it constrains 18 classes (9 short vowels and 9 long vowels like Thai simple vowels grouping). Each speaker speaks 2 times each vowel, the total of the collected voice is 1,800 sound files of Thai vowels included of 900 male's files (18 vowels x 25 males x spoken twice) and 900 female's files (18 vowels x 25 females x spoken twice). The 80% of the total files in each group is used for training and the 20% for testing. The sound is collected in normal environments, thus there are many types of noise such as vehicles, people talking, music, wind, animals, and others. 16,000 Hz sampling rate is used for each record in datasets.

TABLE I. THAI SIMPLE VOWEL IN THE INTERNATIONAL PHONETIC ALPHABET (IPA) [6], [7].

Vowels			
Short		***Long***	
Thai letter	***Phonetic***	***Thai letter***	***Phonetic***
อะ	/a/	อา	/aː/
อิ	/i/	อี	/iː/
อึ	/ɯ/	อือ	/ɯː/
อุ	/u/	อู	/uː/
เอะ	/e/	เอ	/eː/
แอะ	/ɛ/	แอ	/ɛː/
โอะ	/o/	โอ	/oː/
เอาะ	/ɔ/	ออ	/ɔː/
เออะ	/ɤ/	เออ	/ɤː/

After the collecting of noisy Thai vowels speech dataset, a Thai linguist has the sound files cut and selected only vowel sound by using linguistic measurement with PRAAT tool. PRAAT is a computer program for analyzing, synthesizing, and manipulating speech developed by Paul Boersma and David Weenink [32]. It is a formidable research and teaching tool for phonetics that is commonly used by linguists in worldwide phonetic researches because it is probably the most comprehensive toolbox, and it is certainly the most affordable with the top-quality graphic representations of speech.

B. Input features

Previous works, CNNs for speech recognition [19] have been defined input features with a size of #times x #frequencies = 11x40. In the research [20], the default input map size for the model was set to 11x40 as well, and researchers experimented with extending the time and the frequency. They received better results with the full-extension model (21x64) and achieved a WER of 9.8%

In this research, the speech signal of Thai vowels was preprocessing by a package for audio and music analysis calls LibROSA library in python. For input feature extraction, librosa.feature.mfcc was used for extracted MFCC features: sampling rate of 16000, number of MFCCs to return are 40 and 64 was set in parameters.

To experiment with the appropriate input features, we initialized the default MFCC features to 11x40, and we extended both time and frequency to find the appropriate value in this research.

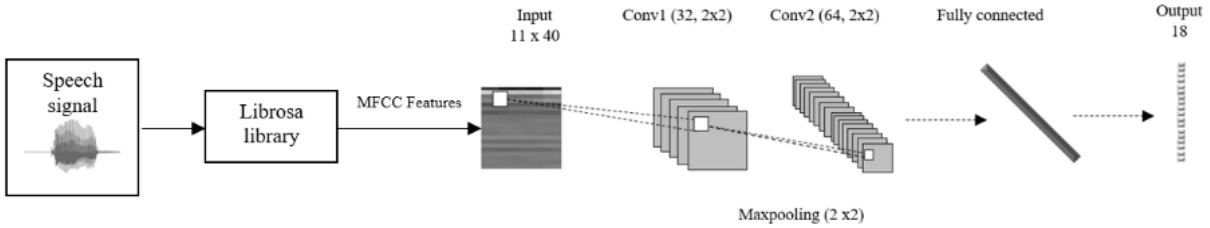


Fig. 1. Feature Extraction and Baseline CNN Architecture

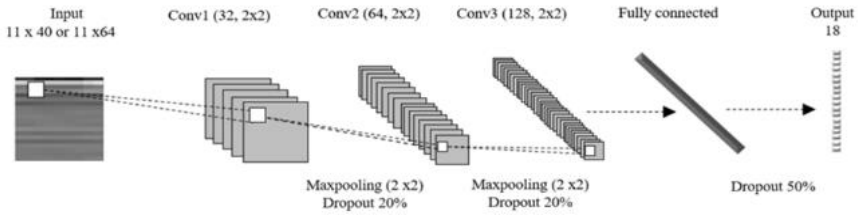


Fig. 2. CNN Architecture of Thai Simple Vowel (CNN_TSV)

C. Baseline structure

In our research, the baseline CNN structure consists of 2 convolutional layers. The first convolutional layer has 32 filters (2x2) followed by max-pooling (2x2), while the second convolutional has 64 filters (2x2) but no pooling layer. ReLU activation function [30] is used in this architecture as it has been used most widely in convolutional neural networks or deep learning and can reduce the calculation time. All of the pooling layers use filter (2x2) and stride of 2. Finally, the fully connected layer uses 64 hidden units and a Softmax activation function. Adam optimizer [11] is used in this model because Adam converges faster and provides high performance. The output classes are 18 class. The Fig. 1. shows details of the baseline CNN model architecture are built for Thai vowels classification.

D. Model architecture

Our CNN architecture of Thai Simple Vowels in this paper is called the CNN_TSV model that derived from the benefits of experimental results section 4.1 to 4.5, and we used it to compare the MLP and SVM models in section 4.6. The CNN_TSV model consists of 3 convolutional layers. The first convolutional layer has 32 filters (2x2) followed by max-pooling (2x2) and dropout 20%, the second convolutional has 64 filters (2x2) followed by max-pooling layer and dropout as the first convolutional layer. While the third convolutional layer has 128 filters (2x2) and uses dropout 20% but no pooling layer. Finally, a fully connected layer uses 64 hidden units, a dropout 50%, and a Softmax activation function. This model uses padding strategy, Adam optimizer, batch normalize is 32. In the female, the appropriate input features are 11x40 and the male are 11x64. This architecture is shown in Fig. 2.

E. Implementation details

For our experiments, we implemented models with python on Keras framework and backend is a TensorFlow. We evaluated our model on Windows 64 bit with Intel CORE i7 CPU, 8 GB memory and Nvidia GeForce GTX 1050 GPU.

IV. RESULTS

This research aims to study the appropriate structure using CNN acoustic modeling for Thai vowels speech recognition. The experimental results are presented in the following tables.

A. Time and frequency extension

The literature review [20], our research had found that using our input features and extending time and frequency were useful for the model. To find the appropriate value for the nosy Thai vowels recognition task, this technique has experimented.

TABLE II. RESULTS OF TIME AND FREQUENCY EXTENSION

Input Features*	Accuracy (%)			
	No padding		*padding*	
	Female	*Male*	*Female*	*Male*
11x40	82.78	76.67	80.00	78.89
11x64	80.00	**80.00**	**80.56**	**80.56**
17x40	78.33	78.33	77.78	78.33
17x64	**83.89**	78.33	78.33	78.89
Avg.	81.25	78.33	79.17	79.17

*Input Features: #times x #frequencies

Table II above presents the experimental results, the appropriate input features for both male and female voices are 11x64. In contrast, female at 17x64 shows the best result at 83.89% .Using padding does not improve the performance (not used with any strategy).

B. Dropout

Based on our experimental results on computer vision that demonstrate effective experiments when using dropout and padding. This paper made more experiments to achieve better performance by using dropout. Moreover, when padding has been used, the results are improved by 5 to 8 % over No padding.

TABLE III. RESULTS OF DROPOUT

Input Features*	Accuracy (%)			
	No padding		*padding*	
	Female	*Male*	*Female*	*Male*
11x40	**87.78**	83.89	87.22	84.44
11x64	87.22	85.56	**88.89**	**87.22**
17x40	85.56	85.56	86.11	83.89
17x64	87.22	**86.67**	86.67	**87.22**
Avg.	86.95	85.42	87.22	85.69

*Input Features: #times x #frequencies

Table III shows the better results by using input features at 11x40, 11x64 and 17x64, therefore these 3 input features are used in the next experiment.

C. Batch normalize

In this section, both with and without batch normalize strategies are tested. Values of batch normalize of 32, 64 and 128 are used respectively.

TABLE IV. RESULTS OF BATCH NORMALIZE (BN)

BN	Accuracy (%)					
	11x40		*11x64*		*17x64*	
	Female	*Male*	*Female*	*Male*	*Female*	*Male*
no	87.22	84.44	**88.89**	87.22	86.67	87.22
32	**88.89**	83.89	87.22	**88.89**	85.56	86.67
64	**88.89**	**86.67**	86.67	86.67	86.11	**88.33**
128	88.33	84.44	86.67	87.22	**87.78**	85.00
Avg.	**88.33**	84.86	87.36	**87.50**	86.53	86.81

Table IV shows the input features (11x40, 11x64 and 17x64) of male and female are taken to find the average value to consider the results. The experiment for the female voice, the appropriate input features are 11x40. Average accuracy at 88.33%. For the male voice, the appropriate input features are 11x64. Average accuracy is 87.50%. The appropriate batch normalize of male and female is 32 which gives 88.89% accuracy.

D. Number of the convolution layer

When extending the convolution layer from 2 to 3, Table V below shows the better results. Especially, the improvement is clear for the female that achieve 90.00%. Although for males voices, the results increase the convolution layer are not different but we believe that adding more convolutional layers will give better results in future experiments.

TABLE V. RESULTS OF NUMBER OF CONVOLUTION LAYER

Number of convolution layer	Accuracy (%)	
	Female (11x40)	*Male (11x64)*
2 layers	88.89	88.89
3 layers	**90.00**	**88.89**

E. Number of hidden units

Our research compares the results of experiments with a different number of hidden units with 3 convolutional layers. Table V provides the effective results that we obtain for both male and female.

TABLE VI. RESULTS WITH DIFFERENT NUMBERS OF HIDDEN UNITS

Number of hidden units	Accuracy (%)	
	Female (11x40)	*Male (11x64)*
64 units	**90.00**	**88.89**
256 units	88.33	87.22
1024 units	86.67	86.11

Table VI, the results conclude that adding hidden units in the fully connected layer does not improve performance, therefore a 64 number of hidden units are used.

F. Comparison between CNN_TSV, MLP and SVM model (k-fold = 10)

The experimental represents comparing the CNN_TSV model with the MLP model and the SVM model. The CNN_TSV model is derived from experimental results section IV (A-E). The result from Fig. 3 shows that the appropriate epochs on the female voice are 500 epochs. The mean of accuracy is 84.86% and the standard deviation is +/- 4.14%. On male voice, the appropriate epochs are 1000 at 89.72% and the standard deviation is +/- 2.79%. The Multilayer Perceptron Classifier (MLP Classifier) is used on a baseline of the MLP model consists of a hidden layer of 256 units. The RELU activation is used in this model, Solver is Adam optimization, 32 batch size and the initial learning rate is 0.001. For the SVM model uses Support Vector Classification (SVC), linear is set to the kernel, and decision function of shape is one-vs-rest ('ovr'). All models use the same input features.

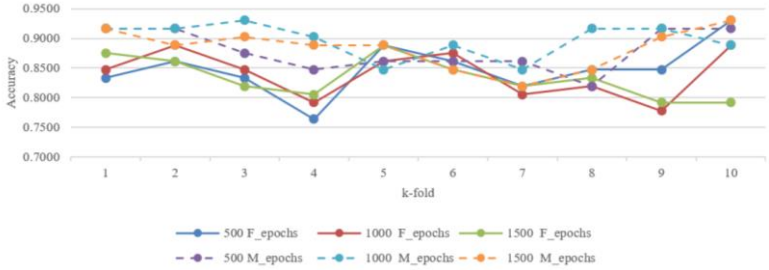


Fig. 3. The result of epochs (500, 1000, 1500) on the female/male voice.

TABLE VII. RESULTS WITH DIFFERENT METHODS

Methods	Mean Accuracy (%)	
	Female (11x40)	*Male (11x64)*
CNN_TSV	**84.86**	**89.72**
MLP	68.12	71.85
SVM	76.00	82.17

As described in Table VII, the results show the CNN_TSV model provides the highest mean accuracy and efficiency for both female and male, 84.86% and 89.72% mean accuracy respectively.

G. The confusion matrix, Precision, Recall, and F1-score of the CNN_TSV model

For error analysis, the confusion matrix of the CNN_TSV model on female and male voice is shown in Fig. 4. From the confusion matrix, the most confusing pair of Thai vowels on female voice is (‘ไอ’ /o:/ and ‘ไอะ’ /o/). Subordinate confusing pair are (‘อือ’ /ɯ:/ and ‘เออ’ /ɤ:/), (‘เออ’ /ɤ:/ and ‘เออะ’ /ɤ/), (‘อิ’ /i/ and ‘ี’ /i:/), (‘อิ’ /i/ and ‘เอะ’ /e/. On male voice, the most confusing pair of Thai vowels are (‘ไอ’ /o:/ and ‘ไอะ’ /o/), as the female voice, and (‘เออ’ /ɤ:/ and ‘เออะ’ /ɤ/). Subordinate confusing pair are (‘ออ’ /ɔ:/ and ‘เอาะ’/ɔ/. The experiment has found that ‘ไอ’ /o:/ and ‘ไอะ’ /o/ vowels are the most confusing pair of both genders. Moreover, in each of the confusing pairs are short and long vowels which contrast in its duration.

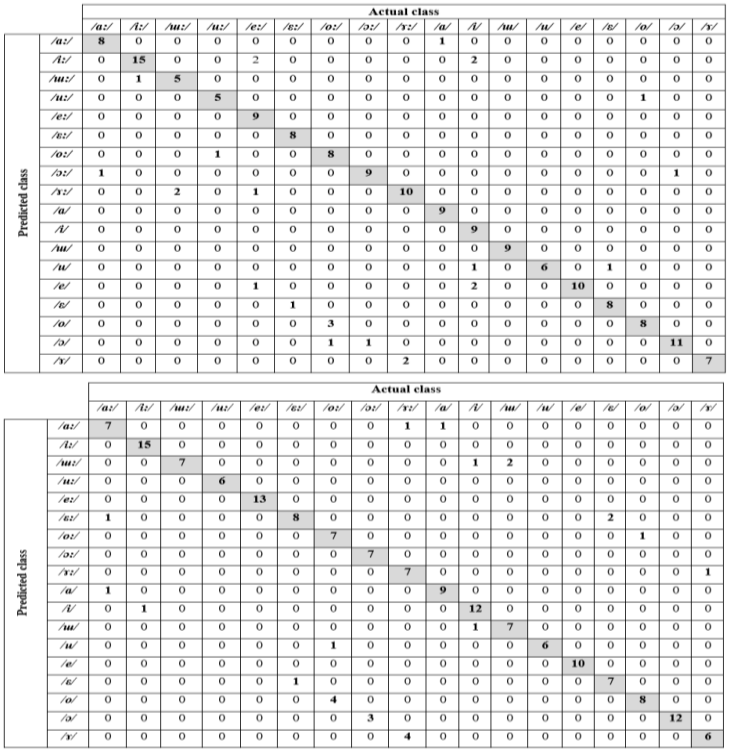


Fig. 4. The confusion matrix of the CNN_TSV model on the female (top)/male(bottom).

TABLE VIII. PRECISION, RECALL, AND F1-SCORE OF THE CNN_TSV MODEL.

Thai Vowels	Female			Male		
	Precision	*Recall*	*F1-score*	*Precision*	*Recall*	*F1-score*
a:	0.89	0.89	0.89	0.78	0.78	0.78
i:	0.79	0.94	0.86	1.00	0.94	0.97
ɯ:	0.83	0.71	0.77	0.70	1.00	0.82
u:	0.83	0.83	0.83	1.00	1.00	1.00
e:	1.00	0.69	0.82	1.00	1.00	1.00
ɛ:	1.00	0.89	0.94	0.73	0.89	0.80
o:	0.89	0.67	**0.76**	0.88	0.58	**0.70**
ɔ:	0.82	0.90	0.86	1.00	0.70	0.82
ɤ:	0.77	0.83	0.80	0.88	0.58	**0.70**
a	1.00	0.90	0.95	0.90	0.90	0.90
i	1.00	0.64	0.78	0.92	0.86	0.89
ɯ	1.00	1.00	1.00	0.88	0.78	0.82
u	0.75	1.00	0.86	0.86	1.00	0.92
e	0.77	1.00	0.87	1.00	1.00	1.00
ɛ	0.89	0.89	0.89	0.88	0.78	0.82
o	0.73	0.89	0.80	0.67	0.89	0.76
ɔ	0.85	0.92	0.88	0.80	1.00	0.89
ɤ	0.78	1.00	0.88	0.60	0.86	0.71

Table VIII presents the precision, recall, and f1-score of the CNN_TSV model for classifying each vowel. The lowest of F1 score on the female voice is ‘ไอ’ /o:/ (0.76) , and the male voice is ‘ไอ’ /o:/ (0.70) and ‘เออ’ /ɤ:/ (0.70). The f1-score results are relevant with the confusion matrix. On the other hand, the highest of F1 score (1.00) on the female voice is ‘อึ’ /ɯ/, and the male voice is ‘อู’ /u:/, ‘เอ’ /e:/, and ‘เอะ’ /e/.

V. CONCLUSIONS

This research presents a nosy Thai vowels speech recognition task using a CNN acoustic model. Our research experiment on a newly collected noisy data set. This dataset consists of 25 male and 25 female voice records. The voices are grouped into 18 classes for each gender. The evaluation is done with time and frequency expansion. We have found that the appropriate input features are 11x40 for female voices and 11x64 for male voices. Padding and Dropout are used by 20%, it improves the performance by 5% - 8%. The appropriate value of Batch normalize strategy is at 32. Increasing the convolution layers to 3 gives a better to be performed in both groups. The appropriate of hidden units is 64. The most accuracy rate of CNN_TSV model is 90.00% and 88.89% on female and male voices respectively.

Finally, the comparison of the CNN_TSV model with MLP and SVM model that the CNN_TSV model is the most effective for both female and male voices. The mean accuracies reached are 84.86% and 89.72% respectively. The most confusing pair of both genders are ‘ไอ’ /o:/ and ‘ไอะ’ /o/ vowels. The highest of F1 score (1.00) on the female voice is ‘อึ’ /ɯ/, and the male voice is ‘อู’ /u:/, ‘เอ’ /e:/, and ‘เอะ’ /e/.

REFERENCES

[1] B. G. Evans and W. Alshangiti, "The perception and production of British English vowels and consonants by Arabic learners of English," *J. Phon.*, vol. 68, pp. 15–31, 2018.

[2] L. Rallo Fabra and J. Romero, "Native Catalan learners' perception and production of English vowels," *J. Phon.*, vol. 40, no. 3, pp. 491–508, 2012.

[3] K. J. Peter Ladefoged, *A Course in Phonetics*, Sixth. Michael Rosenberg.

[4] X. Peng, H. Chen, L. Wang, and H. Wang, "Evaluating a 3-D virtual talking head on pronunciation learning," *Int. J. Hum. Comput. Stud.*, vol. 109, no. August 2017, pp. 26–40, 2018.

[5] M. Tabain and R. Beare, "An ultrasound study of coronal places of articulation in Central Arrernte: Apicals, laminals and rhotics," *J. Phon.*, vol. 66, pp. 63–81, 2018.

[6] L. Jeerapradit, A. Suchato, and P. Punyabukkana, "HMM-based Thai Singing Voice Synthesis System," in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, 2019, pp. 1–4.

[7] S. Aunkaew, M. Karnjanadecha, and C. Wutiwiwatchai, "Constructing a phonetic transcribed text corpus for Southern Thai dialect Speech Recognition," in *Proceedings of the 2015 12th International Joint Conference on Computer Science and Software Engineering, JCSSE 2015*, 2015, pp. 69–73.

[8] A. Munthuli, C. Tantibundhit, C. Onsuwan, K. Kosawat, and C. Wutiwiwatchai, "Frequency of occurrence of phonemes and syllables in Thai: Analysis of spoken and written corpora," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 2015, pp. 3–7.

[9] K. Supphanat, "Syllable Structure Based Phonetic Units for Context-Dependent Continuous Thai Speech Recognition," in *EUROSPEECH*, 2003, pp. 797–800.

[10] R. Grammar, *Thai Reference Grammar*. U.S. Government Printing Office, 1964.

[11] S. K. Gouda, S. Kanetkar, D. Harrison, and M. K. Warmuth, "Speech Recognition: Keyword Spotting Through Image Recognition," 2018. [Online]. Available: http://arxiv.org/abs/1803.03759.

[12] Š. Šimáčková and V. J. Podlipský, "Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility," *Res. Lang.*, vol. 16, no. 2, pp. 169–191, 2018.

[13] S. Sahatsathatsana, "Pronunciation Problems of Thai Students Learning English Phonetics: A Case Study at Kalasin University," *J. Educ.*, vol. 11, no. 4, pp. 67–84, 2017.

[14] P. Ghaffarvand Mokari and S. Werner, "Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels," *Lingua*, vol. 185, pp. 81–95, 2017.

[15] K. Mirzaei, H. Gowhary, A. Azizifar, and Z. Esmaeili, "Comparing the Phonological Performance of Kurdish and Persian EFL Learners in Pronunciation of English Vowels," *Procedia - Soc. Behav. Sci.*, vol. 199, pp. 387–393, 2015.

[16] M. Navehebrahim, "An Investigation on Pronunciation of Language Learners of English in Persian Background: Deviation Forms from the Target Language Norms," *Procedia - Soc. Behav. Sci.*, vol. 69, no. Iceepsy, pp. 518–525, 2012.

[17] S. Newatia and R. K. Aggarwal, "Convolutional Neural Network for ASR," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 638–642.

[18] T. N. Sainath and C. Parada, "Convolutional Neural Networks for Small-footprint Keyword Spotting," in *Interspeech*, 2015, pp. 1478–1482.

[19] T. N. Sainath *et al.*, "Deep Convolutional Neural Networks for Large-scale Speech Tasks," *Neural Networks*, vol. 64, pp. 39–48, 2015.

[20] Y. Qian and P. C. Woodland, "Very Deep Convolutional Neural Networks for Robust Speech Recognition," *IEEE/ACM Trans. AUDIO, SPEECH, Lang. Process.*, vol. 24, no. 12, pp. 2263–2276, 2016.

[21] G. Kovács, L. Tóth, D. Van Compernolle, and S. Ganapathy, "Increasing the robustness of CNN acoustic models using autoregressive moving average spectrogram features and channel dropout," *Pattern Recognit. Lett.*, vol. 100, pp. 44–50, 2017.

[22] T. N. Sainath, O. Vinyals, A. Senior, and N. York, "Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4580–4584.

[23] K. Srijiranon and N. Eiamkanitchat, "Thai speech recognition using Neuro-fuzzy system," in *ECTI-CON 2015 - 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2015, pp. 1–6.

[24] P. Phokharatkul, K. Nantanitikorn, and S. Phaiboon, "Thai speech recognition using Double filter banks for basic voice commanding," in *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering*, 2010, vol. 6, pp. 33–36.

[25] P. Sukhummek, S. Kasuriya, T. Theeramunkong, C. Wutiwiwatchai, and H. Kunieda, "Feature Selection Experiments on Emotional Speech Classification," in *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2015, pp. 1–4.

[26] N. Kurpukdee, T. Koriyama, and T. Kobayashi, "Speech Emotion Recognition using Convolutional Long Short-Term Memory Neural Network and Support Vector Machines," 2017, no. December, pp. 1744–1749.

[27] C. K. Dewa and Afiahayati, "Suitable CNN Weight Initialization and Activation Function for Javanese Vowels Classification," *Procedia Comput. Sci.*, vol. 144, pp. 124–132, 2018.

[28] C. K. Dewa, "Javanese vowels sound classification with convolutional neural network," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2016, pp. 123–128.

[29] N. Suktangman, K. Khanthavivone, and K. Songwatana, "Optimizing vowel recognition in Thai spoken language using reduced LPC spectrum and reduced feature set of critical band intensities," in *2006 International Symposium on Communications and Information Technologies, ISCIT*, 2006, no. 4, pp. 128–132.

[30] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8609–8613.

[31] L. Wenjie, G. Cheng, F. Ge, P. Zhang, and Y. Yan, "Investigation on the Combination of Batch Normalization and Dropout in BLSTM-based Acoustic Modeling for ASR," in *Interspeech 2018*, 2018, vol. 2018, pp. 2888–2892.

[32] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," *Glot Int.*, vol. 5, no. 9–10, pp. 341–347, 2001.

Unsupervised Multilingual Ontology Learning

***Abstract*—Multilinguality poses a big challenge to the growth of semantic web. In order to develop multilingual applications we need to develop ontologies which can be shared across languages. In this paper we propose an unsupervised learning algorithm to automatically learn multilingual ontology from unstructured text. We propose three different approaches for multilingual ontology learning, Dictionary based method,parallel corpus based method and Latent Dirichlet Allocation (LDA) based method. While the first two approaches require availability of dictionary and parallel corpus, the LDA based approach does not require any special resource. We have conducted our experiments for two languages, English and Hindi, however the proposed method is general enough to be adopted for other languages also.**

***Index Terms*—Ontology learning, Expectation-Maximization, multilingual ontology, unsupervised learning**

I. Introduction

Ontology, defined as 'Explicit specification of conceptualization' [12], remains at the heart of Semantic Web. A lot of research is being done to (semi) automatically construct ontology from structured or unstructured text. Conceptualization of an ontology extracted (automatically or semiautomatically) from natural language text is often constrained by lexical space of the language. As the lexico-semantic structure of languages vary, the ontologies constructed (for same domain) from different language may also very. As noted by [19] 'No matter how expressive ontologies might be, they are all in fact lexical representations of concepts'. In order to overcome the boundaries of languages and build a language independent ontology, we need to merge evidences from multiple languages.

In this paper we present a multilingual ontology extraction system. We carry out our experiments with the following multi-fold aims,

- *How to develop an ontology which can be shared across languages?:* Lexicalization of concepts varies across languages. This creates lexical gaps and difference in conceptual structure. In order to share ontology across languages we need to merge lexical and semantic structure of different languages and build a language neutral conceptualization.
- *Can evidences from multiple languages improve automatic ontology learning process?* [17] have shown that evidences from multiple languages improves taxonomy learning. However the evidences used for ontology extraction in monolingual setting are mostly language specific and can not be extended across languages. We aim to develop an ontology learner which can compare and relate concepts across languages and can integrate evidences from heterogeneous sources e.g. WordNet, Lexical Patterns, distributional similarity, etc.
- *Can we improve ontology learning process for a language with limited resources by consulting other language with rich resources?* Ontology learning process often involves use of rich resources like WordNet, Wikipedia, etc and tools like Parser. If some language does not have this resources, can we use resources of languages like English to build ontology for resource poor language?

In order to address above mention issues, we propose an ontology learning system which can learn ontology by merging evidences from different languages. we present three different approaches of multilingual ontology learning: Bilingual Dictionary based, Parallel Corpus based, and Comparable corpus based. The proposed ontology learning algorithm is completely unsupervised and does not require sophisticated NLP tools or resources.

We have conducted our experiments for two language, English and Hindi and for two domains, tourism and health. We have restricted our experiments to detect taxonomic relations and build concept hierarchy, however the proposed system can be adopted to construct more complex ontologies by detecting non-taxonomic relation.

The remaining of the paper is organized as follows, Section 2 describes related work, section 3 describes an expectation-maximization based algorithm for ontology learning, section 4 presents multilingual ontology learning process, experiments and observations are discussed in section 5 and 6.

II. Related Work

In this section, we describe various ontology learning methods for mono lingual setting and multilingual setting.

A. Ontology Learning: Monolingual

Ontology learning approaches can be divided into three categories: heuristic based, statistical and hybrid techniques. Heuristic approach [14], [2], [11] primarily relies on the fact that ontological relations are typically expressed in language via a set of linguistic patterns. [14] outlined a variety of lexico-syntactic patterns that can be used to find out ontological relations from a text. [2] used a pattern-based approach to find out part-whole relationships from the text. Heuristic approaches rely on language-specific rules which can not be transferred from one language to another.

Statistical methods relate concepts based on distributional hypothesis [13]. Context vectors of two concepts are compared using various similarity measures. All these measures give semantic relatedness between a pair of concepts, however they do not identify the type of semantic relation. More recently there has also been work on finding out relation between a pair of concepts using directional distributional similarity [18],

	feature	Description	formula
f_1	Cosine Similarity	Cosine similarity between word w_1 and w_2 is calculated by comparing the vectors of words.	$cosine(w_1, w_2) = \frac{\overrightarrow{V}(w_1) \cdot \overrightarrow{V}(w_2)}{\lvert\overrightarrow{V}(w_1)\rvert\lvert\overrightarrow{V}(w_2)\rvert}$
f_2	Weeds Precision	This measure quantifies the weighted inclusion of the features of a term w_1 within the features of a term w_2. [32], [18].	$WP(w_1, w_2) = \frac{\sum_{f \in F(w_1) \bigcap F(w_2)} w_1(f)}{\sum_{f \in F(w_1)} w_1(f)}$
f_3	cosWeeds	This measure corresponds to the geometrical average of Weeds Precision and cosine similarity between words w_1 and w_2	$CW(w_1, w_2) = \sqrt{cosine(w_1, w_2).WP(w_1, w_2)}$
f_4	ClarkeDE	This measure is a close variation of Weeds Precision, proposed by [6].	$CDE(w_1, w_2) = \frac{\sum_{f \in F(w_1) \bigcap F(w_2)} min(w_1(f), w_2(f))}{\sum_{f \in F(w_1)} w_1(f)}$
f_5	Frequency Ratio	We use frequency ratio to measure degree of generality of a word. The measure is based on following hypothesis, 'A more general term appears more frequently in the corpus, while a more specific term appears less frequently' [28]	$fratio(w_1, w_2) = \frac{f(w_1)}{f(w_2)}$
f_6	Head Word heuristic Pattern	This pattern finds hypernymy relation from noun phrase. e.g. "Heritage Hotel" is a "Hotel"	(NP)*NP is hyponym of (NP)
f_7	Neighbor Pattern	This pattern detects neighbor (Co-hyponymy) relation. e.g. Delhi, Mumbai, Calcutta are cities.	$((NP) * (NP)(CC\|,)) * (NP)$
f_8	WordNet hypernym	This formula calculates probability of hypernymy by consulting WordNet	$\frac{hypernym(w_1, w_2)}{totalRelation(w_1, w2)}$
f_9	WordNet Synonym	This formula calculates probability of synonymy by consulting WordNet	$\frac{synonym(w_1, w_2)}{totalRelation(w_1, w2)}$
f_{10}	WordNet Neighbor	This formula calculates probability of co-hyponymy by consulting WordNet	$\frac{co-hyponym(w_1, w_2)}{totalRelation(w_1, w2)}$

TABLE I
FEATURES FOR ONTOLOGY LEARNING

[20]. [16] performed semantic clustering to find semantically similar nouns. [27] proposed a divisive clustering method to induce noun hierarchy from an encyclopedia.

Hybrid approaches leverage the strengths of both statistical and heuristic based approaches and often use evidences from existing knowledge bases such as WordNet, Wikipedia, etc. Table I shows various measures which can be used as features to compare concepts and learn ontology.

[3] combined the lexico-syntactic patterns and distributional similarity based methods to construct ontology. The calculate cosine similarity between vectors of a pair of nouns and used for hierarchical bottom-up clustering. Hearst-patterns are used to detect hypernymy relation between similar nouns. In a similar approach, [5] clustered nouns based on distributional similarity and used Hearst-patterns, WordNet [10] and patterns on the web as a hypernymy oracle for constructing a hierarchy. [9] used Wikipedia to extract ontology for different languages. [31] have used probabilistic topic modeling framework to learn terminological ontology.

B. Ontology Learning: Multilingual

While a lot of work for ontology learning is done for monolingual setting, not much work is done in multilingual setting. [17] have shown benefits of merging evidences from multiple language. [17] have used hierarchical clustering to learn ontology from comparable corpora. The approach for building the multilingual model presented in this work assumes that a domain-specific bilingual dictionary is available. In a similar approach, [33] have used Affinity Propagation clustering (AP clustering) algorithm to process clustered terminologies.

There has also been efforts to construct multilingual knowledge base by using existing resources. recently, [21] have used multilingual documents of Wikipedia to construct multilingual knowledge base. Similarly, [24] have constructed a multilingual lexical knowledge network by combining multilingual information from WordNet and Wikipedia. Apart from this, [30] have tried to relate words across languages using latent topic vector.

C. Proposed Approach

Almost all of the existing multilingual ontology extraction work relies on availability of bilingual lexicon or parallel corpus and extracts taxonomy using clustering techniques. They do not use pattern based measures or other sources (e.g. Wordnet) to detect semantic relations. The key contribution of our work is as follows,

- The proposed approach is completely unsupervised and does not rely on availability of bilingual dictionary.
- The proposed system does not require availability of sophisticated NLP tools or language specific resource and hence can be easily adopted for resource poor languages.
- We describe three different scenarios with different degree of resource available, any of which can be used based on the resource availability.
- The proposed framework is general enough to accommodate different measures of semantic relation detection.

III. ONTOLOGY LEARNING PROCESS

We have used an Expectation Maximization based algorithm for domain specific ontology learning. The various measures used to detect relation between a pair of concepts are as shown in table I. The expectation maximization algorithm consults these features and predicts relation between the pair of concepts. The general process of monolingual ontology learning is as described in Fig. 1

The input text is first pre-processed to build context vector for each term. Preprocessing involves POS tagging, morph analysis, stop words removal. Key terms from the corpus are extracted using pattern based method. Lexical pattern $(NP)^*(NP)$ is applied to extract key phrases from the corpus. Terms are filtered out using weirdness measure [1]. For each term the context vector is created using bag of word approach. Co-occurrence is calculated using Point-wise Mutual Information [4] measure.

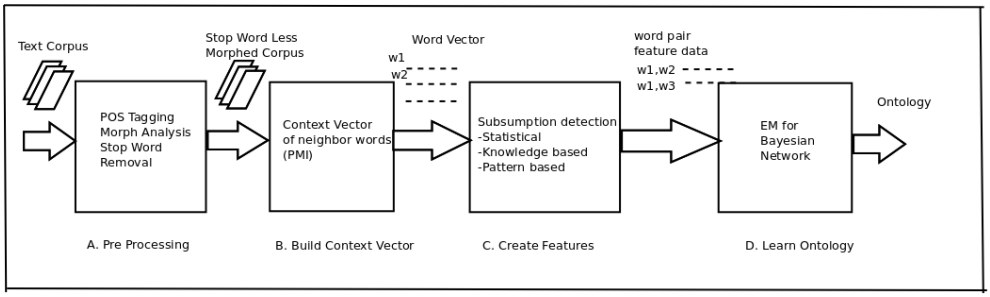


Fig. 1. Ontology Learning Process

After preprocessing and context vector construction, values of different features are calculated as per the formula shown in table I. Table II shows example output after feature calculation. Column f_1 to f_{12} shows the calculated value for the features (defined in table I). columns H, S, N corresponds to relations (to be predicted) Hypernym, synonyms and neighbors respectively. The first row describes caculated features values for word 'haemorrhagic fever' and 'dengue haemorrhagic fever'. As shown in Table II the output of different features does not match and they often contradict. Our aim is to predict the correct relation between the pair of words using the observed values of features.

We model this as a Bayesian Network Learning Problem, where observed variables (x) are the features we calculate and hidden variables (y) are the relation that we want to predict. For any relation label y paired with feature attribute values $x_1...x_d$, likelihood for complete training set n can be calculated using following formula,

$$L(\theta) = \sum_{i=1}^{n} log \sum_{y=1}^{k} (P(Y^{(i)} = y) \prod_{j=1}^{d} P(X_j^{(i)} = x | Y^{(i)} = y)) \tag{1}$$

Here, $P(Y = y)$ and $P(X_j = x_j | Y = y)$ for $j = 1, 2..., d$ are network parameters. Let's define parameter vector θ as a vector consisting values of these parameters.

1) EM Algorithm: The EM algorithm starts by randomly choosing the initial parameter values θ_0. At each iteration value of hidden variable $Y^{(i)}$ is calculated as a function of the training set and the previous parameter values θ_{t-1}; and then new parameter values θ_t are updated using the observed variables and previously estimated hidden variables. [15], [7].

2) E-Step: For the given value of θ, E-Step calculates probability of hidden variable for each example X_i using following formula,

$$\delta(y|i) = p(Y^{(i)} = y | X^{(i)}; \theta_{t-1})$$

$$\delta(y|i) = \frac{P(Y^{(i)} = y) \prod_{j=1}^{d} P(X_j^{(i)} = x_j / Y^{(i)} = y)}{\sum_{y=1}^{k} \left(P(Y^{(i)} = y) \prod_{j=1}^{d} P(X_j^{(i)} = x_j / Y^{(i)} = y) \right)}$$

$\delta(y|i)$ is defined to simplify notation of M-Step. The value for $\delta(y|i)$ is the conditional probability for label y on the ith example, given the parameter values θ_{t-1}. (denominator in the equation of $p(Y^{(i)} = y | x^{(i)}; \theta_{t-1})$ ensures that $\sum_{y=1}^{k} P(Y = y) = 1$).

Expectation of Y is calculated using following formula

$$E(Y = y) = \sum_{i=1}^{n} \delta(y|i) \tag{2}$$

3) M-Step: The second step at each iteration is to calculate the new parameter values (elements of vector θ), as

$$P(Y = y)^t = \frac{\sum_{i=1}^{n} \delta(y|i)}{n} \tag{3}$$

$$where, n = total number of examples \tag{4}$$

$$p(X_j = x | Y = y)^t = \frac{\sum_{i=1:X_j^i = x}^{n} \delta(y|i)}{\sum_{i=1}^{n} \delta(y|i)} \tag{5}$$

IV. Multilingual Ontology Learning

This section presents three different approaches of Multilingual Ontology learning, Bilingual dictionary based, Parallel corpus based and comparable corpus based.

A. Bilingual Dictionary Based Cross Lingual Ontology Learning

In this section we are presenting a cross lingual ontology learning method through which ontology can be learned for a language using the resources available for other language. We present a scenario in which a mono lingual corpus and bilingual dictionary are available. The overall learning process is as shown in the Fig 2.

Let's say, language A is the resource poor language for which we want to learn ontology using resources of language B. The basic conceptual structure is constructed using the contextual evidences of language A. Domain specific terms and their context vector are constructed using the monolingual corpus available in language A. Then, for each pair of word $(w_1, w_2)^A$ in source language a corresponding word pair $(w_1, w_2)^B$ is found in the target language using bilingual dictionary. The knowledge base (e.g. WordNet) and patterns of language B are used to detect relation between the pair of words. Evidences from both languages are then merged to build feature vector.

i	Word Pair	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	H	S	N
1	haemorrhagic fever - dengue haemorrhagic fever	1	1	1	1	1	1	0	0	0	0	0	0	?	?	?
2	leptospirosis-kalaazar	2	0.39	0.24	0.24	0.30	0	0	1	0	0	0	0	?	?	?
3	transplant - transplantation	1.91	0.20	0.14	0.14	0.17	0	0	0	0.33	0	0	0	?	?	?
.	...	...	...	...	...	...	...	...	...	...	...	...	?	?	?	
n	cannabis - marijuana	1.6	0.18	0.14	0.14	0.16	0	0	0	0.25	0	0.25	0	?	?	?

TABLE II
EXAMPLE DATASET

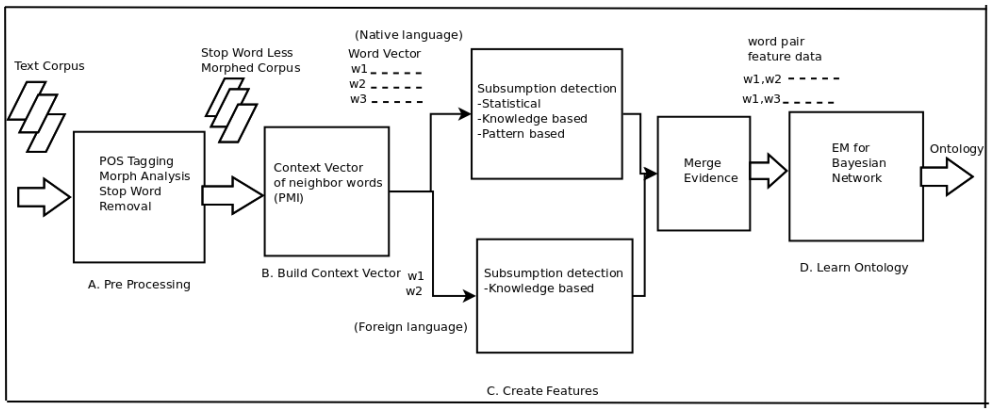


Fig. 2. Cross lingual Ontology Learning Process (bilingual dictionary based approach)

This approach is particularly useful when resources like WordNet, Wikipedia etc are not available for a language. Also, Named entity recognition improves by checking translation of each term. We call this approach cross lingual, because the context vector and statistical relation between word pair is created from a monolingual corpus and then patterns and WordNet of other languages are used to detect semantic relation.

B. Parallel Corpus based Approach for Multilingual Ontology learning

This section describes a method to construct a multilingual ontology by merging conceptualizations of two languages. The proposed approach builds multilingual context vector using parallel corpus. The overall ontology learning process is as shown in Figure 3.

As shown in Figure 3, we construct feature vector for both *native language* and *foreign language* separately. In order to merge these feature vectors we learn word alignment from the parallel corpus. We have used giza++ [26] to learn word alignment. Once the features are constructed in both languages, word alignment is used to find out translation word pair $(w_1, w_2)^t$ for the source word pair $(w_1, w_2)^s$. Then the feature values are added and normalized to construct merged feature set for a concept pair.

The various statistical measures described in previous section are applied on multilingual feature vector to relate concepts. Knowledge based and lexical patterns of both languages are used to identify relationship between concepts.

domain	language	Total Sentences	Unique Words
Health	Hindi	25000	61000
	English	25000	69000
Tourism	Hindi	48000	89000
	English	59000	121000

TABLE III
CORPUS DETAILS

C. Comparable Corpus Based Approach

In order to relate concepts across languages we need to prepare a language independent description for concepts. We rely on Latent Dirichlet Allocation (LDA) based topic model [8] for this. LDA provides a language independent latent topic distribution for each word. We first create word topic distribution from bilingual corpus using BiLDA [25], [22], [29]. BiLDA provides a common latent topic distribution for words of different languages. We use these word topic distribution to compare concepts across languages.

Fig 4 shows the overall learning process. The word topic vector created using BiLDA is used to compare concept across languages and features mentioned in previous section are calculated using this 'word-topic' vector.

V. EXPERIMENT AND EVALUATION

For english we have used *morpha* [23] morph analyzer and stanford POS tagger [1], for hindi we have used morph analyzer and pos tagger developed at IITB [2].

[1]http://nlp.stanford.edu/software/tagger.shtml
[2]http://www.cfilt.iitb.ac.in/Tools.html

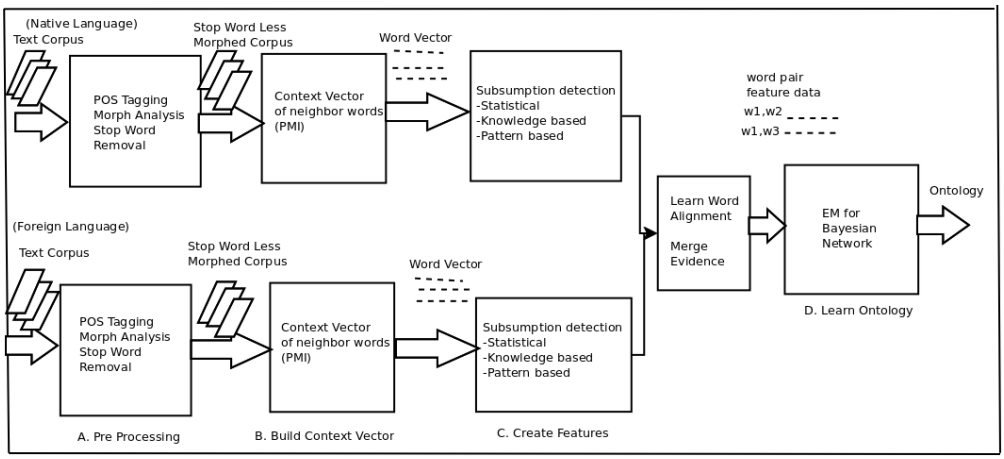


Fig. 3. Multi lingual Ontology Learning Process (Parallel corpus based approach)

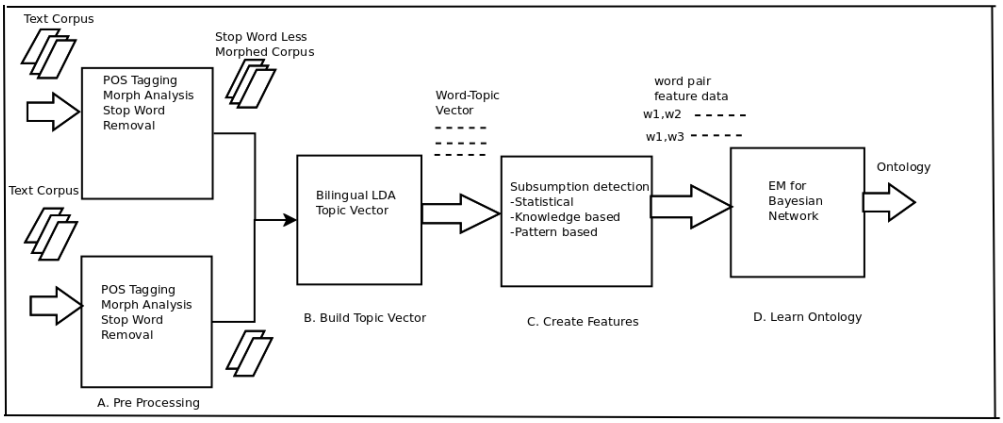


Fig. 4. Multi lingual Ontology Learning Process (Comparable corpus based approach)

We have conducted experiments for two domains, Health and Tourism and for two languages, English and Hindi. We choose English as resource rich language and Hindi as resource poor language. Table III shows the details of the corpora. We preporcess the english corpus using Stanfor POS tagger and Morpha and Hindi corpus with CFILT POS Tagger. After extracting key terms and building context vector the features are extracted as described earlier.

We have carried out four different experiments for both domains, Monolingual, Dictionary based cross lingual, parallel corpus based multilingua and comparable corpus based. The experiments were aimed to detect four different relation classes, hypernym, synonym, neighbor and no relation. Table IV summerizes the results of each experiments.

As shown in Table IV overall precision and recall is improved for dictionary based and parallel corpus based approaches. English monolingual performs better than Hindi monolingual since the health domain contains many borrowed words which are transliterated in Hindi. This is also a reason why cross lingual dictionary based approach performs better than Hindi monolingual. The LDA based approach performs reasonably well considering that it does not require parallel corpora or dictionary. However, LDA based approach does not work for multiword terms or noun phrases.

VI. Conclusion

We have presented three different approaches for multilingual ontology learning. We observe that consulting evidences from more than one language helps overall ontology learning process. Multilinguality helps in building ontology for resource poor language by taking support of resounce rich language, it also helps in better term extraction and relation extraction. We have used limited set of features and expectation maximization algorithm, however the techniques discussed here can be used for any set of features. The proposed method is completely unsupervised and can be used for any language.

References

[1] Ahmad, K., Gillam, L., Tostevin, L., Group, A.: Weirdness indexing for logical document extrapolation and retrieval (wilder). In: The Eighth Text REtrieval Conference (1999)

Domain	Language	Hypernym			Neighbor			Synonym			No Relation			Average		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Health	Hindi	0.59	0.84	0.69	0.66	0.65	0.66	0.42	0.81	0.55	0.82	0.46	0.59	0.62	0.69	0.66
	English	0.62	0.85	0.72	0.61	0.70	0.65	0.58	0.75	0.66	0.81	0.50	0.62	0.66	0.70	0.68
	Dicitonary based	0.68	0.82	0.74	0.69	0.76	0.72	0.62	0.83	0.71	0.79	0.52	0.63	0.70	0.73	0.71
	Parallel Corpus Based	0.58	0.89	0.70	0.64	0.72	0.68	0.47	0.71	0.57	0.76	0.69	0.65	0.61	0.69	0.65
	LDA based	0.46	0.84	0.59	0.57	0.67	0.62	0.41	0.77	0.54	0.80	0.67	0.61	0.56	0.67	0.61
Tourism	Hindi	0.54	0.85	0.66	0.62	0.60	0.61	0.3	0.63	0.41	0.8	0.43	0.55	0.57	0.63	0.6
	English	0.56	0.79	0.65	0.54	0.65	0.59	0.63	0.75	0.68	0.76	0.47	0.58	0.62	0.66	0.64
	Dicitonary based	0.62	0.74	0.68	0.54	0.66	0.59	0.66	0.78	0.72	0.72	0.48	0.57	0.63	0.67	0.65
	Parallel Corpus Based	0.51	0.82	0.63	0.62	0.69	0.65	0.49	0.71	0.58	0.76	0.42	0.54	0.60	0.66	0.63
	LDA based	0.43	0.83	0.57	0.54	0.62	0.58	0.39	0.71	0.50	0.78	0.38	0.51	0.54	0.63	0.58

TABLE IV
EXPERIMENT RESULTS

[2] Berland, M., Charniak, E.: Finding parts in very large corpora. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. pp. 57–64. ACL '99, Association for Computational Linguistics, Stroudsburg, PA, USA (1999)

[3] Caraballo, S.A.: Automatic construction of a hypernym-labeled noun hierarchy from text. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics. pp. 120–126 (1999)

[4] Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. Comput. Linguist. **16**(1), 22–29 (mar 1990)

[5] Cimiano, P.: Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)

[6] Clarke, D.: Context-theoretic semantics for natural language: An overview. In: Proceedings of the Workshop on Geometrical Models of Natural Language Semantics. pp. 112–119. GEMS '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)

[7] Collins, M.: The naive bayes model, maximum-likelihood estimation, and the em algorithm (2013), lecture notes

[8] David M. Blei, A.N., Jordan, M.: Latent dirichlet allocation. JMLR **3**, 993–1022 (2003)

[9] Domínguez García, R., Schmidt, S., Rensing, C., Steinmetz, R.: Automatic taxonomy extraction in different languages using wikipedia and minimal language-specific information. In: Proceedings of the 13th international conference on Computational Linguistics and Intelligent Text Processing - Part I. pp. 42–53. CICLing'12, Springer-Verlag, Berlin, Heidelberg (2012)

[10] Fellbaum, C.: WordNet: An Electronic Lexical Database. Bradford Books (1998)

[11] Girju, R., Badulescu, A., Moldovan, D.: Learning semantic constraints for the automatic discovery of part-whole relations. In: Proceedings of HLT/NAACL-03. pp. 80–87 (2003)

[12] Gruber, T.R.: Towards principles for the design of ontologies used for knowledge sharing. In: Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer, The Netherlands (1993)

[13] Harris, Z.: Mathematical structures of language. John Wiley Sons (1968)

[14] Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th International Conference on Computational Linguistics. pp. 539–545 (1992)

[15] Heckerman, D.: Learning in graphical models. chap. A Tutorial on Learning with Bayesian Networks, pp. 301–354. MIT Press, Cambridge, MA, USA (1999)

[16] Hindle, D.: Noun classification from predicate-argument structures. In: Proceedings of the 28th annual meeting on Association for Computational Linguistics. pp. 268–275. ACL '90, Association for Computational Linguistics, Stroudsburg, PA, USA (1990)

[17] Hjelm, H., Buitelaar, P.: Multilingual evidence improves clustering-based taxonomy extraction. In: Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008) (2008)

[18] Kotlerman, L., Dagan, I., Szpektor, I., Zhitomirsky-geffet, M.: Directional distributional similarity for lexical inference. Natural Language Engineering **16**(4), 359–389 (oct 2010)

[19] Leenheer, P.D., Moor, A.D.: Context-driven disambiguation in ontology elicitation. In: Context and Ontologies: Theory, Practice, and Applications. Proc. of the 1st Context and Ontologies Workshop, AAAI/IAAI 2005. pp. 17–24. AAAI Press (2005)

[20] Lenci, A., Benotto, G.: Identifying hypernyms in distributional semantic spaces. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics - Proceedings of the Sixth International Workshop on Semantic Evaluation. pp. 75–79. SemEval '12, Association for Computational Linguistics, Stroudsburg, PA, USA (2012)

[21] Mahdisoltani, F., Biega, J., Suchanek, F.: YAGO3: A Knowledge Base from Multilingual Wikipedias. In: 7th Biennial Conference on Innovative Data Systems Research. CIDR 2015 (2015)

[22] Mimno, D., Wallach, H., Naradowsky, J., Smith, D.A., McCallum, A.: Polylingual topic models. In: EMNLP (2009)

[23] Minnen, G., Carroll, J., Pearce, D.: Applied morphological processing of english. Nat. Lang. Eng. **7**(3), 207–223 (sep 2001)

[24] Navigli, R., Ponzetto, S.P.: Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence **193**(0), 217–250 (2012)

[25] Ni, X., Sun, J.T., Hu, J., Chen, Z.: Mining multilingual topics from wikipedia. In: Proceedings of the 18th international conference on World wide web. Association for Computing Machinery, Inc. (April 2009)

[26] Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. Computational Linguistics **29**(1), 19–51 (2003)

[27] Pereira, F., Tishby, N., Lee, L.: Distributional clustering of english words. In: Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics. pp. 183–190 (1993)

[28] Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. J. Artif. Intell. Res. (JAIR) **11**, 95–130 (1999)

[29] Smet, W.D., Moens, M.F.: Cross-language linking of news stories on the web using interlingual topic modelling. In: King, I., Li, J.Z., Xue, G.R., Tang, J. (eds.) CIKM-SWSM. pp. 57–64. ACM (2009)

[30] Vulic, I., Moens, M.F.: Probabilistic models of cross-lingual semantic similarity in context based on latent cross-lingual concepts induced from comparable data. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, October 2529, 2014. pp. 349–362. ACL (Oct 2014)

[31] Wang, W., Barnaghi, P., Bargiela, A.: Probabilistic topic models for learning terminological ontologies. IEEE Transactions on Knowledge and Data Engineering **99**(RapidPosts) (2009)

[32] Weeds, J., Weir, D.: A general framework for distributional similarity. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. pp. 81–88. EMNLP '03, Association for Computational Linguistics, Stroudsburg, PA, USA (2003)

[33] Zhang, C.: Generating a multilingual taxonomy based on multilingual terminology clustering. Chinese Journal of Library and Information Science (CJLIS) **4**(2), 27–40 (2011)

Combining Extreme Multi-label Classification and Principal Label Space Transformation for Cold Start Thread Recommendation

Kantarakorn Jitharn
Big Data Engineering Program
College of Innovative Technology and Engineering
Dhurakij Pundit University
Bangkok, Thailand
605162020009@dpu.ac.th

Eakasit Pacharawongsakda
Big Data Engineering Program
College of Innovative Technology and Engineering
Dhurakij Pundit University
Bangkok, Thailand
eakasit.pac@dpu.ac.th

***Abstract*—The recommendation system has been widely used in various areas, e.g., entertainment, education, and travel. However, this technique faces two main challenges which are Cold Start and High-Dimensionality problems. The cold start happens when the system does not have enough profile of new users; therefore, the system cannot recommend products to them. The second issue comes from the fact that there are a lot of distinct products or users to be recommended. Recently, Extreme Multi-label Classification (XMLC) has been applied to the recommendation system and addressed the Cold Start issue. However, the previous method still has a high-dimensionality issue. In this paper, we proposed a new approach, namely XMLC-PAO, which integrated label space reduction with XMLC. In more details, we transformed the recommendation problem to XMLC and applied Singular Value Decomposition (SVD) to generate reducing operator of label space (products' or users' label space). For the feature space, Deep Learning technique has been used to extract features from texts. From the experiments with Stackoverflow online forums dataset, we have found that the XMLC-PAO showed better performance in terms of RECALL@M and NDCG@M when the dimensions were reduced to 50% and 80% of the original size.**

***Index Terms*—extreme multi-label classification, recommendation system, cold start problem, singular value decomposition - SVD**

I. INTRODUCTION

Recommendation system is a tool for recommending users of what items they are looking for. It has been used in a variety of area such as movies, music, books, and research articles. However, there is a classic problem in recommendation system called "Cold Start" which is about a system that could not recommend new items to any users or could not suggest any items to new users as there is no previous interaction.

Examples of traditional cold start solutions are 1) 'Statistical model-based approach' which is using the corresponding probability distribution statistics user for projecting and initializing rates and high probability items to recommend, 2) 'Average approach' in which the original rating matrix is filled using the average of all ratings of the items before collaborative filtering, and 3) 'Mode approach' that user's predicted result score is calculated from the most often user rating. However, those methods still have an area of improvement in recommendation system [1].

Multi-label classification is a classification of items which have multiple correlated labels. It was initially studied by Schapire and Singer in text categorization [2]. Later many techniques in multi-label classification have been proposed for various applications such as semantic scene classification, music emotion categorization, automated tag recommendation, bioinformatics research and sentiment analysis [3]. Extreme multi-label classification is a multi-label classification that has extremely large set of target label [4] such as online forums which have been created for asking opinions or questions in any community.

The eXtreme Multi-Label Classification Method (XMLC) has been used in predicting a set of users (labels) who will want to respond to any new items in online forums. Stackoverflow online forum dataset could be used as items of the recommendation system. It is in a form of 'thread' starting by posting questions or asking others for opinions on a certain topic. While community members would be enable users to ask questions, the key of system is to ensure that the members find questions relevant to their interest to get them answered. Selecting a subset of users from the set of all users in the community poses significant challenges due to scalability and sparsity [5].

One possible approach to deal with High-Dimensionality issue which leads to scalability and sparsity in extreme multi-label classification is to transform into subset of original label hypercube view which has important correlations between labels before learning. Principal Label Space Transformation (PLST) is a simple and efficient method which relies on only singular value decomposition as the key step [6].

This paper aims to apply combination of eXtreme Multi-Label Classification and Principal lAbel space transfOrmation (XMLC-PAO) with the Stackoverflow online forum dataset. The input thread is processed by using stacked bi-directional

Gated Recurrent Units (Bi-GRU) architecture for text encoding along with cluster sensitive attention (CSA) for exploiting correlation along the large label space [5]. Application of Principal Label Space Transformation in reducing extreme multi-label space can reduce high-dimensionality problem before creating prediction network; however, deep neural network structure needs to be transformed from a classification model to a regression model. The XMLC-PAO models of different reducing dimensional ratio would be evaluated by values of MRR, RECALL@M, and NDGC@M comparing with the original model architecture [5].

II. BACKGROUND

Recommendation system is a system that tries to match user's profile (content-based filtering) or user's social environment and past behavior (collaborative filtering) with some reference characteristics that are related to item characteristics. User-item can be associated with various kinds of interactions such as ratings, bookmarks, purchase frequency, number of 'likes', number of page visits etc.

A. Cold Start Recommendation Problem

Cold start problem is a problem that recommendation systems could not match user's characteristic to any reference for recommending any items. There are three cases of cold start which are:

1. **New community:** It refers to a new recommendation system that has already had a catalog of items but has no users.
2. **New user:** If a new user has just registered, there is no interaction provided to the system.
3. **New item:** If a new item is added, there is no content information in the system.

The common problem of those cases is a lack of user-item interaction which makes it challenges to provide reliable recommendations (Fig.1). However, our framework study would only focus on a new item cold start scenario and create recommendation model to predict potential users to that item.

B. Definition of multi-label classification

Extreme multi-label classification (XMLC) refers to a task of assigning items into their most relevant subset of labels from an extremely large collection of class labels. The fundamental difference between multi-label classification and traditional binary or multi-class classification is that in multi-class classification only one among the possible labels applies to an item, whereas in multi-label classification the labels can be correlated with each other or have a subsuming relationship, and multiple labels can apply for an item (e.g., 'politics' and 'White House' for news articles, 'electron- ics', 'Samsung' and 'smartphone' for products, 'Eiffel tower' and 'vacation 2017' for images) [5].

Let $X = \mathbb{R}^P$ and $Y = \{0,1\}^K$ be an P-dimensional feature space and K-dimensional binary label space, where P is the number of features and K is a number of possible labels, i.e. classes. Let $D = \{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle, ..., \langle x_q, y_q\rangle\}$ is a

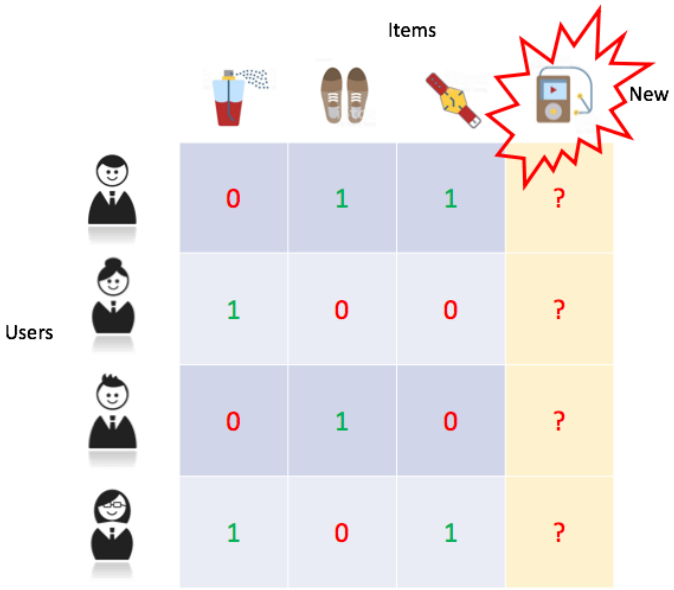


Fig. 1. Interaction Matrix of Cold Start Problem: '1' ⇒ interaction, '0' ⇒ no interaction

set of q objects (e.g. documents, images, etc.) in a training dataset, where $x_i \in X$ is a feature vector that represents an i^{th} object and $y_i \in Y$ is a label vector with the length of $K, [y_{i1}, y_{i2}, ..., y_{iK}]$. Here, y_{ij} indicates whether the i^{th} object belongs (1) or not (0) to the j^{th} class.

In general, 2 main phases are exploited in a multi-label classification problem: (1) model training phase and (2) classification phase. The goal of the model training phase is to build a classification model that can predict the label vector y_t for a new object with the feature vector x_t . This classification model is a mapping function $F : \mathbb{R}^P \rightarrow \{0,1\}^K$ which can predict a target value closest to its actual value in total. The classification phase uses this classification model to assign labels. For convenience, $X_{N\times P} = [x_1, ..., x_N]^T$ (i.e. D) denotes the feature matrix with N rows and P columns and $Y_{N\times K} = [y_1, ..., y_N]^T$ represents the *label matrix* with N rows and K columns, where $[-]^T$ denotes matrix transpose [3].

C. Input and Output of our system

Examples of input (thread) and output (user list) are showing in Table I. For the input part, there is a deep neural network architecture (Fig. 3) in extracting features which starts from word embedding, BiGRU and Cluster Sensitive Attention [5] before feeding into the predicting network. In the output part, user list is needed to reform to interaction matrix which has rows as items or threads and columns as user id. The interaction matrix would have value of 1 for interacting and 0 for not interacting between items and user ids corresponding for rows and columns. It is obviously sparse (Fig.2) which is a reason for trying to decrease dimension. In the next section, the interaction matrix would be called as label Y which would be referred in the experiment.

TABLE I
EXAMPLE OF RAW THREAD (INPUT) AND USER LIST (OUTPUT)

OwerID	Thread	User List
120	I am using CCNET on a...	[12734]
3400	How do you specify that ...	[419, 383, 60, ...]
580	Is it possible to do image...	[149, 34, 116, ...]
4320	Both the jQuery and Proto...	[17, 493, 598, ...]
5170	I have just started working ...	[225, 162, 667, ...]

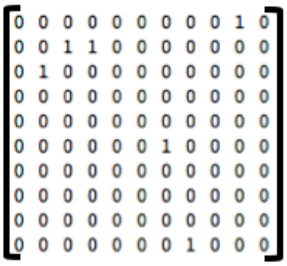


Fig. 2. Example of label Y: The column number is User ID and the row number is thread number

III. PROPOSED METHOD

A proposed method is a combination of Extreme Multi-Label Classification and Principal lAbel space transfOrmation (XMLC-PAO). Deep neural network in this paper would be created by using these following structures.

A. Principal Label Space Transformation

As mentioned in the previous section, the multi-label Y can be correlated with each other and the application of hypercube view can be applied to XMLC. The method of hypercube view used in this article is adjusted from original Principal Label Space Transformation (PLST) method. The Label Y hypercube with K-dimension would be decomposed into 3 parts through Singular Value Decomposition (SVD) for finding a proper projection matrix P and the decoder D for an M-flat label H [6].

This method is adjusted from original PLST method [6] to fit with experimental label Y dimension. In particular, the matrix Y (interaction matrix) is formed with each row being y_n, the occupied rows. Then, SVD would be performed on the $N \times K$ matrix Y to obtain three matrices [7].

$$Y = U\Sigma V^T \tag{1}$$

Here U is a N by N unitary matrix, Σ is a N by K diagonal matrix, and V^T is a K by K unitary matrix. Through SVD, each row y_n can be represented as a linear combination of singular vectors v_m^T in V^T. The vectors form a basis of a flat that passes through all the y_n. The matrix Σ is a diagonal matrix containing singular values σ_m that corresponds to the singular vectors v_m^T. It could be assumed that the singular values are ordered such that $\sigma_1 >= \sigma_2 >= ... >= \sigma_K$.

The equation (1) can be rewritten as

$$YV = U\Sigma \tag{2}$$

where the orthogonal basis V can be seen as a projection matrix of Y that maps each y to a different coordinate system. Since the largest M singular values correspond to the Principal directions of the original label space, the rest of the singular values could be discarded and their associated basis vectors in V obtained a smaller projection matrix

$$P = V_M = \left[\begin{array}{cccc} v_1 & v_2 & ... & v_M \end{array} \right] \tag{3}$$

that maps the vertices y to the M-flat label H which can be calculated as

$$H = Y.V_M \tag{4}$$

The transformed label H would be replaced label Y in the model structure which would change classifier model to regression model as value in H is not [0, 1].

An efficient decoder D for PLST can be calculated because $P = V$ is an orthogonal matrix, $P^{-1} = P^T$. This means V_M^T can be used to map any vector h on the M-flat space back to a point $h.V_M^T$ in label Y space.

TABLE II

Principal Label Space Transformation
1. With a parameter M, perform SVD on Y and obtain
$V_M = \left[\begin{array}{cccc} v_1 & v_2 & ... & v_M \end{array} \right]$
2. Using $P = V_M$ for transform Y to H by $H = Y.V_M$
3. Using $D = V_M^T$ for transform H back to Y by $Y = H.V_M^T$

B. Feature Engineering

As each input of system is a post text, the feature engineering sections of deep neural network is needed. Main sections for extracting features are Text Encoding and Cluster Sensitive Attention.

1) Text Encoding: Each input post is consists sequence words $(w_1, w_2, ..., w_n)$. The first step of feature engineering is to embed each post into a lower-dimensional space which would be represented as a sequence of word vectors $\{q_1, q_2, ..., q_n\}$ where $q_i \in \mathbb{R}^d$. The word vectors use pre-trained GloVe embedding [8].

The post is then encoded using Bi-directional Gated Recurrent Unit (Bi-GRU) [9]. Input of this part is a sequence of word vector $\{q_1, q_2, ..., q_n\}$ and output is a sequence of $\{p_1, p_2, ..., p_n\}$ where $p_i \in \mathbb{R}^g$. A Bi-GRU reads the sequence of word vectors q_i from left to right in the forward stage to create p_i^f. The backward stage read p_i^f from forward stage in reverse order to create p_i^b. The result from forward and backward states are concatenated to create the encoded hidden state of a post $p_i = [p_i^f; p_i^b]$ considering all its surrounding

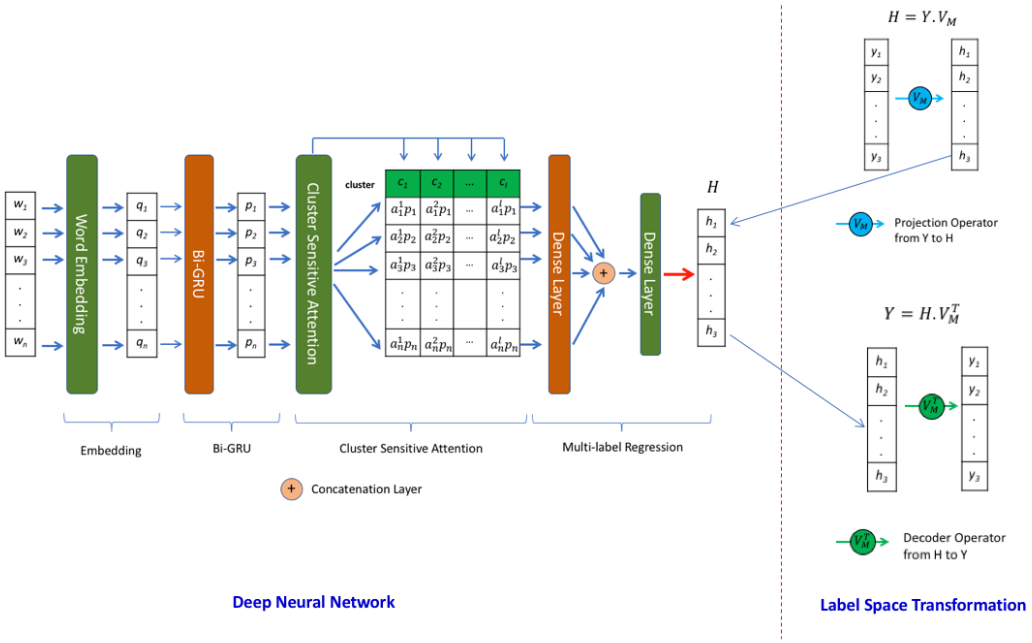


Fig. 3. Overall model architecture of XMLC-PAO

words.

2) Cluster Sensitive Attention: Cluster Sensitive Attention is a component in the network that can help focus on parts of post for different users [5]. To achieve this, it needs an attention mechanism that can give different weights to words of the posts and generate an encoded text representation using the weighted words, thus focusing on important parts. For each p_i from Bi-GRU component, a weight a_i for its corresponding word sequence w_i is computed for generating and attention vector $a = \{a_1, a_2, ..., a_n\}$ as:

$$a_i = \frac{exp(e_i)}{\sum_{j=1}^{n} exp(e_j)}, where \tag{5}$$

$$e_i = tanh(W_i p_i + b_i) \tag{6}$$

where W_i is a weight matrix of $1 \times g$ and b_i is the bias term. The text representation for a single attention layer (c_j) is then computed as:

$$c_j = \sum_{i}^{n} a_i^j p_i \tag{7}$$

The attention weights (a_i^j) should be dependent on different users' interests which can be softly clustered in a finite number of clusters and called it as "Cluster Sensitive Mechanism". From the text representation p_i, it can generate l different attention weight vectors $a_i^1, a_i^2, ..., a_i^l$. Thereafter, by using the different attention weights on p_i, a cluster sensitive encoding of the post text is $C = [c_1, c_2, ..., c_l]$.

C. Extreme Multi-Label Regression

For a post text, the l encoded texts are concatenated and fed through a fully connected layer with U output neurons (corresponding to each user). The fully connect layer learns the weights for its l inputs (corresponding to the different encoded texts).

$$z = tanh(W.C + b) \tag{8}$$

where W and b are weight and bias matrices respectively and $tanh$ is an element-wise non-linear activation function. The output of this feed-forward layer $z \in \mathbb{R}^U$ is then passed through output layer which having H label as output without any activation function to make it as regression model. Our model is trained using mean square error as the loss function. The network is end-to-end trainable and is optimized with Adam optimizer.

IV. EXPERIMENT

A. Dataset

Stackoverflow: is a dataset which has been used in this paper. It is a CQA website for programming related questions. It has been dumped from kaggle. All questions are posted during 2008 to 2010. The dataset would be used by removing all the code snippets (encapsulated within the tags $\langle code \rangle \langle /code \rangle$) from the question texts as showing in "Threads" column of Table I.

B. Evaluation

Before evaluation, the prediction from trained model need to be transformed from predicted H_{pred} to label Y space by applying decoder V_M^T as mentioned in proposed method

section. The evaluation would be the comparison on Y and Y_{pred}.

$$Y_{pred} = H_{pred}.V_M^T \tag{9}$$

The label Y is huge and very high sparsity ratio (Fig.2). Therefore, it is a reason for not using overall accuracy as evaluation metric. It would be better to evaluate the positive instances i.e. the users who actually participated in a thread.

There are 3 following metrics considered as the better choices to evaluate the recommendation quality of the competing methods.

- Mean Reciprocal Rank (MRR) : indicates the position of the first relevant user in the ranked list.This measures the ability of a system in identifying an interested user at the top of the ranking. Let rt be the rank of the highest ranking relevant user for a test thread t. MRR is just the reciprocal rank, averaged over all threads in test set, n:

$$MRR = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{r_t}$$

- Recall@M : considers how many top-M users actually interacted with the thread (the higher is better). Recall for the entire system is computed as the average recall value for all threads in test data.

- Normalized Discounted Cumulative Gain (NDCG@M) : is well suited for evaluation of recommendation system, as it rewards relevant results ranked higher in the returned list more heavily than those ranked lower. NDCG@M for a thread t is computed as:

$$NDCG_t = Z_t \sum_{j=1}^{M}\frac{2^{r(j)}-1}{log(1+j)}$$

  where Z_i is a normalized constant so that a perfect ordering would obtain NDCG of 1; each $r(j)$ is an integer relevance level (for used case, $r(j) = 1$ and $r(j) = 0$ for relevant and irrelevant recommendations, respectively) of result returned at the rank $j \in \{1, ..., k\}$. Then, for each M value, $NDCG_t$ is averaged over all (n) threads in the test set to get the overall NDCG@M.

The evaluation at $M = [5, 10, 30, 50, 100]$ are used to determine the quality of recommendation at different thresholds of the ranked list.

The Y_{pred} is a predicted interaction matrix of recommendation system (Fig. 2). The evaluation methods would rank columns of each row form each element value descending (the higher value is higher chance to recommend that user). The evaluation values would use ranked column index in the calculation as previously mentioned.

V. Results

The first 5,000 threads of data have been selected for training (4,000) and testing (1,000). By applying XMLC-PAO for reducing label Y dimension with ratio of [1.0, 0.8, 0.5] from original K-dimension (6,651 users), the results are verified with 1,000 testing data and compared with original model [5] (Table III). As mentioned before, that accuracy is not used in evaluating result in the experiment. Performance of predicting is based on value of MRR, RECALL@M, NDCG@M and training time. However, the training time and MRR for result of XMLC-PAO ratio 0.5 and 0.8 are not obviously better than ratio 1.0 and result without XMLC-PAO; the RECALL@M and NDCG@M of M = [30, 50, 100] for XMLC-PAO ratio 0.8 are much better than the original result.

All evaluation values have been adjusted to vary between 0 (min) and 100 (max) which means they are quite small. It is a nature for this type of data as it is difficult in processing and there are still some noises which could not be removed. There are not so many users that could answer for any kind of questions. The higher RECALL@M means the more numbers of correct prediction results comparing to true recommendation and the higher NDCG@M means the more correct prediction rank matching with true result. It is a good enough indicator for demonstrating better result from this experiment.

By adjusting Adam learning rate for deep neural network model of XMLC-PAO ratio 0.8 label (Table IV), all evaluation values are varied with different learning rate. The best result is a model trained by Adam learning of 0.008. Almost all RECALL@M and NDGC@M are better than others and are also better than that from original model.

VI. Conclusions

Extreme multi-label (Y) of online Stackoverflow forums dataset has been transformed to subset of hypercube of original label (H) and used it for generating deep neural network. The experimental results of using different XMLC-PAO ratio label are compared with model using full original label and verified that the combination of Extreme multi-label and Principal Label Space Transform is successful in reducing the computational effort and achieves better performance.

The model still can be turned to give better prediction such as adjusting Adam learning rate. There are many points inside deep neural network architecture which can be changed and might give better results.

This method can be applied to other kinds of data set such as product description, movie synopsis, or book content for solving in cold start recommendation.

Acknowledgment

Thanks to Kishaloy Halder for sharing his source used in 'Cold Start Thread Recommendation as Extreme Multi-label Classification' research which can make this experiment started easier.

TABLE III
COMPARISON RESULTS OF TESTING DATA (1,000 THREADS)

Activation function	**sigmoid**	**regression**	**regression**	**regression**
Loss function	**binary crossentropy**	**mse**	**mse**	**mse**
XMLC-PAO Ratio	**XMLC org**	**1.0**	**0.8**	**0.5**
Training time	00:26:00	00:26:28	00:25:58	00:25:39
MRR	0.0302	0.0076	**0.0307**	0.0282
		RECALL@M		
RECALL@5	1.1006	0.0704	1.0239	**1.1780**
RECALL@10	**1.6109**	0.0704	1.4403	1.4308
RECALL@30	1.9900	2.1058	**2.9463**	2.0056
RECALL@50	3.1718	3.4573	**4.2110**	2.8578
RECALL@100	4.4196	5.1964	**5.9378**	4.8524
		NDCG@M		
NDCG@5	**1.1802**	0.0719	1.1274	1.1683
NDCG@10	**1.4150**	0.0719	1.3316	1.3077
NDCG@30	1.5474	0.7337	**1.8050**	1.4868
NDCG@50	1.8527	1.10317	**2.1325**	1.7113
NDCG@100	2.1727	1.5379	**2.5371**	2.1812

TABLE IV
COMPARISON RESULTS OF TESTING DATA (1,000 THREADS) WITH VARY ADAM LEARNING RATE IN COMPILING MODEL

Activation function	**regression**	**regression**	**regression**	**regression**	**regression**	**regression**
Loss function	**mse**	**mse**	**mse**	**mse**	**mse**	**mse**
XMLC-PAO Ratio	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**
Adam Learing Rate	**0.0015**	**0.003**	**0.006**	**0.008**	**0.01**	**0.02**
Training time	0:18:02	0:17:55	0:17:54	0:17:47	0:17:44	0:17:38
MRR	0.0143	0.0257	0.0261	**0.0311**	0.0234	0.0216
			RECALL@M			
RECALL@5	0.4167	0.8547	0.8571	**1.1409**	0.8663	0.9365
RECALL@10	0.72016	0.9947	1.0847	**1.3648**	1.2813	1.1351
RECALL@30	0.9234	2.1380	1.5775	**2.5082**	1.6429	1.6362
RECALL@50	1.8689	2.9016	2.1940	**3.1395**	2.6179	2.4456
RECALL@100	**5.2004**	4.6890	3.6647	4.5454	4.0976	3.8060
			NDCG@M			
NDCG@5	0.4572	0.9799	0.9983	**1.2160**	0.8870	0.8735
NDCG@10	0.5833	1.04168	1.1004	**1.3309**	1.0661	0.9865
NDCG@30	0.6480	1.4149	1.2585	**1.7478**	1.2065	1.1587
NDCG@50	0.9052	1.6173	1.4101	**1.9190**	1.4932	1.4049
NDCG@100	1.6629	2.0450	1.7521	**2.2448**	1.8636	1.7273

REFERENCES

[1] M. Chen, C. Yang, J. Chen, and P. Yi, "A Method to Solve Cold-Start Problem in Recommendation System based on Social Network Sub-community and Ontology Decision Model," Atlantis Press, pp. 159–166. 2013.

[2] R. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization," Mach. Learn, Vol.39, pp. 135-168. 2000

[3] E. Pacharawongsakda and T.Theeramunkong, "Multi-Label Classification Using Dependent and Independent Dual Space Reduction," The Computer Journal, Vol. 56 No. 9, pp. 1113–1135. 2013

[4] R. Babbar and B. Schölkopf, "Adversarial Extreme Multi-label Classification," stat.ML, 2018

[5] K. Halder, L. Poddar, and M.Y. Kan, "Cold Start Thread Recommendation as Extreme Multi-label Classification," Extreme Multilabel Classification for Social Media, pp. 1911–1918. 2018.

[6] F. Tai and H. Lin, "Multi-label Classification with Principal Label Space Transformation," 2nd International Workshop on Learinng from Multi-Label data, Haifa, Israel, 2010.

[7] B. Datta, "Numerical Linear Algebra and Applications," Brooks/Cole Publishing, 1995.

[8] J. Pennington, R. Socher, and C. D Manning, "Glove: Global vectors for word representation," In Proc. of EMNLP, 2014.

[9] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," Proc. of NIPS Deep Learning and Representation Learning Workshop, 2014

Analysis of Detecting and Interpreting Warning Signs for Distance of Cars using Analyzing the License Plate

Patiyuth Pramkeaw
Media Technology Program
King Mongkut's University of Technology Thonburi, Thailand
patiyuth.pra@kmutt.ac.th

Mahasak Ketcham
Department of Information Technology Management, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand
mahasak.k@it.kmutnb.ac.th

Warissara Limpornchitwilai
Media Technology Program
King Mongkut's University of Technology Thonburi, Thailand
piannerypianopin@gmail.com

Narumol Chumuang
Department of Digital Media Technology, Faculty of Industrial Technology,
Muban Chombueng Rajabhat University Ratchaburi, Thailand
Lecho20@hotmail.com

***Abstract*— The purpose of this research was to The Develop a System for Detecting and Interpreting Warning Signs, The researcher has proposed to develop a system for detecting and interpreting warning signs by using the theory of color from the sign of the traffic sign. This can be used in analyze and detect traffic signs through the design of applications on smartphones in order to be able to create the traffic sign support system by detecting and interpreting traffic signs. In this case, the driver can clearly and correctly understand the meaning of traffic signs. As a result, the driver can follow the traffic rules. Moreover, it helps reducing accidents on the road as well. The efficiency of the system was tested with test result, it is found that the program was able to detect traffic signs and interpret the objects within the image as traffic signs accurately. Additionally, it is found that the light and distance of the camera affected the quality of data processing in this test to measure performance. In addition, the accuracy of the system is totally 93.5.**

Keywords— Warning Signs; Traffic sign; Detection

I. INTRODUCTION

Road accidents cause both losses for life and property of Thai people. In each year, there are more than 13,000 road deaths, excluding those injured people that the numbers cannot be estimated officially. When considering the value of economic loss, it is estimated that road accidents cause a loss of 232,855 million baht per year (Department of Highways (a), 2007), mainly due to driving behaviors of 69.03% of total causes. Accidents caused by vehicle condition is calculated as 1.11 percent, and the environment (animal surpassed) only 0.48 percent for the cause of road accidents from the statistics in 2008.

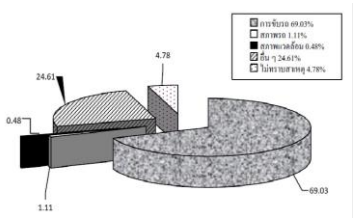


Figure 1. Causes of road accidents nationwide in 2008

Road accidents across the country caused by the top 5 driving [1] include:

No. 1 driving speed exceeds the legal rate of 23.41 percent

No. 2 driving passes the other cars in the close distance 21.07 percent

No. 3 driving behind another car in the close distance, 11.69 percent

No. 4 driving surpassing illegally for 8.73 percent

No. 5 violating of traffic light / traffic signs for 6.12 percent

Most of the problems of road traffic accidents occurred from driving behavior. At night time, it is risk of accident as well. One of the causes is the reduction of visual acuity, so the distance from the front car will be too less or the action that driving close to another car can cause accident.

From the causes of these problems, there is an idea to develop a system to help reduce accident. Most of the problems come from driving behind other cars in close range. This is the top cause of the problem in the accident. There is research related to the calculation of driving simulators based on Californian safety driving criteria, The United States of America, that is, the safe driving distance should be at least about one car length at 10 mph. For example, if driving at 40 mph keep driving distance at 4 car lengths, from this rules, the minimum time between two cars when running through the

Minimum Time Headway can be calculated as a function of vehicle speed or approximately 1.36 seconds. When compared to field observations, it is found that the minimum headway obtained from the Pipes model is slightly less than the field value at speeds of 15-35 mph, but is different from the value obtained from the airport in the low speed and high speed [2]. There is a research studied on the Reaction time that it has a direct effect on driving behavior. Therefore, the time between the rear bumper of the car and the rear of the car behind the car should be no less than the response time. It means that the minimum time headway is equal to the response time, plus the time for the preceding car to travel at a distance equal to the length of the car. Forbes has conducted several field experiments to find that the minimum time gap is different for each experiment. In summary, Forbes' driving model provides safe driving distances similar to those of Pipes [3].

The distance between the cars must be determined by the distance between the first vehicle and the second vehicle driving on the road. In order to find the distance between the cars can be detected from the registration plate of vehicle, and such plate is important to the government to force all cars to have a license plate to show the identity of the car owner. The size and color of the car is clearly defined according to the Motor Vehicle Act (No. 10) BE 2542. Therefore, the license plate is suitable for calculating the distance. Research on the Analysis and Identification System of Recognized License Plates can be divided into two main sections: finding the license plate, and recognizing the characters inside the license plate. The finding of a license plate is an important step because it is the first step in the recognition system to detect the plate position. If this process is accurate and quick for the recognition system, it can make the accuracy and speed of the overall system more effective.

Research on car license plate location can be divided into two major groups which are the group using the color image of the plate, and the group that finding location of the car plate by using gray image. The research in car license plate location using color image will be used to determine the color of the car plate, which will be different from the color in other areas of the car. The license plates will have the texture and appearance of the background color scheme within the same plate. After that, it is processed to check for the image frame and the background color of the license plate to verify the location of the license plate [4]-[7]. However, the problem of the group finding the license plates using color images is the difference in light intensity, the color of the license plates changed, or any other area of the car with the same color as the license plate will result in erroneous positioning [3]. It also takes a long time to process because it requires three color images (red, green, blue) [8]. In the second group, the finding of the license plates is conducted by using gray scale images from the research [9] - [10]. Images obtained from the camera will be changed to gray levels. After that, the image is changed to binary image to get rid of the background noise contract. Then, the image is pulled vertically. The median in the image filtering plate is projected horizontally and vertically to obtain the location of the license plate. The research on the vehicle license plate reading system from the distance from which the camera can detect the license plate is in good criteria at about 0.5-1.25 m [11]. The speed of the car is also important as well because if the car uses high speed traffic, the camera will not be able to take photos immediately or the image quality is not good. As a result of the study, it is found that reading the license plate is effective when the vehicle is running at low speed [12], and the environment while shooting is also important, as the quality of the photos that can be taken is depending on the environment, darkness, and brightness while capturing the image. It can be seen that nighttime images are blurry, resulting in poor system performance [13, 14].

The researcher has proposed to develop a system for detecting and interpreting warning signs by using the theory of color from the sign of the traffic sign. This can be used in analyze and detect traffic signs through the design of applications on smartphones in order to be able to create the traffic sign support system by detecting and interpreting traffic signs. In this case, the driver can clearly and correctly understand the meaning of traffic signs. As a result, the driver can follow the traffic rules. Moreover, it helps reducing accidents on the road as well.

II. BACKGROUND AND NOTATION

The way to detect and interpret traffic signs is based on a similar research samples, including the pros and cons of each method. The results of the study found that the system will be used to develop the system, and using the related theories in the development of traffic sign interpretation systems, digital image processing, and converting analog video data into digital signals in order to process the results through the computer system. The use of Region-of-interest (ROI) is a theory that is interesting. It may be in any positions within the image by encircling the area of interest, and the Haar-like Features Theory is introduced. It is a way to detect and interpret objects within images by using the principle of Haar Wavelet.

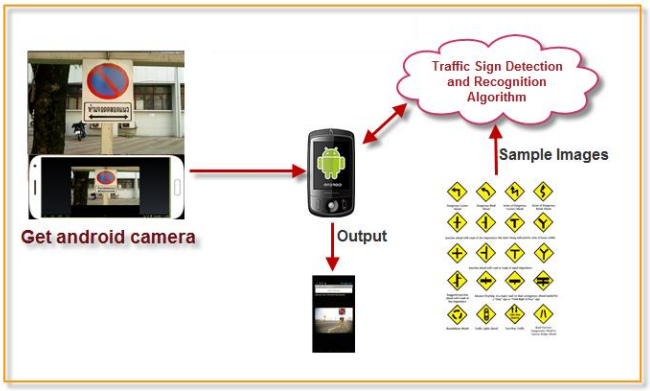


Figure 2. Shows the overall workflow of the system

The Detecting and Interpreting Warning Signs process is divided into 6 steps as follows:

Step 1: Image conversion

When the camera is receiving a picture on the device, it enters the digital image conversion process and sends it to the processor, which is a RGB image.

Step 2: Traffic Signs Placement

The Haar like-Features of Viola and Jones are used to locate traffic signs.

Step 3: Comparison of data

Send data to compare the format of the pixel position to the processing order to calculate the sum at the gray (X, Y) pixel position within each pixel to indicate the position that indicates the traffic sign. AdaBoost (Adaptive Boost), a process that looks similar. And different with imported images for the classification of images.

Step 4: Output

The system will display a detachable traffic sign frame and display a meaningful text of the traffic sign, along with a warning tone as the type of sign.

III. THE PROPOSED ALGORITHM

Operation of Traffic Signal Detection and Interpretation System. The processes of detecting and interpreting traffic signs are as follows.

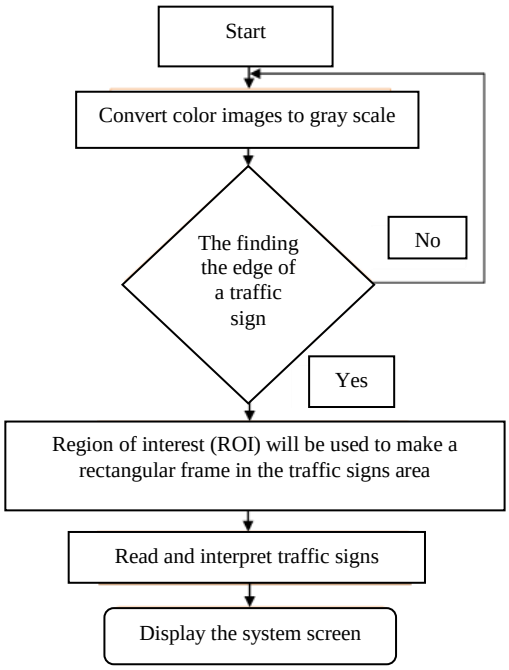


Figure 3. Shows the work of Flow Chart of traffic detection and interpretation system

A. Receive input from a smartphone's camera

When the system starts up, it is captured from the camera on a smartphone. In the detection, when receiving data from a camcorder on a supplied smartphone, it is necessary to separate the video into frames for each frame to detect the desired traffic signal. It depends on the frame rate used to record, and here are 10 frames per second (fps). In 1 second, 10 images can be extracted, and the image size is 640x480 pixels according to the size of the camera capture on the device.

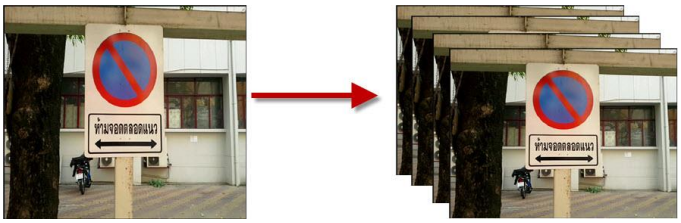


Figure 4. Shows the images splitting from a camera on a smartphone device

B. The process to convert color images to gray scale

Haar like-Features according to the method of Viola and Jones are methods of detecting and interpreting objects within the image by creating a feature that shows the difference between the white area and the black area which can change the size and the position. It is used for detecting the various types of images such as straight lines, circles, etc.

It converted the received image by using an adjustable breakpoints technique to convert RGB images to Gray Scale and convert them from the intense value of white and black from the appropriate image into white-black image in Binary type.

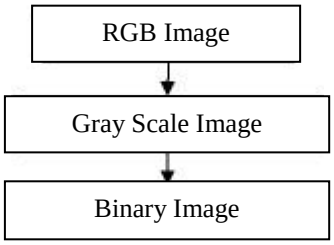


Figure 5. Shows the conversion steps of the adjustable breakpoints.

Convert the RGB image to Gray-scale image from the following equation:

$$Gray = 0.299 \times R + 0.587 \times G + 0.114 \times B$$

Whereas

Gray = Intense value of gray color with the value from 0-255

R = Intense value of red color with the value from 0-255

G = Intense value of green color with the value from 0-255

B = Intense value of blue color with the value from 0-255

C. The finding the edge of a traffic sign by using the Detecting Haar Classifier

AdaBoost (Adaptive Boost), which is a process to find similar and different features from the imported images for the image classification has the following processes.

- First, weight is determined for the Feature that runs within the preview.
- Find areas that consisting of the needed points.
- Increase weight for the remaining parts only the needed features that haven't been divided.
- Repeat this process until the last one brings the whole area together. Then, the needed object and features in parts within the object will be obtained

Cascade Classifiers are interpretations of images according to the features of the image by cutting the negative sub-window, and the positive will be circulated within the image by changing the way of detecting until it can specify what the image is.

By calculating the sum at the pixel (X, Y) of the gray area within each square of the pixel to determine the location of the traffic signs as it is in the fig. 6.

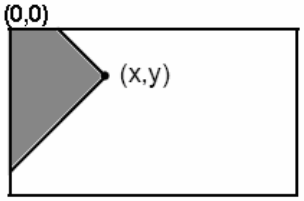


Figure 6. Example of calculating the sum of (x, y)

$$AR[x,y] = \sum_{x' \le x, x' \le x - |y - y'|} A(x', y') \qquad (1)$$

Whereas

A[x,y] = The original image

AR[x,y] = Image with only interesting features

The obtained XML file from the training process will be used to detect traffic signs within the imported image using the sub window mask in order to look for the same characteristics pursuant to the principle of the Haar Classifier.

Since the traffic signs are from 5 pixels to 75 pixels, the minimum size of the window mask is 5x5 pixels and the size of the window mask is 5%. Once it is detected, the location, size, and the type of the detected image will be recognized.

D. Detecting Eage by ROI

Region of interest (ROI) will be used to make a rectangular frame in the traffic signs area, covering only the areas of interest.

Figure 7. Shows the sample of location detecting by using Region of interest (ROI)

E. Read and interpret traffic signs

Haar like Feature technique is a way to detect and interpret objects within images based on Haar Wavelet's principle to bring the traffic signs with square frame to be processed by using the function named Detector.java. It is a Class used for detecting traffic signs. The values obtained from the images in each frame are compared to the values obtained from the training by using the Haar like Feature, and the obtained value will be used to display the results with the interpretation of traffic signs, and then use the function named itemadaptor.java, which is a class used to display the image of a traffic sign that the system detects and displays the meaning of the sign and the audio to read the name of the traffic sign and display on the screen.

Figure 8. Shows the detection and interpretation of the objects within the image

IV. METHODOLOGY

The process of calculating the distance of cars by analyzing the license plate's image.

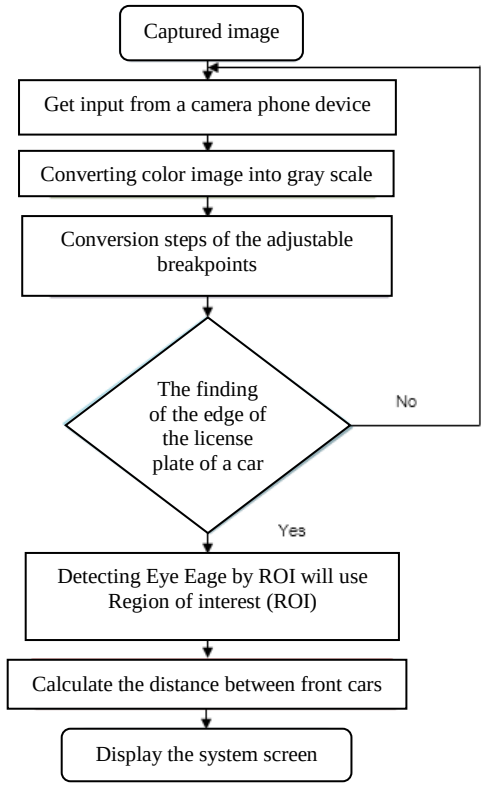


Figure 9. The Flow Chart of the distance of cars' calculation system by analyzing the plate image.

The process of calculating the distance of the vehicle by analyzing the plate image is as follows.

Step 1: Get input from a camera phone device. When the system starts up, it is captured from the camera on a smartphone.
Step 2: The processes of converting color image into gray scale image by using adjustable breakpoints technique to convert RGB images to Gray Scale and convert them from the intense value of white and black from the appropriate image into white-black image in Binary type.
Step 3: The finding of the edge of the license plate of a car by using Detecting eye by Haar Classifier can be calculated from the sum at the pixel (X, Y) of the gray area within each square of the pixel to determine the location of the traffic signs.

It can be calculated by using the following equation:

$$AR[x,y] = \sum_{x' \le x, x' \le x - |y - y'|} A(x', y') \quad (2)$$

Whereas

A[x,y] = The original image
AR[x,y] = Image with only interesting features

Step 4: Detecting Eye Eage by ROI will use Region of interest (ROI) to make square frame at the rear of a car, covering over only the interesting parts.

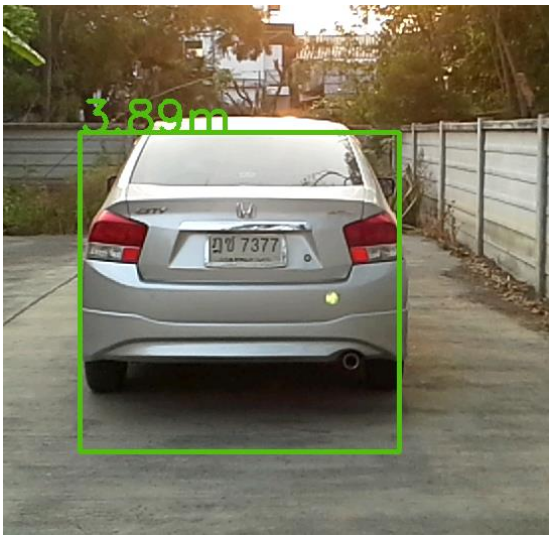


Figure 10. Shows the determination of location by using Region of interest (ROI) Technique

Step 5: Calculate the distance between front cars and the display by using Haar like Feature technique is the way to object within an image. By having the Haar Wavelet principle, the rear of car with square frame shall be taken for the processing by using the function named SensorActivity.java, which is a class used to process images, with a method to draw a frame of interest in an image, and then the obtained value shall be calculated to fine the distance of the object using the formula M (x) = a * exp (b * x) + c * exp (d * x) and display on the device screen.

Display Range: When the distance between cars is obtained in the number of distance between cars, the obtained value shall be an estimated value only, so the comparison over time will be presented in a strip of color with different distance values.

TABLE 1. Table shows distance values

Color bar	Distance
Red	Very close to 0-3 meters
Green	The distance is more than 3-25 meters

Figure 11. Shows the detection and calculation of distance between the front cars within the image

V. Experimental Results and Discusssion

A. *Test the system performance and system performance evaluation*

The system developer has performed a total of 100 system tests to determine accuracy, and the details of the test are as follows:

Step 1: Test to frame the rear part of the car: Test the camera image on a smartphone, and it can frame the interesting area.

Step 2: Test of the distance between the front cars: The test displays the system result that can display the distance between the car by displaying the light frame presenting the distance of the car and presenting numbers to tell distance between the front cars.

B. *System performance test results*

In this study, the developer has performed a total of 100 system tests, and the details of the test are as follows:

Test to frame the Traffic Signs: Test the framing at the traffic signs, covering only the interesting part of the traffic signs in different characteristics as in the following figures:

Figure 12. Shows the square framing on traffic signs and selected areas of interest

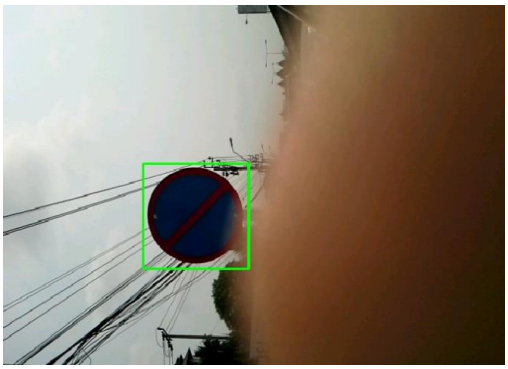


Figure 13. Shows the circular framing on traffic signs and selected areas of interest

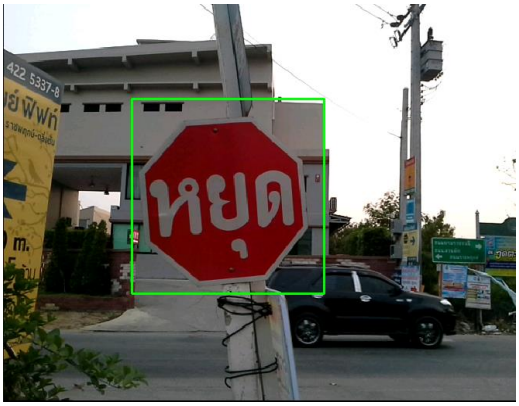


Figure 14. Shows the octagon framing on traffic signs and selected areas of interest

C. Interpretation of Traffic Signs: Displaying the Traffic Signs and meaning of the traffic sign detector as in the following figures:

Figure 15. Example shows the detection and interpretation of the object within the traffic sign warning that there is a construction work ahead.

Figure 16. Example shows the detection and interpretation of the object within the traffic sign warning that there are construction workers working ahead.

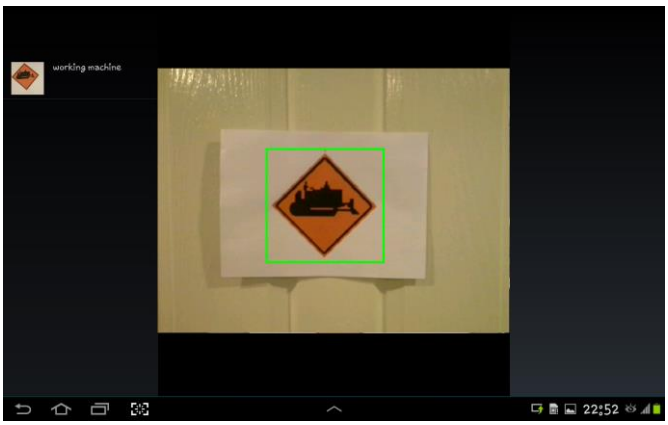


Figure 17. Example shows the detection and interpretation of the object within the traffic sign warning that the machine is working ahead.

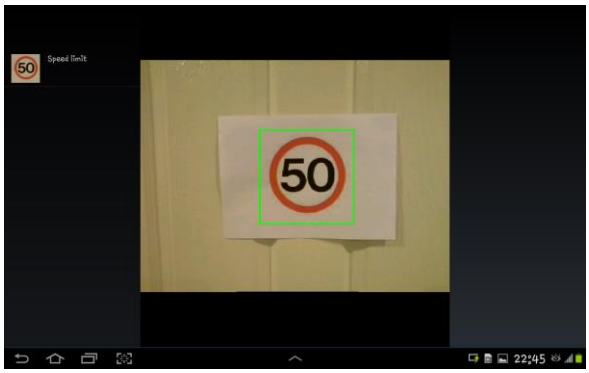


Figure 18. Example shows the detection and interpretation of the object within the traffic sign warning about the limited speed.

TABLE 2. Test results to measure system performance and accuracy

The details of the test	Amount of test	Result	
		Detected	Not detect
1. The system detected traffic signs on the real road.	100	90	10
2. Interpret traffic signs and beeps.	100	97	3
Total	**100**	**93.5**	**6.5**

From the table 2, the summary of the results of the system performance and accuracy measurements from the test by performing the test of system by using method of analyzing the accuracy of detecting and interpreting the objects within the images of the traffic signs according to the Haar like-Features total of 100 times. The traffic signs were 93.5 percent accurate from the 100 test. The system detected 90 traffic signs on the real road and could not detect traffic signs on the real road for 10 times. Moreover, the system can interpret traffic signs and beeps 97 times and cannot interpret the traffic signs and beeps 3 times. Hence, the system developed is very effective and work well in the very good level.

In this study, the developer has performed a total of 100 system tests, and the details of the test are as follows:

Testing the license plate: Test the framing at the rear part of the car, covering only the interesting rear part of the car as in the following figures.

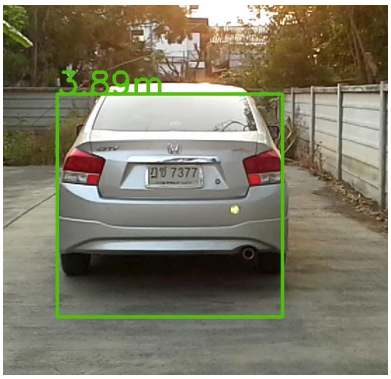


Figure 19. Shows the result of selecting automotive license plate and pick the interesting rear part of the car.

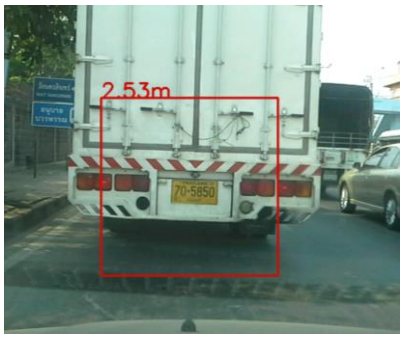


Figure 20. Shows the result of selecting automotive license plate and framing the interesting rear part of the car on the real road.

Test the distance between the front cars: Test the display of the system that can display the distance between cars by presenting the distance between cars in the color frame and presenting numbers of distance of the front car.

Figure 21. Example of measuring the distance between cars by presenting distance of cars in color frame.

Figure 22. Example of measuring the distance between cars by presenting distance of cars in color frame on the real road.

TABLE 3. Table of the test result to measure system performance and accuracy

The details of the test	Valid amount
1. Test of parking vehicles	96
2. Test while the cars were on the real road	94

From Table 3, the results of the performance measurement and system accuracy test is summarized by testing a total of 100 system tests, divided into the detection while parking and while on the real road for 50 times each, and the test results are as follows:

From the test of parking vehicles for 50 times, the system can display the result as a color frame and display the car's distance for 48 times, and it cannot detect the display as a color frame and show the distance of the car for 2 times

From the test while the cars were on the real road for 50 times, the system can display the result as a color frame and display the car's distance for 47 times, and it cannot detect the display as a color frame and show the distance of the car for 3 times

From the 100-time test, it can present the result of display by presenting in a color frame and presenting a distance of a car in two situations, and the test results are calculated as a percentage of 70% from the test results of both situations. Therefore, the developed system has the efficiency to work at a satisfactory level.

VI. CONCLUSIONS

The researcher focused on developing programs to assist older people to reduce the risk of road accidents, and the signs are important to guide and warn drivers to follow traffic rules as well. If the traffic signs are not clear or the driver cannot interpret correctly, it may be caused by traffic signs that are unclear or the driver neglects and does not pay attention to the displayed traffic signs, and it may cause an accident. From the recent research, the use of computer technology to detect or observe the traffic signs will result in increased safety in driving. The performance of the computer system and the design of mobile applications have a potential system for using with the acceptable work, and the performance can be compared with the computer system. The researcher is thinking to develop a system to detect and interpret traffic sign by using the theory of color from traffic signs to analyze traffic signal detection through the design of mobile applications, and the image processing such as Haar-like feature, the Region of Interest (ROI) are used to help calculating the position of the traffic signs, detect, and interpret objects within images by using OpenCV Library and JAVA with the function named Detector.java which is a class used to detect traffic signs. The values obtained from the images in each frame are taken to fine Pixel with the similar value to the values obtained from the training by using the Haar like Feature. Then, the obtained value shall be displayed and the traffic signs shall be interpreted by using the function named itemadaptor.java, which is a class to display the traffic signs at the detecting system with the meaning of signs and audio sound that reads the name of traffic signs on the displayed screen. It is developed into the form of prototype application to have a test as a basis. From the test result, it is found that the program was able to detect traffic signs and interpret the objects within the image as traffic signs accurately. Additionally, it is found that the light and distance of the camera affected the quality of data processing in this test to measure performance. In addition, the accuracy of the system is totally 93.5 percent, which is counted as very good.

REFERENCES

[1] Pipes, L.A. An Operational Analysis of Traffic Dynamics. Journal of Applied Physics, 24, 1953.

[2] Forbes, T.W. Human Factor Considerations in Traffic Flow Theory. Highway Research Record 15, 1963.

[3] Faisal XIAO JIANG; YONG-DONG HUANG, GEN-QIANG LI FUSION OF TEXTURE FEATURES AND COLOR INFORMATION OFTHE LICENSE PLATE LOCATION ALGORITHM School of Information and Computation Science, Beifang Univesity of Nationalities, Yinchuan 750021, China, 2012.

[4] Mohammad Ghazal, Hassan Hajjdiab License Plate Automatic Detection and RecognitionUsing Level Sets and Neural Networks. Abu Dhabi University Abu Dhabi, UAE.

[5] Abdollah Amirkhani Shahraki, Amir Ebrahimi Ghahnavieh, Seyyed Abdollah Mirmahdavi License Plate Extraction From Still Images. Dept. of Electrical Engineering, Iran Univ. of Science and Technology, Narmak, Tehran, Iran, 2013.

[6] E. R. Lee, P. K. Kim and H. J. Kim Automatic Recognition of A Car License Plate using Color Image Processing. Proc. Of International Conference On Image Processing (ICIP'94), vol. 2, pp. 301-305, 1994.

[7] Hsin-Fu Chen, Chang-Yun Chiang, Shih-Jui Yang and Chian C. Android-Based Patrol Robot Featuring Automatic License Plate Recognition. Embedded SoC Lab, Department of Electrical Engineering National Yunlin University of Science and Technology Douliou, Yunlin 64002, Taiwan, ROC

[8] Honghai Liu, Xianghua Hou The Precise Location Algorithm of License Plate Based on Gray Image. College of Information and Engineering Huzhou Teachers College Huzhou, Zhejiang, China, 2012

[9] R. Belaroussi, P. Fouchery, et al. Road Sign Detection in Images: A Case Study. IEEE, 2010.

[10] Stefan Toth. Usage Smartphones for Traffic Sign Recognition and Data Acquisition. Paris, France, 2009.

[11] Bobdavies2000, bornet, garybradski, markasbach, neurosurg, relrotciv, vp153. (2008). [online] Computer Vision Library. [cited 21 May 2014]. Available from: URL: http://sourceforge.net/projects/opencvlibrary/.

[12] Haar wavelet. (2009). [online] [cited 19 May 2014]. Available from: URL: http://en.wikipedia.org/wiki/Haar_wavelet.

[13] Open Computer Vision Library Reference Manual. Intel Corporation. USA, 2001.

[14] Paul Viola and Michael J. Jones. (2001). [online] Rapid Object Detection using a Boosted Cascade of Simple Features. [cited 21 May 2014]. Available from: URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.7597.

Interactive LED Table

*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st PhD.Dr. Sirimonpak Suwannakhun
line 2: *The Project of Administrative Cooperation in Media Arst and Media Technology Curriculum*

(MTA)
line 3: *King Mongkut's University of Technology Thonburi (KMUTT)*
line 4: Bangkok, Thailand
line 5: sirimonpak.sut@kmutt.ac.th

***Abstract*—This the purpose of this research is to design, create, interact and evaluate the performance of Interactive LED table. This table can display the temperature of glass of drink and mix the colors of the light that represent the temperature on both sides of the table where the objects were placed to enhance the enjoyment, colorful, relax for user. In addition, it can show the status of the glass on the table with cold light tones or warm light tones. Evaluation of this project was separated for 3 experiments (25 times per set). First place a glass of drink with kind of different temperatures on LED light drinks holder beside the table and observe the results. Second place object on the interactive LED table then observe the results. Third place a glass of drink with kind of different temperatures on LED light drinks holder beside the table and place object on the interactive LED table at the same moment after that observe the results which were correlate in evaluation performance. We found that the first experiment succeeds 100%, the second experiment succeed 80% and the third experiment succeed 91.05%. In conclusion, the interactive LED table was successful pass all criteria.**

Keywords— interactive LED table, performance, cold tone, warm tone

I. INTRODUCTION

Nowadays, technology has been developed rapidly and it is recognized as if the 5th necessity. Humans apply technology in many aspects in order to make use of it. However, furniture and technology are developed to facilitate in some groups only. Such development does not cover features as the era not only differs development, but also alters humans' daily lives. Some people work at night time, some people normally spend time at night, some people live their lives rapidly while others work hard even during relaxing time. This can cause a problem among people spending nightlife and a problem of urban people in a hurry to find things. An important problem for such people is the danger of picking up things in a low light condition or the danger of not being able to recognize a status of such thing.

As mentioned above, the researchers have designed and developed an interactive table with light color processing system based on temperature level. Related studies and literature were investigated to develop a LED interactive table, using an infrared sensor to detect movement or objects on the LED table. The researchers applied an Arduino program to control the infrared sensor and LED light to change its colors based on glass temperature through cool or warm colors to identify status of a glass on the table to ensure safety use. The project designed interesting and attractive colors of light and the interactive table with light color processing system based on temperature level. This can be a guideline to develop a more interactive tables among users with more devices.

II. METHODOLOGY AND FRAMEWORK PROPOSED

The Interactive LED to study the theory related to the design and creation of the Interactive LED by the color processing of lights according to the rating and suitable for use in Consists of various theories.

A. Cognitive Behavior

Perception has a root from a Latin word "Percipere" which means Per referring to Through, and Cipere which means To Take. This is an important psychological process of a person. Without perception, people cannot recognize or learn anything. Therefore, perception is a way a person looks at the world around himself or herself. Two individuals can perceive a stimulant under the same condition. However, the two individual may Recognize, Select, Organize and Interpret such stimulant differently. This is consistent with the idea of Schiffman & Kanuk (1991), stating that Perception means "a process individual selects, organizes and interprets a stimulant through a meaning". Kast & Rosenzweig stated that perception is an interpretation of a stimulant and a response. It is different among individuals, depending on existing experience. This differs individual behaviors and it's called "Concept".

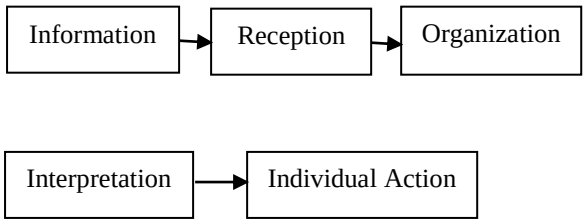


Fig. 1. Recognition process

B. Design work

The Figure 2. shows the general electronic circuit design system It is a working system, starting with researching to identify problems and solutions for the circuit to be designed. With various functional designs in which circuit testing is performed with error correction and there are 10 steps as follows

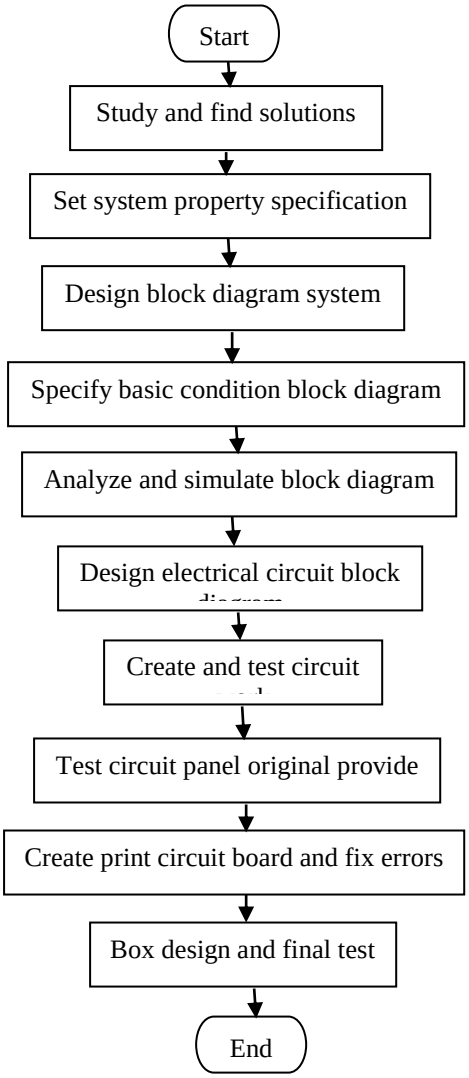


Fig. 2. Functional design

III. PRODUCTION PROCESS AND RESULTS

An intrusion prevention system in the building with an indicator controlled by a microcontroller [18] is a process of development for an intrusion prevention system within an office building, a residence and important places. This system can identify whether a sensor is cut or short and shows an error at each point. They system will enter a warming mode and all devices including an electrical bell, a flashing light and a buzzer will start. There are 8 sensors including a magnetic switch, an infrared and a Pyroelectric Infrared Sensor (PIR) etc. This research applied infrared technology.

A. Design and lighting color processing system based on temperature levels

A Study of Lighting Design Performance Using High Power LED [19] developed lighting design. For years, LED lights have been developed and become high power LED lights which are more efficient than existing ones. The study compared advantages of high power LED lights with other lights. The study consisted of 2 main parts including a high power LED set and a drive circuit. According to a lighting test, applying the high power LED set could result in better lighting.

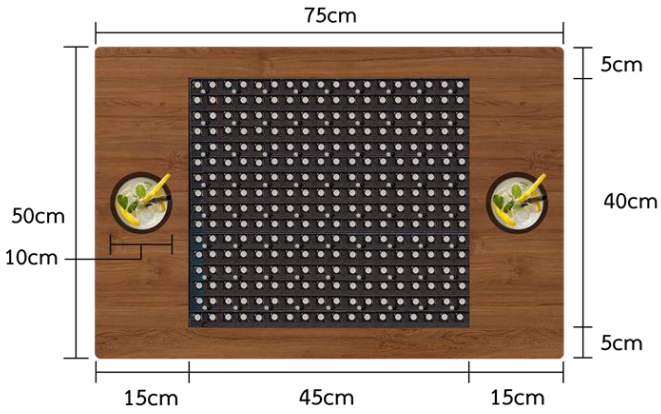


Fig. 3. Design of the interactive table

B. Designing the placement of the device

The interactive table with light color processing system based on temperature level contains devices that work cooperatively together. The devices include an infrared sensor and Neopixel. Positioning such devices shall be balance between table size and the number of devices. The researchers defined the size of 1 channel as 5 centimeters, containing 4 Neopixel lights. The space between each light is 1.5 centimeters. One circuit of an IR Emitter and an IR Receiver or an infrared is connected.

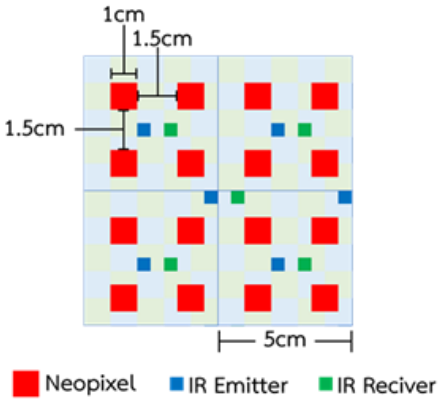


Fig. 4. Infrared and Neopixel sensor position design

The design of the table Which consists of infrared sensor circuit, MCP23017 circuit, temperature sensor circuit and Neopixel circuit. All circuits use Power Supply Switching as the power supply to the circuit.

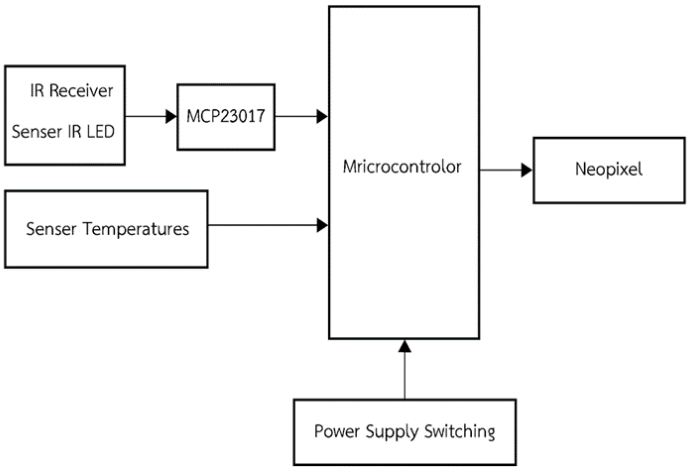


Fig. 5. Design of the interaction table circuit with the light color processing system according to the temperature level

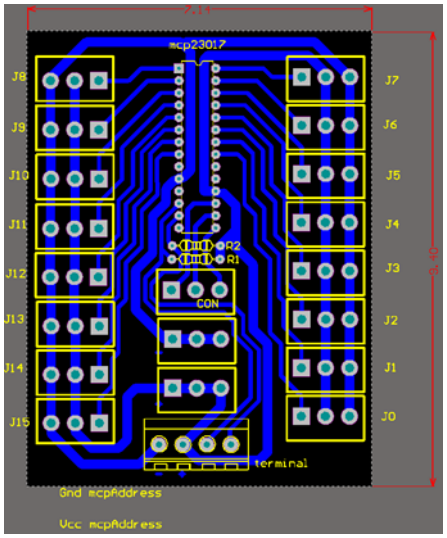


Fig. 6. PCB design of IC circuit MCP23017

C. *Development and construction of software and hardware of interaction with light color processing systems based on temperature level*

Performance of the interactive table with light color processing system based on temperature level begins with infrared checking. If no object blocks an infrared signal, Input value equals 1 and it will return to the beginning. If an object blocks an infrared signal, Input value equals 0. Then, temperature is checked to identify its color. The color is displayed at Neopixel.

An infrared sensor signal circuit detects an object blocking infrared, using a principle of transmitting light wave to reflect on an object and detect light reflection. An IR emitter and an IR receiver are main devices that detect as far as object, depending on a resistant and an infrared signal circuit.

Identify applicable funding agency here. If none, delete this text box.

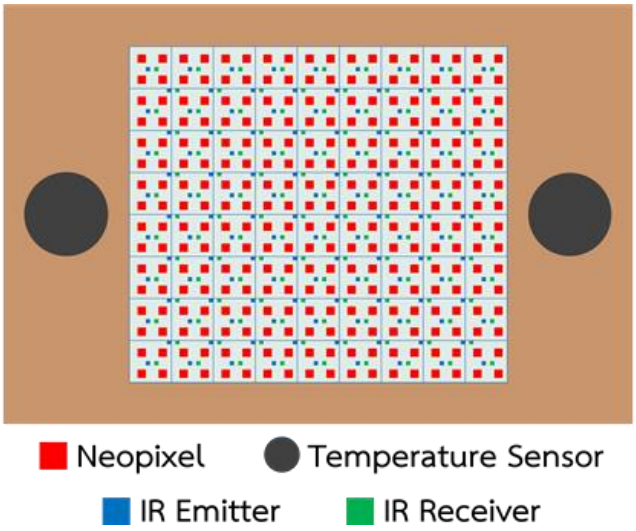


Fig. 7. Components of the interaction table

D. *Results*

Test of glass positioning showed that 100 percent passed the set criteria. A glass holder consists of a temperature sensor and stable Neopixel. From the test, the temperature sensor could receive accurate value as commanded. From the interoperability testing of the temperature sensor and the Neopixel, a glass could be placed and the display can be shown according to set temperature color smoothly. The tests were conducted 25 times and 100% passed the tests.

Test of placing an object in front of the table showed that there were 20 times of accurate indicator lights identifying object position. There were 5 times of inaccurate indicator lights due to work interruption. There were 20 times of accurate lights moving along a glass position. There were 5 times of inaccurate lights due to work interruption. The tests were carried out 25 times or 80 percent. 20 percent failed the tests because positions of objects were not in line with the infrared sensor signals and the objects moved to fast ompared to the control command.

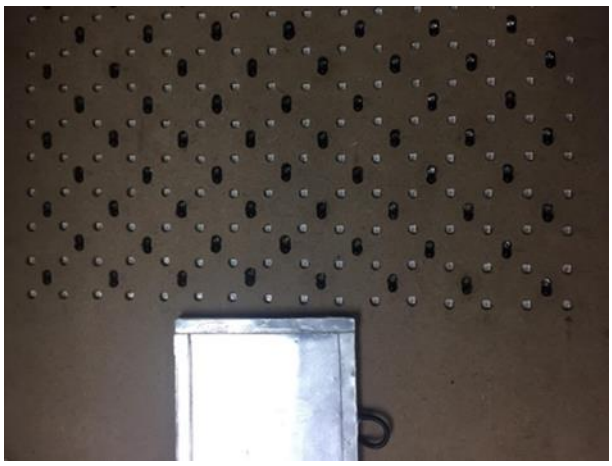


Fig. 8. Neopixel infrared sensor circuit and temperature sensor circuit on a piece of wood at the specified position

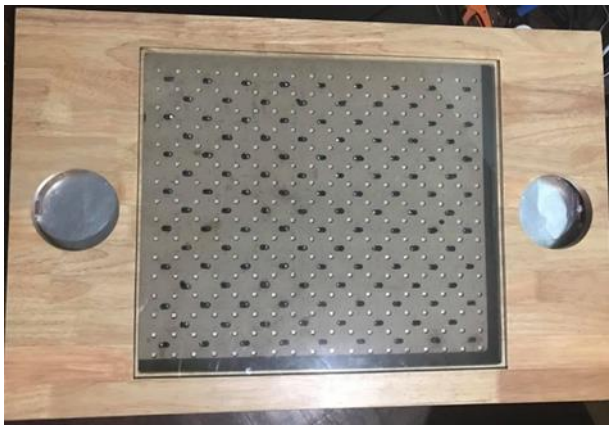


Fig. 9. Assembly and connection of the circuit to the table

IV. SUMMARY

In this paper proposed tests of interoperability of placing a glass and an object in front of the table showed that there were 15 times of color mixing according to colors from temperature in both sides. The test of color mixing applied RGB concept. The mixed colors were stable and became a single color. There were 25 tests or 100 percent. Errors in color mixing may occur because some Neopixel was outstanding or faded over other Neopixel, because the circuit consumed too much power and because Neopixel was a drive circuit that may lead to errors. There were 20 times of accurate performance evaluation of object indicator lights and moving lights or 80 percent. 20 percent failed because positions of objects were not in line with the infrared sensor signals and the objects moved to fast compared to the control command. The tests of interoperability of placing a glass and an object in front of the table was 91.05 percent. It could be concluded that it passed the test criteria effectively.

ACKNOWLEDGMENT

Media Technology Curriculum, King Mongkut's University of Technology Thonburi (KMUTT), Thailand.

REFERENCES

[1] N. Methiphiwat, "Factors Affecting the Use of Bank Account Payments for Residents of Residents in Water Housing Case Study," Nonthaburi Provincial Waterworks Authority, Public Management Program College of Commerce Burapa university, 2010.

[2] M. Khaosriwong, "Student Ethics as Perceived by Students Teachers and parents of Chiang Mai University Demonstration School," Educational Psychology and College Counseling Program, Chiang Mai University, 2008.

[3] Artaddesign, "RGB and CMYK are different and are suitable for any type of design," 22 September 2017
Available: http://www.artaddesign.com/frontend/RGB- and -CMYK-Difference- and Suitable for any type of design --article-10.php, 2009.

[4] C. Tangsirivorakul, "Study of Lighting Design Performance by Using High Power LED," College of Industrial Technology, King Mongkut's University of Technology North Bangkok and King Mongkut's Institute of Technology Ladkrabang, 2009.

[5] L. Shop, "Color Mixing (Color Cycle)," 22 September 2017
Available: http://ananauo.lnwshop.com/article/3/ Color Mixing Color Cycle, 2015.

[6] Blogger, "Tone of Color," 22 September 2017
Available: http://piyadacolortheory.blogspot.com/2014/01/tone-of-color.html, 2014.

[7] In House Practical Training (IHPT), "PCB Making ," 22 September 2017
Available: dk.coe.psu.ac.th/lecture/ihpt345/Report/Doc/4310458.doc, 2001.

[8] Know2learning, "PCB Type," 22 September 2017
Available: http://know2learning.blogspot.com/2014/06/5.html, 2014.

[9] K9APE,"Circuit Board: print circuit board (PCB)," 4 November 2017
Available: http://worthatry.tistory.com/361, 2009.

[10] Ourpcbte, "4 Layer PCBA PCB Assembly Immersion Tin Finishing," 4 November 2017
Available : http://www.ourpcbte.com/products-detail/4-layer-pcba/, 2007.

[11] Wenkm, "Flex Circuit Material Flexible Circuit Board With Connector Panasonic Flex Circuit Material," 6 January 2018
Available: http://wenkm.com/flex-circuit-material/flex-circuit-material-flexible-circuit-board-with-connector-panasonic-flex-circuit-material/, 2018.

[12] A. Shoppen, "Arduino Mega 2017 R3," 6 January 2018
Available: https://arduinoshoppen.dk/produkt/arduino-mega-r3, 2018.

[13] T. Quads, "IR LED - For Repairing Trackmate Transponder," 6 January 2018
Available: http://www.twistedquads.com/IR-LED--For-Repairing-Trackmate-Transponder_p_1845.html, 2016.

[14] W11Shop, "IR Receiver," 6 January 2018
Available: http://www.w11stop.com/ir-receiver, 2016.

[15] R. Channel, "Implementing IR Infrared Obstacle Avoidance Sensor Module," 11 January 2018
Available: https://robotsiam.blogspot.com/2016/10/ir-infrared-obstacle-avoidance-sensor.html, 2016.

[16] Raspberrysource Shop, "Stainless steel package Waterproof DS18b20 temperature probe temperature sensor," 11 January 2018
Available: http://www.raspberrysource.in.th/product/90/stainless-steel-package-waterproof-ds18b20-temperature-probe-temperature-sensor, 2014.

[17] Arduino All, "NeoPixel LED WS2812B RGB Matrix 10x10 IC DRIVER Built-In 5Vdc Black Board," 11 January 2018
Available: https://www.arduinoall.com/product/1142/neopixel-led-ws2812b-rgb-matrix-10x10-ic-driver-built-in-5vdc- Black board, 2017.

[18] Aduino All, "16 Pin I / O Expansion IC for Arduino," Number MCP23017, 11 January 2018
Available: https://www.arduinoall.com/product/24/ic-leg extension-16-leg-i-o-for -arduino-number-mcp23017, 2018.

[19] K. Phliwethaisong, " Intrusion prevention system in the building With the status of each operation point Controlled by micro controller," Master's thesis Computer Science Program, King Mongkut's University of Technology North Bangkok, 2007.

[20] Geek, "PC817," 11 January 2018
Available: http://geek.kg/optron/, 2013.

[21] Relong, "LM358P LM358N LM358 DIP-8 Chip IC Dual Operational Amplifier - Op Amp," 11 January 2018
Available: https://www.lelong.com.my/lm358p-lm358n-lm358-dip-8-chip-ic-dual-operational-amplifier-op-amp-robotedu-194236242-2019-07-Sale-P.htm, 2017.

[22] Obd2, "BC547B NPN Transistor TO-92 Type," 25 Pieces, 11 January 2018
Available : http://www.fabotronix.com/product/bc547-npn-transistor/, 2003.

[23] M. Young, "The Technical Writer's Handbook. Mill Valley," CA: University Science, 1989.

AUTHORS

Sirimonpak Suwannakhun is works also as a visiting professor of media technology curriculum at the Universities of King Mongkut's University of Technology Thon-buri (KMUTT), Thailand.

Predicting Chance of Success on Epiretinal Membrane Surgery using Deep Learning

Suvimol Reintragulchai[a], Thanaruk Theeramunkong[a, b], Paisan Ruamviboonsuk[c], Vorarit Jinaratana[c], Natsuda Kaothanthong[d, *]
[a]Information and Communication Technology for Embedded System, SIIT, Thammasat University, Pathum Thani, Thailand
[b]Associate Follow, Royal Society of Thailand, Bangkok, Thailand
[c]Department of Ophthalmology, Rajavithi Hospital, Bangkok, Thailand
[d]School of Management Technology, SIIT, Thammasat University, Pathum Thani, Thailand
Email: {meteogon38, coeplus.rajavithi}@gmail.com, {thanaruk, natsuda}@siit.tu.ac.th, Art_jinaratana@hotmail.com

***Abstract*— A preliminary study on predicting chance of success on an epiretinal membrane surgery is studied. Given an optical coherence tomography image, the study shows that the multilayer perceptron neural network can achieve 91.0% accuracy. Due to an unbalance of the images of success and failure classes, under-sampling and over-sampling are applied. For over-sampling, the images in the failure class are duplicated to balance the number of images compared to the success class. Utilizing the balance dataset, the prediction performance is improved from 91.0% to 93.0% for over-sampling. With the exploitation of, the salient region for training the model and predicting the outcome. The salient region is manually segmented to express the fovea in the OCT. The experimental results evidence an improvement of 1.0% with achievement of 94.0% accuracy.**

Keywords—Epiretinal Membrane, Machine learning, retinal OCT image.

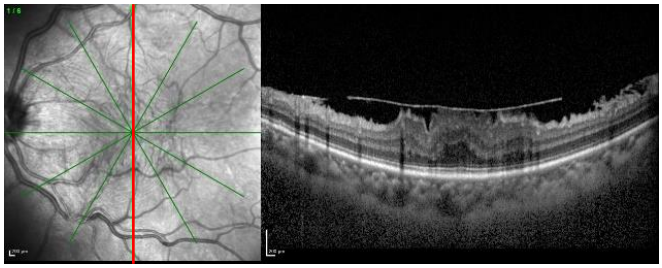
I. INTRODUCTION

Machine learning is applied in many researchers including in medical image analysis [1] to assist in the analysis to be faster. There are 2 methods for the diseases related to a macular, which are medication and surgery. Age-related macular degeneration (AMD) and Diabetic macular edema (DME) can be cured by using medicines. Epiretinal Membrane (ERM) is cured by a surgery.

Epiretinal Membrane (ERM) is condition that a very thin layer of scar tissue forms on the surface of the retina as shown in Fig. 1. This thin layer causes a deformation in the retinal. To treat ERM, a surgeon operates on the affected eye to remove the membrane on the retina. However, the outcome of the operation may not be successful. In this way, being able to predict the surgery outcome is able to reduce the pain and recovery time for the patient that the outcome will not be successful.

Many papers Optical Coherence Tomography (OCT) image such as Reza et al. [2] applies a random forest (RF) and a convolutional neural network (CNN) to extract a feature and classify the abnormal macula from OCT images. Qingge et al. [3] applied convolutional neural network (CNN) to find salient layers for the analysis of age-related macular degeneration and diabetic macular edema. Other researches can be found from [4], [5], [6] and [7]. There are a few researches for predicting the operation outcome Agnieszka et al. [8] propose an algorithm to find ERM surface by using pixel intensity analysis and graph search.

This paper presents a preliminary study of applying machine learning techniques to predict the operation outcome of the epiretinal membrane operation. In addition, the salient location on the epiretinal and the performance of the classification result is being study by cropping a fovea region for the machine learning. Three machine learning techniques, which are a multilayer perceptron neural network (MLP), support vector machine (SVM) and random forest (RF) are used to compare the prediction outcome of an original OCT image and fovea as an input.



(a) (b)

Fig. 1. Example of a retinal OCT image with ERM (a) a radial reference lines scan of OCT image and (b) a cross-sectional image of the retina of red-line scan

II. LITERATURE REVIEW

To segment the layer of the membrane for analysis, Lang et al. [4] classify the retinal layer from OCT images using RF with Candy-inspired boundary (CAN) and random forest (RF) with Optimal graph search (GS) algorithm to classify retinal layer of OCT images. The two methods give an error of the layer's boundary below 3.5 micro-millimeter and 4.5 micro-millimeter as compare to the ground truth.

Bhavna et al. [5] segments the surface of macula in SD-OCT images using MLP and graph theory. The outcome shows that the graph theory can give more precise segmentation result.

* Corresponding author.

For a classification task, Reza et al. [2] applies a convolutional neural network (CNN) and random forest (RF) to find a feature and classifier an abnormal macula in retinal OCT. The outcome average precision was 99.33% on classify between diabetic macular edema (DME) and normal and 98.67% on classify between Age-related macular degeneration (AMD) and normal.

In addition, many researches focus on the feature extraction. Qingge et al. [3] applies a convolutional neural network (CNN) to learn the abstract semantic feature of the retinal OCT B-scan of Age-related macular degeneration (AMD) and Diabetic macular edema (DME). The outcome average precision was 98.2%.

Motozawa et al. [7] propose OCT-based Deep-learning Models for classifying Normal and AMD and Exudative AMD and Non-Exudative AMD change. The outcome of classifying Normal and AMD OCT images was 100% sensitivity, 91.8% specificity and 99.0% accuracy. The outcome of classifying AMD having Exudative or not exudative change with 98.4% sensitivity, 8803 specificity and 93.9% accuracy.

Agnieszha et al. [8] propose an algorithm to find ERM surface by using pixel intensity analysis and graph search. The outcome show that the graph search technique gives better ERM segmentation results than the pixel intensity analysis.

III. DATA & METHOD

A. *Pre-processing and Original Image*

In this work, the OCT images of patient were taken before the surgery. For each image, there are 2 parts, 1) a radial reference lines scan and 2) a cross-sectional image of the reference line scan as in Fig. 1.

The image showing a reference line was removed during the preprocessing. The outcome of the preprocess is shown in Fig. 2 and it is denoted as an original image.

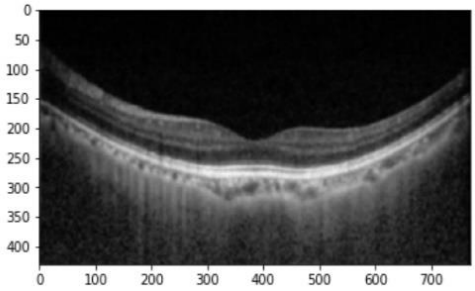
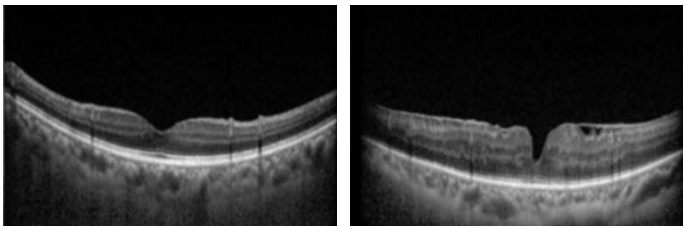


Fig. 2. Example of original image after cropped only cross-sectional image from OCT image

The label of each image is assigned according to the outcome of the operation of each patient. In other words, all the images of the same patient are assigned using the same operation outcome. A pair of the OCT image of 'success' and 'failure' outcome is shown in Fig. 3. All retinal OCT images in this experiment come from 42 patients including of 27 patients were improved after taking an operation and 15 patients were not improved as shown in TABLE I. The retinal OCT image of each patient may not equal depended on type of radial line scan as shown in TABLE II. The total original OCT image of retina is 380 images including 255 success images and 125 Not-success image with 430 x 770 pixels resolution.



(a) (b)

Fig. 3. (a) Example of Success class image and (b) Example of Failure class image

TABLE I. TOTAL NUMBER OF PATEINT AND THE NUMBER OF SCANS IN THE ORIGINAL DATASET.

Spectral Domain OCT image	**Number of Patients**	
	Success	***Failure***
Radial scan 6 lines	23	13
Radial scan 16 lines	-	1
Radial scan 24 lines	1	-
Radial scan 31 lines	3	1
Total in each class	27	15
Total		42

TABLE II. TOTAL NUMBER OF RETINAL OCT IMAGE IN EACH CLASS OF ORIGINAL DATASET.

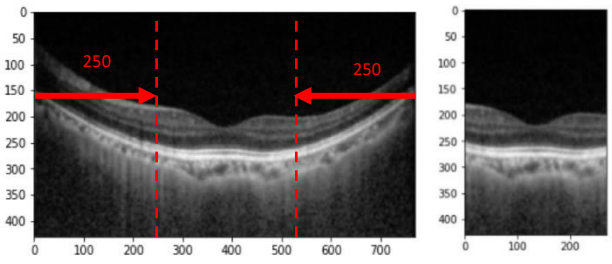
Spectral Domain OCT image	**Number of images**	
	Success	***Failure***
Radial scan 6 lines	138	78
Radial scan 16 lines	-	16
Radial scan 24 lines	24	-
Radial scan 31 lines	93	31
Total in each class	255	125
Total		380

B. *Cropped image*

To study an effect of the selection of the salient area to the classification result, a fovea area is cropped from the original image. The fovea is a small depression in the retina of the eye where visual acuity is highest. The center of the field of vision is focused in this region, where retinal cones are particularly concentrated.

The fovea area is obtained by cropping the center area of the original OCT image. In this work, the salient area is cropped by moving the left and the right borders of the original image

by 250 pixels from both sides. The outcome of the cropped image is shown in Fig. 4. The size of the cropped image is 380 pixels with 430 x 270 height and width.



(a) (b)

Fig. 4. (a) Delete 250-pixels area from left-side and right-side of the original OCT image and (b) Example of Cropped images.

C. Balanced Image Data Set

Regarding to the total number of images in each class as shown in TABLE II, the number of images of 'Failure' class is less than half of the 'Success' class. In this way, we also applying a sampling method to avoid the unbalanced problem. Two approaches are utilized: over-sampling and under-sampling.

For over-sampling, the number of images in 'Failure' class is increased by randomly selected to obtained six images per patient (125 OCT images become 90). After that, duplicate each image twice. The number of 'Failure' class after applying an over-sampling is 180 images.

For under-sampling, the OCT images of the patients with more than six reference lines as shown in TABLE I are randomly selected to obtained six images per patient. The number of 'Success' class after applying an under-sampling is 162 images.

After this process, the total retinal OCT image of Success class become 162 images and Failure class become 180 images

D. Feature Extraction

The image's intensity is used as a feature. For each pixel in an image, the intensity value is a positive integer from 0 to 255, where 255 is the brightest. To allow a classifier to understand the image's content, the two-dimensional array of intensity of each pixel in the image is transformed into a vector of size $1 \times N$, where N is the total number of pixels in an image.

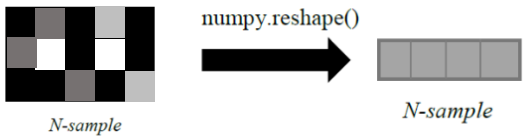


Fig. 4. 2-dimensional array of intensity of each pixel in image converted to 1-dimentional array.

IV. Experiment

A. Dataset

In this experiment, the data is organized into 6 group as follow:

1). Original dataset: the original retinal OCT images.

2). Cropped dataset: the original retinal OCT images after cropped image process.

3). Under-sampling dataset: the original retinal OCT images after Balanced image process.

4). Under-sampling-cropped dataset: the retinal OCT image of down-sampling dataset after cropped image process.

5). Over-sampling dataset: the original retinal OCT images after Balanced image process.

6). Over-sampling-cropped dataset: the retinal OCT image of up-sampling dataset after cropped image process.

TABLE III. TOTAL NUMBER OF RETINAL OCT IMAGE IN EACH DATASET.

Image type	Dataset	Class		
		Success	***Failure***	***Total***
Original image	Original	255	125	380
	Under-sampling	162	125	287
	Over-sampling	162	180	342
Cropped image	Original	255	125	380
	Under-sampling	162	125	287
	Over-sampling	162	180	342

In this experiment, a cross-validated is applied. Given each data set, it is split into five subsets, also called fold. The training and test sets using the ratio of 4:1, where 4 folds are applied for training the prediction model. The experiment is repeated for five times to ensure that every as been used as a test data. See Fig. 5 for illustration. The accuracy of each model is an average score of 5 consecutive experiments.

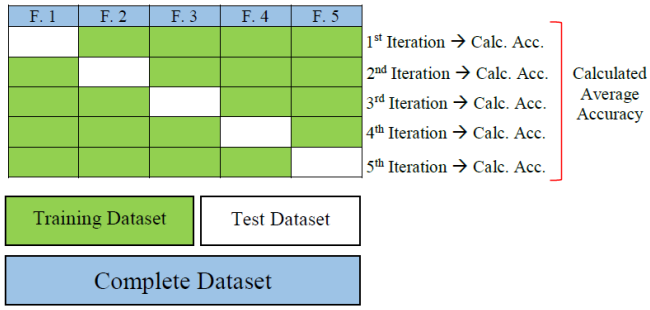


Fig. 5. 5-Fold cross-validation scheme.

B. Machince Learning Model

Three machine leaning techniques are applied in this experiment: a multi-layer perceptron neural network (MLP), a support vector machine (SVM), and a random forest (RF). In this work, the available library Sci-learn of Python is being used.

For MLP, 100 hidden layers setting is applied. For SVM, the C value is set to 1, A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly. Lastly, 100 trees are set as the maximum number of the trees for RF.

C. Evaluation

A confusion matrix is applied to measure the efficiency of prediction model of each dataset. True positive (TP), False positive (FP), False negative (FN), and True negative (TN) values as shown in TABLE IV shows the performance of the model. TP and TN show the correctly predicted result. On the other hand, FN shows incorrect predicted in term of predicted as normal while it is abnormal, while FP incorrect predicted in term of predicted as abnormal while it is normal.

TABLE IV. CONFUSION MATRIX.

		Predicted	
		Success	***Failure***
Actual	***Success***	True Positive (TP)	False Positive (FP)
	Failure	False Negative (FN)	True Negative (TN)

Sensitivity or true positive rate is defined as the proportion of people with the disease who will have a positive result. In other words, a highly sensitive test is one that correctly identifies patients with a disease. It can be computed as follow:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

Specificity or true negative rate is the proportion of people without the disease who will have a negative result. In other words, the specificity of a test refers to how well a test identifies patients who do not have a disease. It can be computed as follow:

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (2)$$

Accuracy also calculate from result of confusion matrix as follow:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

V. EXPERIMENTAL RESULT

The experimental results of the three machine learning techniques using each data set are shown in TABLE V. The best accuracy of the three models is MLP. The highest accuracy is 0.94 with over-sampling-cropped dataset. Using the same data set, SVM achieves 0.92 accuracy. The highest accuracy for RF is 0.77 using the original-cropped dataset. Considering the accuracy of each dataset presented in TABLE V., the highest accuracy of the original dataset is 0.91, under-sampling dataset is 0.90, and over-sampling dataset is 0.93. By cropping the fovea area, the experimental result shows an improvement of the accuracy as compared to the original image. See TABLE V.

TABLE V. ACCURACY OF EACH MACHINE LEARNING MODEL FOR EACH DATASET.

Image type	**Dataset**	**Accuracy of each model**		
		MLP	***SVM***	***RF***
Original image	Original	0.91	0.91	0.75
	Under-sampling	0.90	0.90	0.74
	Over-sampling	0.93	0.91	0.73
Cropped image	Original	0.92	0.91	0.77
	Under-sampling	0.92	0.91	0.70
	Over-sampling	0.94	0.92	0.72

The sensitivity and the specificity values of MLP of each data set is shown in TABLE VI. The result shows the over-sampling of cropped image achieve the sensitivity of 0.95 while its specificity is 0.88. In this way, this dataset can predict the positive outcome rather than the negative ones. On the other hand, the under-sampling of the original image achieves the highest specificity of 0.96. The comparison of the other datasets can be found in TABLE VI.

TABLE VI. SENSITIVITY, SPECIFICITY AND ACCURACY OF MLP MODEL.

Image type	**Dataset**	**Sensitivity**	**Specificity**	**Accuracy**
Original image	Original	0.90	0.94	0.91
	Under-sampling	0.83	0.96	0.90
	Over-sampling	0.91	0.90	0.93
Cropped image	Original	0.93	0.92	0.92
	Under-sampling	0.94	0.84	0.92
	Over-sampling	0.95	0.88	0.94

TABLE VII shows the confusion matrix result of MLP of each dataset of the last cross validation. Considering the situation that the patient with a good operation outcome misses the operation, the model wrongly predicts the outcome. On the other hand, the patient with a worsen outcome were wrongly predicted as a good outcome.

TABLE VII. CONFUSION MATRIX OF MLP MODEL WITH OVER-SAMPLING CROPPED DATASET.

		Predicted	
		Success	***Failure***
Actual	***Success***	41	7
	Failure	2	53

Sensitivity = 0.95, Specificity = 0.88

TABLE VII shows the confusion matrix of MLP model using over-sampling cropped image dataset. The model predicted that 43 patients will have an improved eyesight, but 2 patients were having a worsen outcome. On the other hand, 60 patients are predicted as a worsen eyesight after the surgery. However, 7 of them have a success outcome. This means that they miss the chance of having an improved eyesight after the surgery.

Example of the images that archives a good and a bad result are shown in TABLE IX. The first three images are correctly predicted using the MLP model with over-sampling cropped image dataset. The last three images are the result that the model wrongly predict the outcome. Similarly, example of the result using over-sampling original images is shown in TABLE VIII.

TABLE VIII. EXAMPLE OF THE PREDICTED RESULT USING MLP AND OVER-SAMPLING ORIGINAL IMAGE .

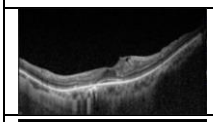
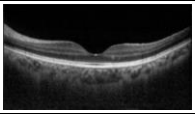
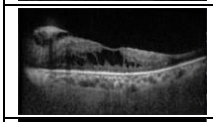
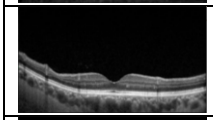
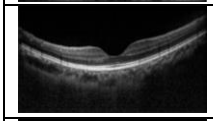
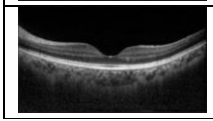
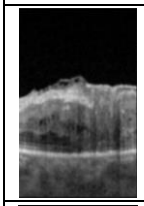
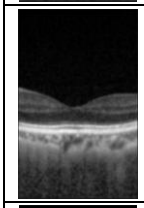
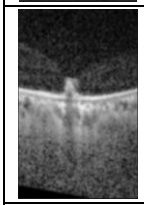
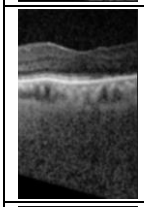
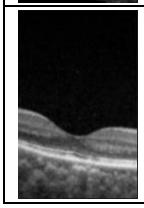
Input	**Actual class**	**Probability**		**Predicted result**
		Success	***Failure***	
	Success	0.9939	0.0061	Success
	Success	0.9872	0.0128	Success
	Failure	0.0020	0.9980	Failure
	Failure	0.8250	0.1750	Success
	Success	0.0004	0.9996	Failure
	Success	0.0010	0.9990	Failure

TABLE IX. EXAMPLE OF THE PREDICTED RESULT USING MLP AND OVER-SAMPLING CROPPED IMAGE .

Input	**Actual class**	**Probability**		**Predicted result**
		Success	***Failure***	
	Success	0.9901	0.0099	Success
	Success	0.9977	0.0023	Success
	Failure	0.0004	0.9996	Failure
	Failure	0.9996	0.0004	Success
	Failure	0.9996	0.0004	Success
	Success	0.3551	0.6449	Failure

VI. CONCLUSION

The preliminary result of the outcome prediction for unmet treatment for an epiretinal membrane shows that the MLP prediction model achieves the highest accuracy of 0.94 using over-sampling of the cropped image dataset. The result also shows that the modification of the input image by forcing the model to learn from the salient region, i.e. fovea, achieves a better accuracy. In the future work, the best location of the salient for the epiretinal membrane will be study.

ACKNOWLEDGMENT

This paper is partially supported by Thailand Science Research and Innovation (TSRI) under the contract number RTA6280015.

REFERENCES

[1] L. Greet, K. Thijs, E. B. Babak and S. A. A. Arnaud, "Medical Image Analysis," pp. 60-88, 2017.

[2] R. Reza, M. Alireza, R. Hossein and H. Fedra, "Automatic disgnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forest classifier," *Journal of Biomedical Optics*, vol. 23, 2018.

[3] J. Qingge, H. Wenjie, H. Jie and S. Yankui, "Efficient Deep Learning-Based Automated Pathology Identification in Retinal Optical Coherence Tomography Image," *Algorithm*, 2018.

[4] A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. A. Calabresi, H. S. Ying and J. L. Prince, "Retinal layer segmentation of macular OCT images using boundary classification," *BIOMEDICAL OPTICS EXPRESS*, vol. 4, pp. 1133-1152, 2013.

[5] A. J. Bhavna , A. D. Michael, H. M. Mattew, J. Woojin , S. H. Elliott, K. H. Young, K. Randy and G. K. Mona, "A combined machine-learning and graph-based framework for the segmentation of retinal surfaces in SD-OCT volumes," *BIOMEDICAL OPTICS EXPRESS*, vol. 4, 2013.

[6] G. K. Mona, A. D. Michael , W. Xiaodong, R. R. Stephen, B. L. Trudy and S. Milan , "Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 28, pp. 1436-1447, 2009.

[7] N. Motozawa, G. An, S. Takagi, S. Kitahata, M. Mandai, Y. Hirami, H. Yokota, M. Akiba, A. Tsujikawa, M. Takahashi and Y. Kurimoto, "Optical Coherence Tomography-Based Deep-Learning Models for Classifying Normal and Age-Ralated Macular Degeneration and Exudative and Non-Exudative Aeg-Related Macular Degeneration Changes, " *Ophthalmol Ther*, 2019.

[8] S. Agnieszka, M. Tomasz, D. Adam , S. Marcin, R. Piotr and M. Elzbieta, "Novel Full-Automatic Approach for Segmentation of Epiretinal Membrane from 3D OCT Images," in *Signal Processing Algorithms, Architectures, Aprangements and Applications*, Poznan, POLAND, 2017.

Classification of Depressed Speech Samples with Spectral Energy Ratios as Depression Indicator

Thaweewong Akkaralaertsest
Department of Electronic and Telecommunication,
Faculty of Engineering,
Rajamangala University of Technology Krungthep Bangkok, Thailand
thaweewong.a@mail.rmutk.ac.th

Thaweesak Yingthawornsuk
Media Technology Program,
School of Architecture and Design,
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
thaweesak.yin@kmutt.ac.th

***Abstract*— This research study aimed to investigate the characteristics of the Spectral Energy Ratios (SER) determined from the Power Spectral Density (PSD) of the spoken speech samples used to represent the severity level of emotional illness such as Depression in quantitative measure. Situation could be getting worst for a person who suffers from such illness with the elevated severity of symptom. When the symptom of severe depression strikes, a depressive person might be at high risk of committing suicide. The prevention of suicide is necessary for depressed persons to save life by admitting them in time and providing the proper treatment under supervision of clinical specialist. Prediction is primarily one of the most important tasks in the prevention of life-threatening risk from suicide. Researcher has attempted to adapt the speech processing techniques into a clinical diagnosis of emotional illness. In this study a full-band energy and further several sub-band energies estimated from the four frequency bands with each 625-Hz bandwidth were computationally extracted from the categorized speech samples and consequently formed the parameter models for classifications. As result shown, the averaged value of correct classification was obtained to be effectively approximate 80%, when training and validating classifiers with 35% and 65% of the extracted SER features, respectively.**

Keywords—speech, depression, spectral energy ratios, classification,

I. INTRODUCTION

Depression is known as a common outcome in persons with serious mental disorders. However, it remains a phenomenon that is under-researched and barely understood. Moreover, methods to identify persons who are at elevated risk of suicide are highly needed in clinical practice. This study addresses on how to characterize the acoustical related spectrum in spoken sound of persons with imminent depressive potential in case of severe recurring depression to gain more comprehensive understanding. The contribution earned from study could lead us to develop more effective techniques of the assessment of severe depression potential.

In previous published research studies [1-3], the analytical techniques have been studied and developed to identify if subjects were in one of three mental states: healthy (control), depressed or high-risk suicidal. In the past the vocal cues have been studied and employed in diagnosis of the syndrome underlying a person's emotional state by physicians with expertise [4-5], but not widespread in clinical use nowadays. Researches have shown that the emotional arousal produces changes in the speech production scheme by affecting the respiratory, phonatory, and articulatory processes encoded into the vocal signal [6]. Such arousal produces a tonic activation of striated musculature, and the sympathetic, and parasympathetic nervous systems [7]. Changes in human's bioelectric signals such as ECG, EMG, blood pressure and respiratory patterns, including phonatory, and articulatory functions in speech production system [8] directly relate to emotions. Therefore, the emotional illness can be the cause of quantitative changes in related acoustical parameters. At some measurable level, change in acoustical parameters possibly indicates the near-term suicidal state. Emotional content of the voice can be associated with the acoustical parameters like amplitude level, dynamic range, contour of fundamental frequency, prosody of speech, vocal energy, distribution of energy in frequency range, location of dominant spectral peaks, formant frequencies, and a variety of temporal measures [9]. As well recognized from research studies in the past depression has been a major impact on the characteristics of acoustic in voice when compared to normal speakers. It has been concluded that prosodic properties in speech are different such as pitch, timbre, loudness, duration are slower in a spoken sound of depressive speakers. Moreover, the spectral energy in depressed speech distributes differently over a frequency range of 0-5KHz as compared to the healthy speakers.

The purpose of study is to investigate more on energy contained in speech and further to determine the relationships of spectral energy among several different subdivided bandwidths within a similar frequency range of 0-2.5KHz at which the most energy in speech spectrum are distributed and concentrated. The further step is classification state based on the extracted energy from different groups of speech samples and the most importance of improving in the accurate prediction or the assessment of the symptom level is faster than the current standard of care. Technology that has developed in the field of computers such as IoT, AI, mobile applications, and machine learning helps us to accelerate a work process. Machine learning technique like support vector machine (SVM) was used in this study for more accurate classification. In this study speech data obtained from the previous recording in the past will be further continuously analyzed with permission granted on data accessibility.

This paper is orderly organized as follows: Section II discusses the related work. Section III provides a detailed description of all categorized speech database, the subject populations described, the extraction of PSD based energy and its ratios performed on all speech dataset. Section IV presents the results of study and compares the acoustical properties among categorized speech groups and discussion on results.

Sections V concludes the findings on the salient feature and result.

II. Related Work

Moore et al. [11, 12] proposed the acoustic speech features based on prosodics of speech, which relate Fo, speaking rate and short-time energy, including glottal spectrum slope and spectral features as the statistical quantifiers in classifying both male and female speech collections of normal healthy speakers and depressive speakers. In female study, he reported that the energy contour estimated from a frame by frame basis was found to be highly effective in classifying the female speech samples with the maximum classification. Cummins et al. [13, 14] presented the mel-frequency cepstrum coefficients (MFCC) showing the statistical significance in classifying the depression from the healthy control. Low et al. [17] identified depression with the low-level descriptors and statistical characteristics.

III. Methodology

A. Database

All speech recordings were collected from three different categorized groups of depressed, high-risk suicidal and remitted female speakers who were volunteers in a suicide prevention research program and clinically evaluated by psychiatrist and specialist in depression. The recording database used in this re-investigation consists of total thirty females clinically categorized into three groups of speech samples and the age range of all female speakers is from 25 to 65 years old. The spoken sound of each subject was collected from two different sessions of recording; first is the main interviewing session made between the psychiatrist and subject, and second is the recording made in the post session while a subject reads out the predetermined part of book. The passage in post session is composed of the standardized texts commonly used in speech science it has all normal sounds of texts and it is phonetically balanced [10].

In this study all speech records collected from first recording session were used in processing as off-line analysis throughout the entire study procedure. In recording process, each speech sample was digitized via a 16-bit analog-to-digital (A/D) converter at a 10-kHz sampling rate with an anti-aliasing filter (i.e., a 5-kHz Lowpass Filter). The background noise and any artifact sound rather than the subject's original sound were removed manually through monitoring speech signal waveform on the audio editor. Before the extraction of the PSD based energy features, all speech samples were first detected for their voiced, unvoiced and silent segments in each speech file using the Dyadic Wavelet Transformation (DWT) in computing the weight scale to identify any segments of 25.6ms as voiced segments. After all voiced segments were tested and detected, they were then stored as files for further off-line processing.

B. Extraction of Energy Ratios

Power spectral density (PSD) of the voiced segment were obtained by using the traditional technique of PSD estimation based on Welch method with a 50% overlapping on segments. The algorithm to estimate PSD was written in MATLAB using 256-point fast Fourier transforms (FFT) to estimate the spectra within a 25.6-ms analyzing window over entire voiced speech sample. Six main parameters were estimated from all voiced segments. The first four parameters are the spectral energies calculated from four different frequency bands: first band from 0-625Hz, second band of 625Hz-1.250KHz, third band of 1.250-1.875KHz, and finally fourth band of 1.875-2.5KHz. As observed from the spectrum of voiced speech signal, most energy distributed and concentrated within this range of frequency and not beyond this upper frequency bound.

For each sub-band energy calculated from each 625Hz sub-band, the percentages of sub-band energy to the total energy along a frequency range from 0Hz to 2.5KHz, or called "Spectral Energy Ratios" (SER's) were calculated and stored as a set of input parameters to state of classification. The other two features are the value of peak power and the frequency of the peak power. The following code shows steps to calculate all spectral energy in each frequency band and its ratio respective to total spectral energy.

Pseudo code to estimate Spectral Energy and Ratios:

1. Setting a frequency range of 0-2.5KHz to estimate PSD;
2. Separating a frequency range into four 625Hz bandwidths;
3. For (all frames of segmented voiced signal){
4. Separate a whole voiced signal into frames with 25.6ms length;
5. Estimate PSD of each frame using Welch method;
6. Calculate a total energy over a 0-2.5KHz range using "trapz" function;
7. Calculate sub-band energies from all four 625Hz bandwidths;
8. Calculate all ratios of band energy from band 1-4;
9. Collect all estimated energy parameters;
10. }end

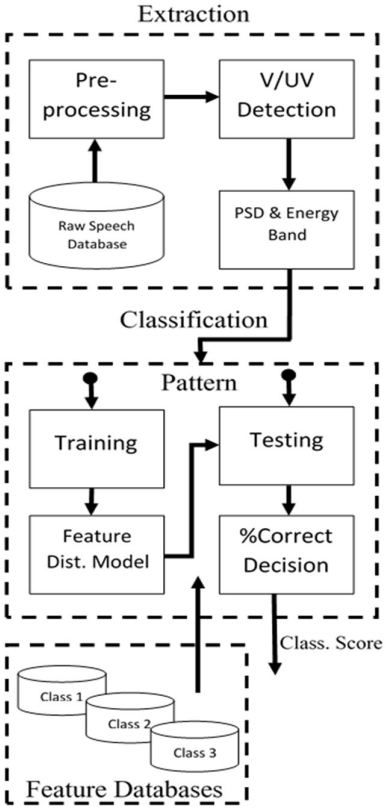


Fig. 1. Overall procedure of extraction of subband spectral energy and energy ratios, and classification

C. *Statistical Analysis and Classification*

All SER parameters were arranged and stored in a matrix form for statistical analysis. Each of the acoustical features (i.e., SER_1, SER_2, SER_3) formed an output vector of features representing all 25.6ms voiced segments for each subject. The SER_4 was not taken into this comparative statistical classification due to the property of linear dependency among sub-bands over a whole 2.5HKz frequency range. Each output matrix contained N rows and M columns (N x M matrix), where N is a number of voiced speech frames with 25.6ms length/frame and M is a number of acoustical features. Mathematically, all SER parameters representing the suicidal, depressed, and remitted speech classes were defined into three large matrices. The parameter matrix representing each class were imported and implemented in matlab. In a state of classification of the between-groups were designed, (i.e. suicidal/depressed, depressed/remitted, remitted/suicidal). Classification accurate score and performance of selected classifiers were tested and evaluated with using the hold-one-out method, and 95% confidence interval were used in statistical analyses. The hold-one-out method was used in this discriminant analysis to compensate for the small size of speech databases used in this study. First, the input SER parameters were randomly selected for 35% of SER sample set to train classifiers, and 65% the rest of the same sample set to validate the same classifiers. The K-fold cross-validation on several trials on random selection of samples for training and testing approximately hundred times are employed for the average performance of classifications among several classifiers.

IV. Experimental Result and Discussion

The whole procedure of our proposed research study is depicted in figure 1. It clearly separates a whole research procedure into two main parts which are feature extraction and classification. Three categorized speech samples of remitted, depressed and high risk suicidal subjects were analyzed and compared for the characteristics of class-separation power in the spectral energy parameters that were extracted directly from the spectrum of speech signals. Figures 2 and 3 show band energies estimated by using two different analyzing windows with lengths of 25.6ms and 51.2ms over a 0-2.5KHz frequency range at which the most energy of speech signal concentrates. These two various analyzing windows were tested to see there is any significant difference appearing on the quantitative energy extracted based on these windows. The band energies are purposely estimated from four separated frequency bands equally divided with each bandwidth of 625Hz and stacking up to the upper frequency bound at 2.5KHz. Both mentioned plots have showed the same tendency of shifting band energies from band no.1 to band no.2 and continued consistently in order for higher frequency bands no.2, 3 and 4 respectively. As seen from plots, the normal speech group (remitted speech) has the highest energy level in frequency band no.1 compared to those respective lower energy level of depressed and suicidal speech groups. But in all higher frequency bands the energies of depressed and suicidal speech groups are inversely higher than that of normal speech group as lines switch around. It means that energy in depressed and suicidal speech starts shifting from band no. 2 through all higher bands. All of energy shifts in frequency bands suggested that the abnormal speech samples modulated with emotional illness such as depression and severely high-risk suicide have more power spectral density grown at the higher frequency bands. In this case the depressed and high-risk suicidal speech sample groups contained more energy concentration at higher frequencies above 625Hz or band no.2 when compared to normal speech group within the same frequency bands.

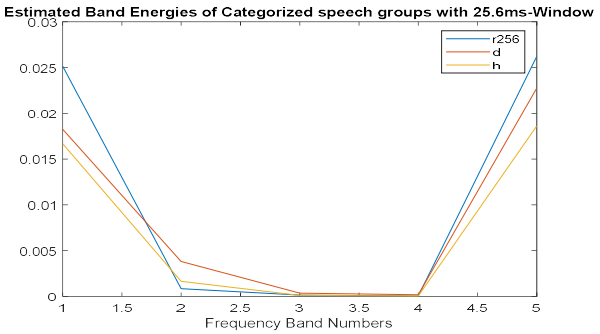


Fig. 2. Estimated band energies from 0 to 2.5KHz with 25.6ms-window

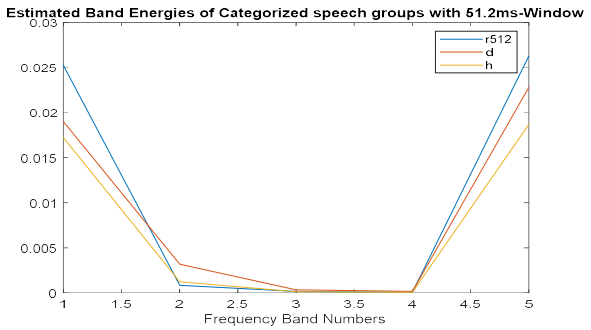


Fig. 3. Estimated band energies from 0 to 2.5KHz with 51.2ms-window

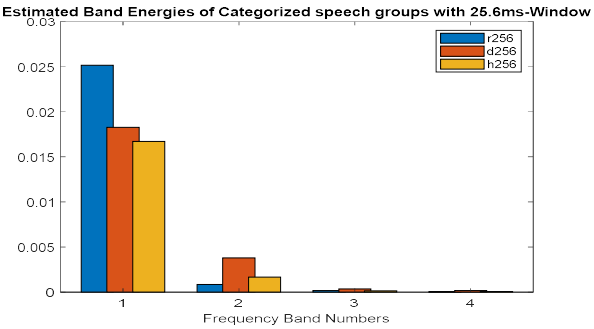


Fig. 4. Estimated band energies of categorized speech groups over a 0-2.5KHz range

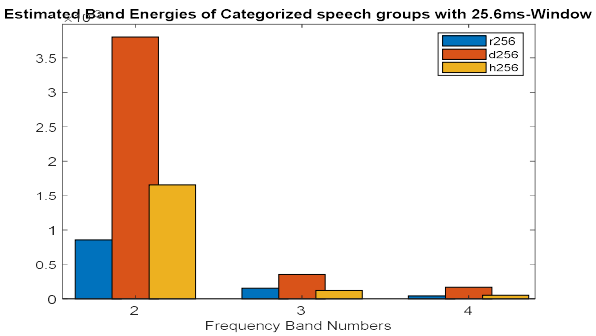


Fig. 5. Magnified plot of band energies over a 625Hz-2.5KHz range

In figures 4 and 5 all extracted energies were represented in bar plots which can provide the visual distinguishing observation on energy distributions among categorized speech sample groups. The significant difference in energy level can be clearly seen among them. It clearly shows the dominant energy found in depressed and high-risk suicidal speech at band no.2 or the higher frequency bands respective to normal case. Similar trends of energy shifting are pretty much the same as already discussed in aforementioned line plots in figure 2. Figure 6 displays the comparative tendencies of spectral energy change along frequency sub-bands over the entire frequency range of 0 -2.5KHz. The shifting among estimated energies of three categorized speech groups can be notified starting in band no. 2 which is the location of shifting occurred and stay the same above the energy level of remitted case for the higher bands no. 2 to 4 without any further shifting being found. Figure 7 shows the different distributions of energy in band no. 1 to 4 and most concentration of energy locates in band no. 1 and less concentrated energy can be notified in bands no. 2 to 4, respectively.

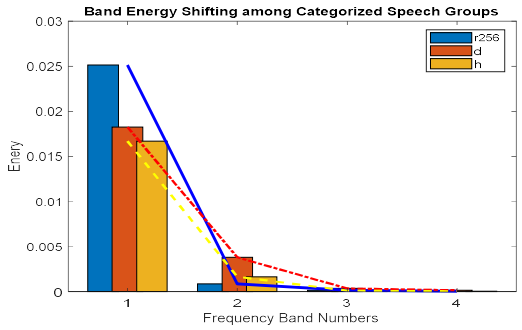


Fig. 6. Shifting in spectral energies among categorized speech groups

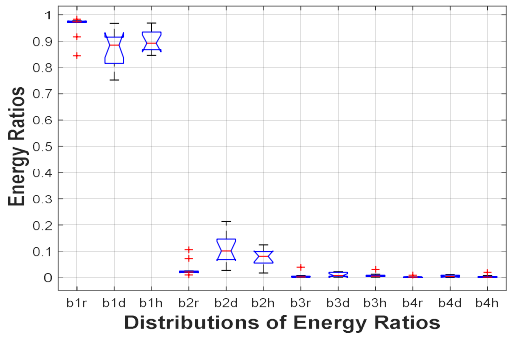


Fig. 7. Distribution of energy ratios among categorized speech groups

TABLE I. % ERROR IN CLASSIFICATION FROM TESTING

Classifiers	%Error in Classification Testing					
	Spectral Energy			Spectral Energy Ratios		
	Class1	Class2	Total	Class1	Class2	Total
LS	0.23	0.30	0.28	0.23	0.13	0.20
ML	0.22	0.39	0.31	0.12	0.32	0.22
SVM	0.36	0.65	0.50	0.26	0.30	0.28
RBN	0.14	0.68	0.39	0.13	0.31	0.22

TABLE II. % ERROR IN CLASSIFICATION FROM TRAINING

Classifiers	%Error in Classification Training					
	Spectral Energy			Spectral Energy Ratios		
	Class1	Class2	Total	Class1	Class2	Total
LS	0.18	0.25	0.18	0.15	0.08	0.10
ML	0.12	0.18	0.15	0.08	0.07	0.07
SVM	0.35	0.63	0.45	0.20	0.19	0.18
RBN	0.00	0.46	0.20	0.01	0.19	0.07

The performance measures summarized in Table I and II indicate that the combined SER features obtained from all categorized speech databases provide the most powerful discrimination in separating between depressed and remitted speech samples effectively in state of training and testing features with several classifiers. The highest correct classification scores from the pairwise study between remitted and depressed speech groups can be found for average value higher than 80% in accuracy. This average is the result of percentage calculated from 100% - % Error in classification using SERs as input parameters in training and testing classifiers shown in Tables I and II. This can be interpreted in that there is no shifting in energy occurred in any frequency bands between depressed and high-risk suicidal speech samples and make all energy ratios no conflict among input parameters to classification due to reversing quantity in parameters, but not for remitted speech samples which have some shift in energy appeared at frequency band no. 2. In order to continue further more on this study the larger collected database may improve the statistical interpretation on results to be more accurate assessment of acoustical characteristics which relate to the psychological states speaker experiences. In addition, the improvement of accurate classification can be achieved by using multi-parameter classifiers. Therefore, the further task will involve with the improved assessment of depression via effective speech processing techniques.

V. CONCLUSION

The spectral energy and energy ratios extracted from speech samples representing the vocal output characteristics related emotional illness are found to be most effective in differentiating between remitted and depressed speech samples. The shifting in energy was found in the higher frequency bands above band no. 1 for both depressed and suicidal speech samples. As result shown, the highest correct classification score was found to be approximately 80% in average among several trial of classifiers. It suggested that the studied feature in frequency domain can be taken in account of potential acoustics for assessment of depression and even high-risk suicide. The adaptation of speech analysis technique can provide an objective indication of mental illness states in terms of an assistive clinical assessment. Further testing of these similar acoustical parameters may produce reliable, clinically useful, adjunctive tools for assessments of depression and even suicidal risk.

REFERENCES

[1] H. Stassen, "Modeling affect in terms of speech parameters", Psychopathol., Vol. 21, pp. 83–88, 1988.

[2] D.J. France, R.G. Shiavi, S. Silverman, M. Silverman, and D.M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk", IEEE Trans. Biomed. Eng., 47(7):829-837, 2000.

[3] G. Fairbanks, Voice and Articulation Drillbook, Harper & Row, New York, 1960.
[4] A. Ozdas, R. G. Shiavi, D. M. Wilkes, M. K. Silverman and S. E. Silverman, "Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment", Methods of Information in Medicine, Vol. 43, pp. 36-38, 2004.
[5] A. Ozdas, R. G. Shiavi, D. M. Wilkes, M. K. Silverman and S. E. Silverman, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk", IEEE Trans. Biomed. Eng., Vol. 51, pp. 1530-1540, 2004.
[6] K. Scherer, Nonlinguistic Vocal Indicators of Emotion and Psychopathology, in C. E. Izard, ed., Emotions in Personality and Psychopathology, Plenum Press, New York, 1979, p 493-529.
[7] K. R. Scherer, Vocal correlates of emotional arousal and affective disturbance, in H. Wagner and A. Manstead, eds., Handbook of social psychophysiology, Wiley, New York, 1989.
[8] J. K. Darby, Speech and voice studies in psychiatric populations, in J. K. Darby, ed., Speech Evaluation in Psychiatry, Grune & Stratton, Inc., New York, 1981.
[9] K. R. Scherer, Speech and emotional states, in J. K. Darby, ed., Speech Evaluation in Psychiatry, Grune and Stratton, Inc., New York, 1981.
[10] K. R. Scherer, Vocal affect expression: A review and a model for future research, Psychological Bulletin, Vol. 99, pp. 143-165, 1986.
[11] E. Moore II, M. Clements, J. Peifert and L. Weissert, "Analysis of Prosodic Variation in Speech for Clinical Depression", the 25th Annual International Conference of the IEEE EMBS, Mexico, 2003.
[12] E. Moore, M. Clements, J. W. Peifer, and L. Weisser, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 96–107, 2008.
[13] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An investigation of depressed speech detection: features and normalization," in Proceedings of Interspeech, pp. 2997–3000, ISCA, Italy, August 2011.
[14] N. Cummins, J. Epps, V. Sethu, and J. Krajewski, "Variability compensation in small data: oversampled extraction of i-vectors for the classification of depressed speech," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2014), pp. 970–974, Italy, May 2014.
[15] L. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2010), pp. 5154–5157, USA, March 2010

Characterizing Depressive Related Speech with MFCC

Sirimonpak Suwannakhun
Media Technology Program
School of Architecture and Design
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
sirimonpak.sut@kmutt.ac.th

Thaweesak Yingthawornsuk
Media Technology Program
School of Architecture and Design
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
thaweesak.yin@kmutt.ac.th

***Abstract*— The experimental results from comparative study of acoustical properties in speech as emotional indicator based on spectral characteristics of speech signal have formerly been studied and reported for its quantitative information in association with the emotional states in persons suffering depression. This symptom affects speech production system of speaker, which modulates in spoken sound. MFCC has been reported for its characteristic change corresponding to severity of depression. The sixteenth MFCCs from remitted, depressed and suicidal patient groups were extracted, statistically tested and classified in pairwise fashion by using ML, LS and LMS classifiers. The best score of classification can be obtained at 0.2487 in error based on ML classifier with 80% of MFCC samples in testing phase. Results suggest the dominant property of MFCC in separation between suicidal and recovering speakers from depression.**

Keywords—speech, MFCC, depression, maximum likelihood

I. INTRODUCTION

Nowadays, it has been an increasing rate on population growth which is climbing up every year. But natural resources have inversely decreased regarding the daily basis consumption of world population. This could have impacted on our daily living in terms of less healthy life, more stressful working to earn income, poverty in lack of basis personal needs. In some serious situation this impact can cause a risk to someone' s life and steadily increasing that risk without any feeling to be aware of life threatening. This affective has been clinically known as depression or in some severe case this emotional disorder can result in suicidality. Many reports on media or social media can be seen about risk of suicide in depressed persons and suicide is popularly the public health problem associated with high population. And there is also an increasing rate of hotline call-in, which is simultaneously monitored by physicians or psychiatrist to evaluate the callers' behavior in speaking and their voice quality while having conversation on a phone line. This could be partially achieved on early decision made solely by highly experienced psychiatrist on judging that person is depressed or non-depressed. How accurate the diagnosis made by physician could impact on their family and lethal risk to that caller if he/she might be planning to commit the suicide.

Apparently, if psychiatrist can diagnose the symptom of depression or high-risk suicidality correctly, it could be high beneficial to patients who have agonized from emotional illness to be admitted to the health care program in time and have the precise treatments right away from the beginning state of depression. In previous works proposed, it has been shown that the acoustical features of human speech can be employed which are associated with recognizing pattern of emotional affection and prediction of the mental state in depressive speakers [1-3, 5-10]. The most common methods to assess, if patients were at severe state of depression or even at elevated risk of suicide, are self-scored patient survey, report by other, clinical interviews and rating scales [4]. Diagnosis and decision making on clinical categories of subjects belong to are clinical procedure with time consuming in which practitioners have to get involved in several steps such as information gathering, background profile checking, hospital admission and visiting records, diagnosing with simultaneous response in judging if patient were psychologically safe from suicidal risk or clinically identified for one of symptom categories, dramatically necessitates for physician to conclude the diagnosing result with the correct decision making on admission and treatment for that patient. As reported in the published studies, several analytical techniques have been proposed for achievement of measuring the particular changes, as a result of affection from the underlying symptom of depression, in acoustics of speech of depressed patients. It has been concluded that the suicidal speech in severely depressed speaker is very similar to that of common depressive one, but the tonal quality of speech significantly changes when the symptom of near-term suicidal risk highly strikes at the moment.

Sections in paper are organized with their details provided within each section. Section II relates works related in the past. Section III describes on method, database, feature extraction, PCA and classification. Section IV deals with experimental result and discussion. Section V provides conclusion and future research direction at the end

II. RELATED WORKS

Speech processing has been successfully adapted in validation of the correlation between acoustic features in speech and depression. The statistical analysis of several acoustic parameters demonstrated the significantly statistical difference in speech acoustical characteristics in between major depressed and non-depressed adolescents [12]. Sonawane et al. [13] and H. Aouani et al. [14] have used Frequency Cepstral Coefficients (MFCC) and subband based coefficients extracted from emotion sound database in multiple SVM classifiers and they discovered that the nonlinear kernel SVM and DSVM achieved the greater accuracy than linear SVM. M. N. Stolar et al. [15] presented the classification between depressed and non-depressed speech samples and used the 13 MFCC coefficients as feature vector in classifications with using both SVM and GMM. The MFCC worked well with the GMM classifier with 5 Gaussian Mixtures. Smitra et al. [16] extracted the 19 MFCC

coefficients and fed into ANN with varying numbers of neurons in the hidden layer to classify between healthy and pathological voices. The best accuracy of classification was obtained for the highest 99.96% based on the frame size of 512 datapoints. The various frame sizes at 256, 512, 1024 and so on were experimented to evaluate for the effectiveness of ANN by classification. It should be noted that the MFCC and extended derivatives (delta MFCC) were popularly employed as the acoustic speech features in most of these previous studies and also the speech databases studied and analyzed in aforementioned works were collected from participants who have their mother's tongue in western languages.

III. METHODOLOGY

A. Database

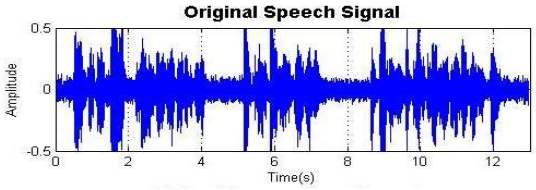
The database consists of speech samples recorded from interviewing session with psychiatrist. It is categorized into three groups of 10 remitted, depressed and high-risk suicidal female subjects. The pre-processing is carried out by first digitizing all speech signals through a 16-bit analog to digital converter at a sampling rate 10 KHz via a 5 KHz anti-aliasing low-pass filer. Prior to detection of voiced, unvoiced, silent segments in each speech files, the monitor and screening on any sound artifact possibly appeared during interviewing are offline implemented by using the Goldwave, including the silences longer than 0.5 seconds are manually removed. All speech signals of remitted, depressed and high-risk suicidal speakers are carefully processed under the same condition of pre-processing and the similar acoustical environment control is made during the period of recording speech sample in interviewing conversation.

B. Speech Segmentation

Based on the exploiting fact that the unvoiced segments of speech signal are very high frequency component compared to the voiced speech which is low frequency and quasi-periodic. To classify which segments of speech signal based on their energy and then weighted using the Dyadic Wavelet Transform (DWT) of speech samples were computed in each segment of 256 samples/frame. The unvoiced speech segments can be readily detected by comparing the energies of DWTs at the lowest scale $\delta_1 = 2^1$ and the highest energy level is $\delta_5 = 2^5$. Any segment of speech signal with its largest energy level estimated at scale $\delta_1 = 2^1$ is favorably classified as an unvoiced segment, otherwise found voiced segments. The following equation is the energy threshold defined as unvoiced segment;

$$UV = (n|\delta_i = 2^1); \quad n = 1, \ldots\ldots\ldots, N; \qquad (1)$$

where *uv* is speech segment classified as unvoiced at which the *n* segment with energy at scale δ_1 maximized.



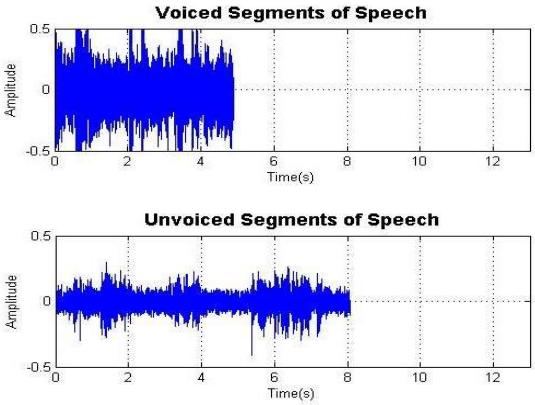


Fig. 1 Original speech signal (upper), voiced segment of speech (middle) and unvoiced segment (lower)

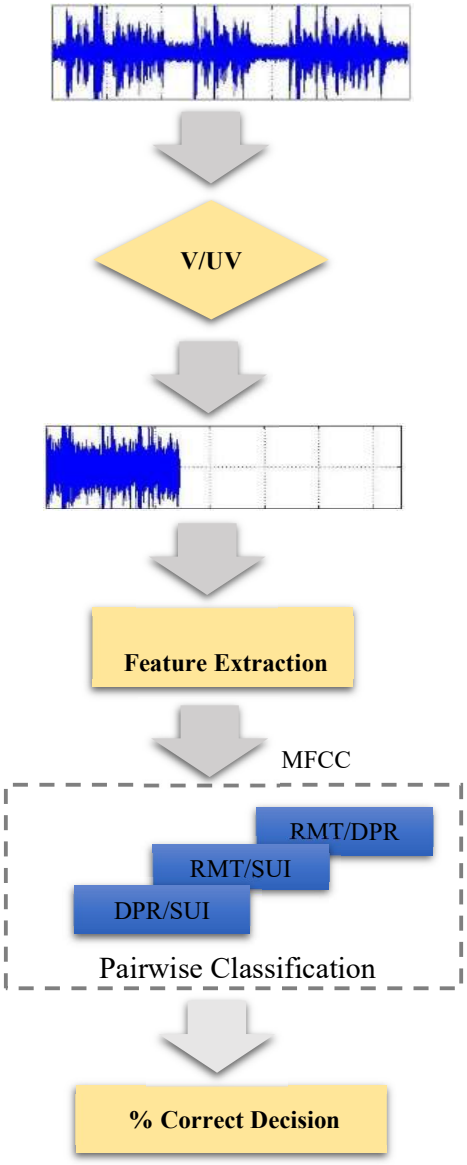


Fig. 2 Speech processing and classification procedure

C. MFCC Extraction

Voiced segments of all speech signals in database are processed for Mel-Scale Frequency Cepstral Coefficients (MFCC) [7-8,10]. The estimation procedure of studied energy parameter is described as follows;

- Windowing each concatenated voiced-segment into 25.6 ms-length frames.
- Computing the logarithm of the discrete Fourier Transform (DFT) for all windowed frames of voiced speech.
- Applying the log-magnitude spectrum through the 16 triangular bandpass filter bank with center frequencies corresponding to Mel-frequency scale.
- Computing the inverse discrete Fourier Transform (IDFT), then calculate the 16-order cepstral coefficients.
- Analyzing all extracted MFCC dataset with two dimensional PCA and then classifying with ML, LS and LMS classifiers.

The purpose of Mel-frequency scale is to map between linear to logarithmic scale for frequencies of speech signal higher than 1 kHz. The characteristics of spectral frequency will correspond to human auditory perception. The Mel-scale frequency mapping is defined [11]:

$$f_{mel} = 2595 * LOG_{10}\left[1 + \frac{f_{lin}}{700}\right]; \qquad (2)$$

in which f_{mel} is the perceived frequency and f_{lin} is the real linear frequency in speech signal.

In filtering phase, a series of the 16 triangular bandpass filters, N_s = 16 is used for a filter bank whose center frequencies and bandwidths are selected according to the Mel-scale. Once the center frequencies and bandwidths of the filter are obtained, the log-energy output of each filter *i is* computed and encoded to the MFCC by performing a Discrete Cosine Transform (DCT) defined as follow:

$$C_n = \frac{2}{N'}\sum_{i=1}^{N_f} x_k cos\left(k_i \frac{2\pi}{N'} n\right) \quad ; 1 \leq n \leq p \qquad (3)$$

Regarding less complexity, the factor $\frac{2}{N'}$ in equation 3 is discarded from algorithm computation.

D. Principal Component Analysis

The PCA technique has been applied to MFCC features to extract the most significant components of feature. This technique helps reduce multi-dimension of dataset down to two dimensions which is adequate for training and testing phases in classification.

E. Pairwise Classification

Several classifiers such as Maximum Likelihood (ML), Least Squares (LS) and Least Mean Squares (LMS) are selected to train and test on two dimensional MFCC dataset and compare among three different subject groups for performances of individual classification. In this study three groups of extracted MFCC samples are arranged into pairwise manners which are RMT/DPR, RMT/SUI and DPR/SUI. First, MFCC samples are randomly selected for 20% from sample dataset, and then used to train classifier, and other 35%, 40% from same dataset for training the same classifier. The reason of doing these is to compare the performances of classification among categorized subject groups, which might be affected from sizes of sample. Several trials on random selection of samples for training and testing approximately hundred times are further proceeded to find the average performance of classification.

IV. Experimental Result and Discussion

Original speech waveform, voiced and unvoiced segments of speech signal are plotted in Figure (1). The difference in amplitude and time interval can be obviously notified between voiced and unvoiced segments. Averaged errors in classification are tabulated in categorized pairwise groups versus types of classifier listed in Table 1-6 for case of 20%, 35%, 40% of training sample, summarized averages from RMT/SUI training as best pairwise with least error, 20% testing, and summarized averages from RMT/SUI testing.

The comparative errors obtained from several trials on selections of MFCC sample in classification are graphically depicted in Figures 3 and 4 for cases of training and testing with LS and ML classifiers. As seen in box-and-whisker diagrams, sampled MFCC represented as class 2 for suicidal group provided very less error of classification approximately 0.15 for all 20%, 35% and 40% of training samples and as well for both LS and ML classifiers. The greater errors can be seen for class1 represented for remitted group in training and testing for all classifiers and percentages of sampling approximately 0.35. More notification can be made on similar results of classification between two classifiers with different sampling percentages.

Based on the first four lower-order MFCC, the fairly high correct classifying scores can be obtained in this study, which are likely productive for its class discriminative property beneficial to emotional disorder assessment. More various acoustical parameters are suggested into the same account with studied MFCC for more accurate classification and improvement of research result toward same golden goals committed to research work.

TABLE 1
Errors of classification with 20% Training

	RMT/DPR	RMT/SUI	DPR/SUI
ML	0.4178	**0.2487**	0.2892
LS	0.3979	0.249	0.2755
LMS	0.49601	0.4857	0.4669

TABLE 2
Errors of classification with 35% Training

	RMT/DPR	RMT/SUI	DPR/SUI
ML	0.423	**0.2497**	0.2903
LS	0.3972	0.2498	0.2764
LMS	0.495	0.4908	0.4898

TABLE 3
Errors of classification with 40% Training

	RMT/DPR	RMT/SUI	DPR/SUI
ML	0.4213	**0.2498**	0.2897
LS	0.3975	0.2499	0.2764
LMS	0.5087	0.5304	0.492

TABLE 4
Summarized errors of classification between Remitted and High-Risk Suicidal

Classification	Percent of sample in training classifier		
	20%	35%	40%
ML	**0.2487**	0.2497	0.2498
LS	0.249	0.2498	0.2499
LMS	0.49061	0.4898	0.5304

TABLE 5
Errors of Classification with 20% Testing

	RMT/DPR	RMT/SUI	DPR/SUI
ML	0.4186	**0.2495**	0.2892
LS	0.3978	0.2498	0.249
LMS	0.4995	0.4849	0.4685

TABLE 6
Summarized errors of classification between Remitted and High-Risk Suicidal

Classification	Percent of sample in testing classifier		
	80%	65%	60%
ML	**0.2495**	0.2499	0.2498
LS	0.2498	0.25	0.25
LMS	0.4849	0.4911	0.5306

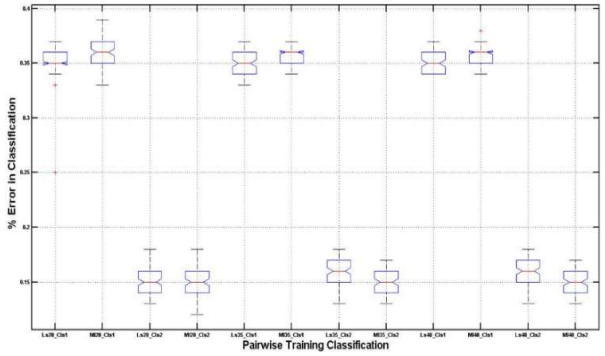


Fig. 3 Comparison of Box plots between LS and ML classification with 20%, 35% and 40% of samples in training phase

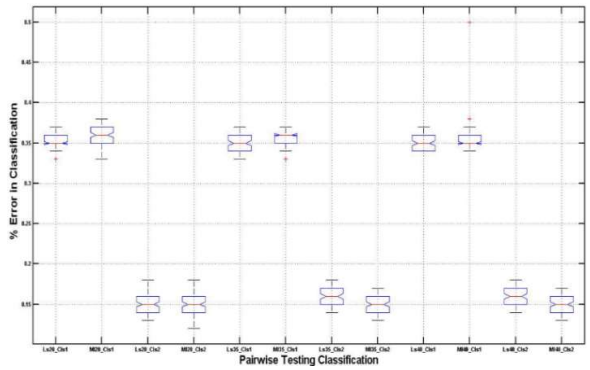


Fig. 4 Comparison of Box plots between LS and ML classification with 80%, 65% and 60% of samples in testing phase

V. Conclusion

Experimental results show the MFCC's property able to indicate speaker's psychiatric state, especially in class separation between remitted and suicidal speaker groups. Different sampling percentages investigated in this study can affect slightly the classification score in some classifiers selected to evaluate vocal samples. Further direction will focus on much more effective acoustics that can be assistive to currently studied MFCC in class separation with highly significantly statistical difference, and larger size of speech sample database required.

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R.

References

[1] T.Yingthawornsuk, "Comparative Study on Vocal Cepstral Emission of Clinical Depressed & Normal Speaker", Int'L Conf. on Control Automation & Systems, Korea, Oct. 26 -29, 2011.

[2] T.Yingthawornsuk et. al, "Comparative Study of Pairwise Classification by ML & NN on Unvoiced Segments in Speech Sample", Int'L Conf. On System & Electronic Engineering, Thailand, Dec. 18 -19, 2012.

[3] T.Yingthawornsuk, "Classification of Depressed Speakers Based on MFCC in Speech Sample", Int'L Conf. on Advances in Electrical & Electronics Engineering, Pattaya, Thailand, April 13 – 15, 2012.

[4] M. Hamilton, "A rating scale for depression", Journal of Neurology, Neurosurgery and Psychiatry, Vol. 23, pp. 56-62, 1960.

[5] France, D.J. et al., "Acoustical properties of speech as indicators of depression and suicide", IEEE transactions on BME, 2000. 47:p 829-837.

[6] F. Tolkmitt, H. Helfrich, R. Standke, K.R. Scherer, "Vocal Indicators of Psychiatric Treatment Effects in Depressives and Schizophrenics", J.Communication Disorders, Vol.15, pp.209-222, 1982.

[7] Godino-Llorente J.I. et al., "Dimensionality Reduction of a pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short Term Cepstral Parameters", IEEE Transaction on Biomedical Engineering, 53(10):1943-1953, 2006.

[8] Lu-Shih Alex Low et al., "Content Based Clinical Depression Detection in Adolescents", 17th EUSIPCO 2009, Scotland, Aug. 24-28, 2009.

[9] T. Yingthawornsuk, R.G. Shiavi, "Distinguishing Depression and Suicidal Risk in Men Using GMM Based Frequency Contents of Affective Vocal Tract Response", ", Int'L Conf. on Control, Automation and System 2008, Seoul, Korea, 2008.

[10] Ozdas, A., Shiavi, R.G., Wilkes, D.M., M. Silverman, and S. Silverman, "Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment", Meth. Info. Med., vol. 43, pp 36-38, 2004.

[11] Koeing, W., "A new frequency scale for acoustic measurements", Bell Telephone Laboratory Record", Vol. 27, pp. 299-301, 1949.

[12] A. Y. Hussenbocus et al., "Statistical Differences in Speech Acoustics of Major Depressed and Non-Depressed Adolescents", Int'L Conf. on Signal Processing and Communication Systems (ICSPCS 2015), Australia, Dec. 4-16, 2015.

[13] A. Sonawane et al.,"Sound based Human Emotion Recognition using MFCC & Multiple SVM", Int'L Conf. on Information, Communication, Istrument and Control, 2017.

[14] H. Aouani et al., "Emotion Recognition in Speech UsingvMFCC with SVM, DSVM andvAuto-encoder", Int'L Conf. on Advanced Technologies for Signal and Image Processing, Tunisia, 2018.

[15] M. N. Stolar et al., "Detection of Depression in Adolescents Based on Statistical Modeling of Emotional Influences in Parent-Adolescent Conversation", the 40th IEEE Int'L Conf. on Acoustics, Speech and Signal Processing (ICASSP 2015), Australia, April. 19– 24, 2015.

[16] Smitra, et al., "Classification of Healthy and Pathological Voices using MFCC and ANN", the 2nd Int'L Conf. on Advances in Electronics, Computer and Communications (ICAECC 2018), Bangalore, India, Feb. 9-10, 2018.

Creating Awareness of Incorrect English Pronunciation in Thai Elementary School Students using the Detect Me English, Natural Language Processing, Application

Ma'ayan Grace
Department of Business English, Faculty of Humanities and Social Sciences, Phetchabun Rajabhat University, Phetchabun, Thailand
maygraceabide@gmail.com

Jeerapan Phomprasert
Department of Business English, Faculty of Humanities and Social Sciences, Phetchabun Rajabhat University, Phetchabun, Thailand
Jeerapan72@pcru.ac.th

Abstract— This paper examines the ability of the Detect Me English application to create awareness of incorrect pronunciations in Thai Elementary Students (identified through Natural Language Processing). Phonetic analysis is a branch of NLP which analyses the sounds of human speech. The English correction software developed for this research uses an android system combined with the Detect Me English application to analyze the phonological sounds produced by the students. The researchers tested 60 students in total from grades 4-6. The results showed that the sample groups were not aware of many of the English phonological rules. Upon further analysis of the data recorded the post-test reveled that students were able to obtain significantly higher results.

Keywords — English pronunciation; English pronunciation awareness; English correction natural language processing software; Phonetics learning,

I. INTRODUCTION

Upon seeing Stephen Hawking's speech synthesizer in action many of us were amazed with the machines ability to convert text to speech. The 'EZ keys' program used could automatically scan each word row and column then produce the respective sounds. The incredible step forward in the field of NLP meant a computer had the ability to process language to the point of giving a voice to a once voiceless a man. NLP is broken up into many different fields. This research focuses particularly on computational phonetics and phonology of which there are two categories. Hawking's machine is an example of speech synthesis, the process of creating electronic signals which can then initiate the output of audio through an electronic speaker [1]. This output will vary according to the phones and prosodic features present in the original text. The other category is speech recognition, where a computer must analyze the phones created and transcribe them into a phonetic transcription in order to generate ordinary text which can then be displayed on the screen. The practical implementation of NLP brought about the 'Detect Me English' application. The application has the ability to transform a regular android device into a device which can both synthesize and recognize speech output.

Thai curriculum requires that English be taught from kindergarten level all the way through to the final year of high school as a fundamental subject. Even so, many students are unable to speak or use English to convey ideas effectively. Ref. [2] says that the key to good speaking begins with good pronunciation. Multiple studies have revealed that the majority of Thai student's pronunciation errors originate from the fact that they often attempt to pronounce English sounds as by using the familiar Thai consonants [3,4,5,6]. The problem is that the phones and phonemes present in Thai language are different from that of English which means that Thai students are prone to mistakes, especially in their pronunciation. The foundations of the Contrastive Analysis Hypothesis show that the elements which are analogous to the learner's native language will be simple for him, and those elements that are dissimilar will be more challenging [7].

Recognizing the difficulty Thai teachers have in teaching phonetics and the lack of native language speakers in many of Thailand's schools, the Detect Me English application was developed. This research was then run to see its effectiveness in facilitating Thai elementary student's English pronunciation acquisition and to measure the effectiveness of NLP in creating awareness of correct English pronunciation.

II. LITERATURE REVIEW

A. Language Transfer

Learning to pronounce English correctly is difficult for ESL students. One of the factors that cause difficulties is the difference between the phonology of their native language (L1) and that of the second language (L2). L2 learners will rely on the knowledge of their first language and their prior experience to deal with L2 problems. So, L1 transfer becomes inevitable in L2 learning. Ref. [8] proposes six factors that may hinder or facilitate learner's pronunciation of L2 are; language transfer, age, exposure to L2, innate phonetic ability, identity and language ego for L1, motivation, and finally the concern for good pronunciation ability. Second language learners may stop short of L2 competence due to fossilization which is the relatively permanent incorporation of incorrect linguistic forms of which are often transferred from L1[9].

B. Contransitive Analysis

In total there are 21 Thai consonant phonemes whereas the number of English phonemes is generally listed as 24. Both languages have phonemes which are not found in the other language or can be found but are used in certain environments which differ from language to language. The English sounds which are not in Thai phonology include /g/, /v/, /z/, /ð/, /ʒ/, /θ/, /ʃ/, /tʃ / and /dʒ/.

The voiced velar plosive /g/ sound is often pronounced /k/ which is a voiceless velar plosive sound [10]. The places of articulation are still the same however the sound changes from voiced to voiceless simply because the /g/ sound does not exist in Thai phonology. English voiced fricatives like /v/, /z/, /ð/ and /ʒ/ are often replaced with other consonant sounds that the speaker thinks best resembles the target sound.

Examples

/v/ becomes /w/

/z/ becomes /s/

/ð/ becomes /d/

/ʒ/ becomes /t/

In linguistics when talking about different environments it is referring to aspects like syllable position. For example all English voiceless fricatives positioned as final sounds are very hard for Thai L2 learners to master; /f/, /s/, /θ/ and /ʃ/. This is due to the fact that of the final plosive and nasal sounds in Thai language there are only four plosive sounds (/p/, /t/, /k/ and /ʔ/) and three nasal sounds (/m/, /n/ and /ŋ/) which are as final sounds, all of which are pronounced inaudibly [11,12]. Thai L2 learners will either replace these voiceless fricatives with a familiar Thai final sound or omit them entirely [13].

With further contrastive analysis many more differences can be found as well as curious occurrences where both languages actually have the sounds needed for correct pronunciation yet for Thai L2 learners they have difficulty pronouncing words with /l/ and /r/ sounds. Ref. [14] suggests that this is most likely due to behaviorism in that Thai's use the two sounds interchangeably when speaking Thai which causes the habit to be transferred when they attempt to pronounce English words.

C. *Awareness of Correct English Pronunciation*

Motivation in learning is the strongest factor which contributes to student's success in second language acquisition [15]. The are many things that can cause demotivation for example when an L2 learner attempts to talk to a native speaker only to find out that they don't understand them or when a student finds out they have been pronouncing a word incorrectly all along. This might seem like something that is inevitably going to happen when learning a second language however; there are ways to prevent mispronunciation. That is why it is so important to generate an awareness of phonological rules and correct pronunciation early on in language instruction. If teachers wait too long until the point that fossilization occurs, even with considerable input and instruction, L2 learners will not be able to acquire the challenging sounds which do not exist in Thai phonology.

III. Methodology

A. *Objectives*

Given the difficulty Thai L2 learners have in creating English sounds correctly the main objective of this research was to test the Detect Me English application's effectiveness in creating awareness of correct English phonetic pronunciations. Further objectives were to measure the improvement of the students after being first exposed to the application and being instructed on the correct pronunciation so as to evaluate the efficiency of the application becoming a teaching tool in Thai classrooms.

B. *Sample Group and Instruments*

The sample group used for this study comprised of elementary students from Nong Mae Na School (N=60). There were 20 students taken from grades 4, 5 and 6 accordingly. The school selected for the trail was a rural school with no native language teachers so all English instruction was conducted by Thai teachers. The instruments used were the Detect Me English application as well as the 30 words taken from the English Standard Curriculum and inputted into the application; 10 grade 4 level words, 10 grade 5 level words and 10 grade 6 level words.

C. *Application Design*

The Detect Me English application is compatible with android systems as seen in figure 1.

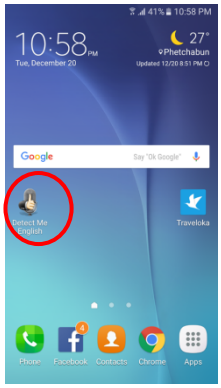


Fig. 1. Application will show up on the main screen, click to open up the main page.

From the homepage (Figure 2) there are two menus to choose from. The START menu links the user to the 30 key words selected for this research.

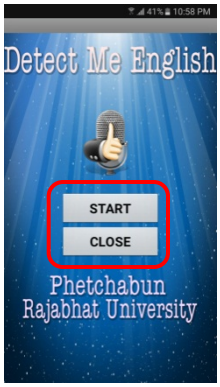


Fig.2 Home page of the Detect Me English application.

Once the program is running the user will see the main screen hub (figure 3). The hub which has functions and displays as follows

1. Microphone Button; this button is pressed when the user is ready to attempt to pronunce the subject word.
2. The Subject Word: where the English word the user needs to pronunce is displayed.
3. Results Box: where the subject word will display upon the user making the correct pronunciation of the word in English. If there is incorrect

pronunciation then the display will not show anything.

4. Incorrect Pronuncation Results Box: pon the user making an incorrect pronunciation, the applicaiton will process the users attempt and display phonetically what the user has pronunced, thus allowing the user to distingush what kind of phonetics they are incorrectly pronuncing.
5. The Demo Button: before the user tests themselves they can listen to the correct pronunciaton by selecting the demo button.
6. 6. Back Button: The user can skip back to the previous word.
7. 7. Next Button. The user can skip to the next word (figure 3.)

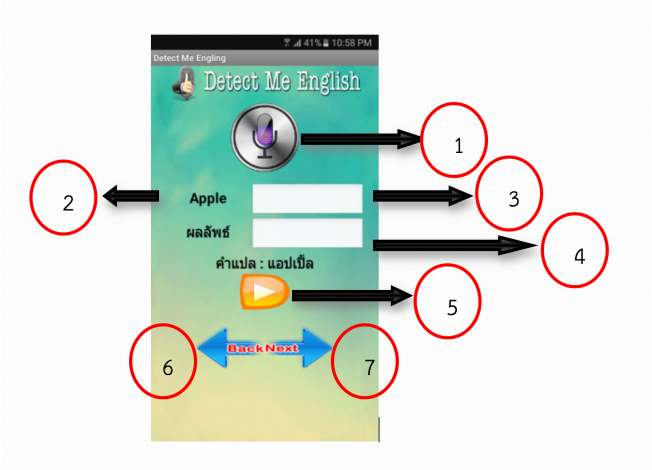


Fig.3. Main on-screen hub for the Detect Me English application

The 30 key words that the sample group worked through for this trail were taken from three different levels; grade 4, 5 and 6 level. Each of these words were chosen based upon the suitablilty to address address the L1 to L2 phonetical barrier as well as their suitability for primary school ages. Each word written into the program can be easily changed as the number of words is not limited and new words can be added or taken away from the library catalogue at any time. The strength of this program is the straight forwardness for teachers to be able to add or change words as desired (see figure 4). The program is also connected with an external hardware which collects the results of the user.

Fig.4. Examples of other word options

D. *Software Protocal Processing*

The recognition software functions by first receiving the input sound through the devices microphone. The information is then processed into its component phones and phonemes and converted into ordinary text which is then displayed in either one of the results boxes depending, of course, on the user pronunciation. The application uses the process of speech synthesis as well as speech recognition in that any words that are inputted into the program can generate for the user to attempt to emulate and any words in the system can also be tested for correct pronunciation by simply having the user speak into the android device's microphone.

E. *Data Collection*

Quantitative data were collected for this study. For both the pre-test and post-test each student was given 3 opportunities to pronounce the 30 selected key words. After the pre-test on the first day, the sample group was given instruction from Thai University students who had been exposed to and using the application already. Educational activities and games both using the Detect Me English application and not using it were devised by the students with the aim to generate a greater awareness of the correct pronunciations. At the end of the second day the sample group was tested again using the same method to see if there was any improvement. The data from both pre-test and post-test were then collected and analyzed by the researchers.

IV. Results

The results showed that the sample group was not aware of many of the English phonological rules. In the pre-test a majority of the words were mispronounced and while in the post-test the sample group's pronunciation improved significantly, the level of correct pronunciations was still low with the overall percentage of correct pronunciations less than 50% for all grades (figure 5).

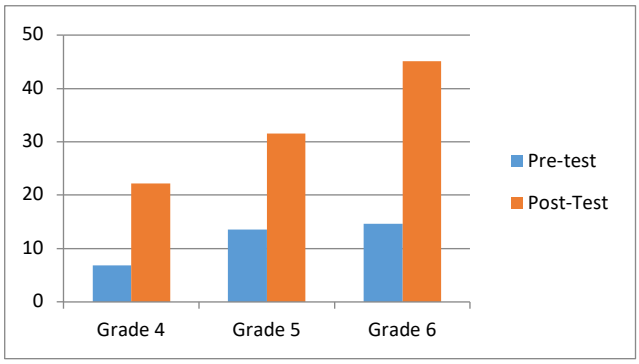


Fig. 5 Graph showing the overall percentage of correct pronunciations according to the Grade levels.

V. Limitations

The limitations of the Detect Me English application, as with most speech synthesizing devices, the sounds generated have a very machine like quality or accent. So, while the students are able to correct their pronunciation with the device suprasegmental factors, like intonation and stress patterns, can't be addressed. Another limitation came from the researches design in that while the application shows which phonemes the user is failing to create accurately the

data recorded by the application simply showed which words were pronounced correctly or incorrectly. This meant that analysis of which phonological rules and pronunciations were difficult or unknown to the sample group was impossible.

VI. Discussion and Conclusion

This research investigated the effectiveness of the Detect Me English application in facilitating Thai elementary student's English pronunciation. Teachers are always looking for new ways to motivate students to learn. The results from this investigation show that the application was an effective way not only to test student's pronunciation but encourage and motivate students to achieve higher academic results. Moreover, the practicality of the device in that it can be used by anyone, teachers or students shows its effectiveness in equipping teachers to be able to create awareness of incorrect pronunciations as well as enable students to self-correct.

The second aim of this research was to measure the improvement of the students after being first exposed to the application so as to evaluate the efficiency of the application becoming a teaching tool in Thai classrooms. The pre-test results showed that the sample group initially was unaware of many of the phonological rules in English. The post-test results however showed that after just two days of the students being exposed to the application each grade showed immense improvement. It is unrealistic to expect every Thai teacher to acquire native like English pronunciation however, this research reveals the need to provide tools for teachers to generate an awareness of the correct pronunciations with the added benefit of not having to spend millions of dollars on training. NLP applications can empower Thai teachers to teach correct English pronunciations confidently. As reveled from the post-test results the Detect Me English application, as with many other NLP systems, is an effective way to aid English education. While the results of the sample groups were still low it was apparent that over the course of two days using the Detect Me English application improved the sample group's percentage of correct pronunciations. It is worth considering how much greater the improvement could be if NLP applications like Detect Me English are used continually throughout the course of a child's education.

Acknowledgment

The results of this research as well as the Detect Me English application were offered in appreciation to Nong Mae Na School to aid the teachers in improving both the teacher's and the student's pronunciation ability and to create an awareness of incorrect pronunciations. Thank you to Phetchabun Rajabhat University for supporting this research so it could take place.

References

[1] Fromkin, V., Rodman, R., Hyams, N., Amberber, M., Cox, F., & Thornton, R. An Introduction to Language. (9th International ed.).South Melbourne, Victoria: Cengage Learning, 2011, pp.392

[2] Ellis, R. Second language acquisition. Oxford: Oxford University Press, 1997.

[3] Lakhawatana, P. A constructive study of English and Thai. Bangkok: The English language center, 1969.

[4] Chanyasupab, T. An analysis of English pronunciation of English major students at higher certificate of education level. M.A. Thesis, Chulalongkorn University, Thailand, 1982.

[5] Malarak, P. The variation of pronunciation of /s/ at the final position, when reading messages of middle school students. M.A. Thesis, Chulalongkorn University, Thailand, 1998.

[6] Mano-im, R. The pronunciation of English final consonant clusters by Thais. M.A. Thesis, Chulalongkorn University, Thailand, 1999).

[7] Lado, R. Linguistics across cultures: applied linguistics and language teachers. University of Michigan Press, Ann Arbor, 1957.

[8] Brown, D. H. Principles of language learning & teaching. (4th ed.). New York:Longman, 2000, pp. 49-58.

[9] Selinker, L. Interlanguage. product information international review of applied linguistics in language teaching, vol.10, pp. 209-241, 1972.

[10] Bowman, M.. A contrastive analysis of English and Thai and its practical application for teaching English pronunciation. The English Teacher, Vol. 4 (1), pp. 40-53, October 2000.

[11] [Abramson, A. S. Word-final stops in Thai. In J. G. Harris & R. B. Noss (Eds.), Tai Phonetics and Phonology Bangkok: Central Institute of English Language. 1972, pp. 1-7.

[12] Tuapharoen, P. (1990). Phonetics and practical phonetics. Bangkok; Thammasat University Press.

[13] Sahatsathatsana, S. Pronunciation problems of Thai students learning English phonetics: a case study at Kalasin University. Journal of Education Mahasarakham University, Vol. 11(4), pp. 67-84, 2017.

[14] Ellis, R. Second language acquisition. Oxford: Oxford University Press, 1997.

[15] Dornyei, Z. Motivational strategies in the language classroom. Cambridge: Cambridge University Press, 2001.

Effective Face Verification Systems Based on the Histogram of Oriented Gradients and Deep Learning Techniques

Sawitree Khunthi
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham, Thailand
sawitri0212@gmail.com

Pichada Saichua
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham, Thailand
pichadakt@gmail.com

Olarik Surinta
Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology, Faculty of Informatics, Mahasarakham University
Maha Sarakham, Thailand
olarik.s@msu.ac.th

***Abstract*—In this paper, we proposed a face verification method. We experiment with a histogram of oriented gradients description combined with the linear support vector machine (HOG+SVM) as for the face detection. Subsequently, we applied a deep learning method called ResNet-50 architecture in face verification. We evaluate the performance of the face verification system on three well-known face datasets (BioID, FERET, and ColorFERET). The experimental results are divided into two parts; face detection and face verification. First, the result shows that the HOG+SVM performs very well on the face detection part and without errors being detected. Second, The ResNet-50 and FaceNet architectures perform best and obtain 100% accuracy on the BioID and FERET dataset. They also, achieved very high accuracy on ColorFERET dataset.**

Keywords— face verification systems, face detection, face verification, ResNet-50, FaceNet

I. Introduction

Face verification is part of the face recognition system that focuses on the one-to-one matching problem [1] to compare whether it is the same person or not the same person. For this reason, face verification is much used in security, surveillance, and immigration, for example, to search for people from closed circuit television (CCTV) or to check if the person is a criminal by comparison of a face captured on camera with faces from a database. Many problems, such as images, low-light images, blurred image, and flare on an image resulting from stray light entering the camera lens, will occur depending on the quality and location of the camera. These effects are of concern for the researchers working on face recognition.

Face verification systems perform two main tasks. The first task is face detection and is essential to any face verification system because the system cannot process if the face is not detected. Many researchers focus on developing algorithms for face detection such as edge detection [2], Haar-cascade classifier [3][4], and histogram of oriented gradients (HOG) [5–7]. These algorithms allow us to find faces even in low-light and blurred images. Moreover, convolutional neural networks (CNNs) that have been proposed [8][9] provide a robust method to detect a face in many conditions such as a small faces, occlusion, or images that do not show the entire face.

The second task of face verification, is the extraction of information from the face (called face encoding) which is sent to the similarity function to calculate and compare the unknown face and detected face. A high similarity value shows that the two faces are the most similar face. Many algorithms have been proposed for the face encoding such as local directional number pattern [10], local binary patterns [11], common encoding feature discriminant [12] and supervised feature encoding [13] are proposed. Nowadays, deep learning approaches are successful in encoding the face, including VGGNet [14], DeepFace [15], FaceNet [16] and ResNet [17].

Contribution: In this paper, we evaluate the performance of face verification systems on three well-known face datasets (BioID, FERET, and ColorFERET). It is quite challenging to verify faces from the ColorFERET because this dataset consists of 3,553 face images of 474 subjects. We divided the experiment into two parts; face detection and face verification. In the face detection part, four different face techniques, including the histogram of oriented gradients combined with the linear support vector machine (HOG+SVM), max-margin object detection with convolutional neural network (MMOD-CNN) [18][19], Haar-Cascade Classifier [20][21] and Faced techniques were evaluated on the BioID dataset. The experiments showed that the HOG+SVM performs very well and without errors of face detection. Moreover, in the face verification part, three robust deep CNN architectures called VGG16, FaceNet, and ResNet-50 architectures were used as the face encoding. The experimental results showed that the ResNet-50 and FaceNet performed best and obtained 100% accuracy on the BioID and FERET dataset. Additionally, both architectures achieved very high accuracy on the ColorFERET dataset.

Paper outline: This paper is organized as follows: In Section II, the face verification systems are described in detail. In Section III, three well-known face image datasets are explained. The experimental results of face detection and verification are presented in Section IV. The last section is the conclusion and suggestions for future work.

II. Face Verification Systems

In the following, we describe the face verification systems used in the experiments; the histogram of oriented gradients and linear support vector machine aimed for face detection. Two face encoding methods; FaceNet and ResNet-50, are computed.

A. Face Detection

For face detection, the Viola-Jones face detector [20][21] is a well-known method that was first proposed for object and

then for pedestrian detection. Nowadays, this technique, called Haar-cascade classifier, has become a standard technique for face detection. The Viola-Jones face detector computes feature vector based on the Haar feature. It calculates from the rectangle detector or sub-window. The detector scans through the image. Then, the set of the feature vector is given to the AdaBoost classifier, which is the weak classifier. This approach can process in real-time and get high precision. However, this approach performs not very well on the BioID dataset.

We proposed to use the histogram of oriented gradients and the linear support vector machine, called HOG+SVM, in face detection experiments.

First, the well-known HOG [22] is proposed to compute a feature vector from sub-images that scans over the whole image. With this method, the oriented gradients are computed using a gradient detector. Then the oriented gradients of each sub-image are weight to the orientation bins and used as a feature vector [23]. The gradient detector is calculated as follows:

$$G_x = I(x+1, y) - I(x-1, y) \quad (1)$$

$$G_y = I(x, y+1) - I(x, y-1) \quad (2)$$

where G_x is the horizontal and G_y is the vertical components of the gradients.

The gradient magnitude (M) and the oriented gradients (θ) are computed as:

$$M(x, y) = \sqrt{\left(G_x^2 + G_y^2\right)} \quad (3)$$

$$\theta(x, y) = tan^{-1} \frac{G_y}{G_x} d \quad (4)$$

where $M(x, y)$ is the gradient magnitude and $\theta(x, y)$ is the orientation of the gradients at the location (x, y).

Consequently, orientation bins are selected based on oriented gradients. The gradient magnitudes for each oriented gradient are weight and summed up to each orientation bin. Then, the orientation bins for each sub-image are normalized using the L2 normalization.

Second, the support vector machine (SVM) [24] algorithm with a linear kernel is proposed in this paper due to the two-class classification. With the SVM algorithm, the hyperplane, which is the maximum distance to the training points, is used to separate training data. The training points that are closest to the calculated separating hyperplane are called support vectors. So, the best hyperplane is the distance between the closest data points of both classes and the hyperplane [25]. The optimal hyperplane is calculated as;

$$g(x) = W^T X + b \quad (5)$$

where W is the weight vector and b is the bias. The decision rule is

$$y = \begin{cases} 1 \; if g(x) > 0 \\ 0 \; otherwise \end{cases} \quad (6)$$

B. Face Encoding

In this research, two deep learning architectures for face encoding; ResNet-50 and FaceNet are proposed as the face encoding.

1) ResNet-50

The residual network architecture, which is a very deep network, was invented by He et al. [27], called ResNet architecture. The deep residual network creates simple stack layers, therefore the network can be set up as 18, 34, 50, 101, and 152-layer. This architecture is quite different from the original convolutional neural network (CNN) that each layer feedforward to the next layer. A deep residual learning block is implemented in the ResNet architecture (see Fig. 1). Hence, each layer allows to feed the output to feed into the next layer and directly into the next 2-3 forward blocks. This architecture known as shortcut connections.

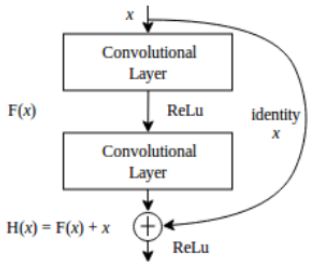


Fig. *1*. The residual network [26].

In this paper, we applied ResNet architecture with 50 layers for the face encoding (called ResNet-50). The lower-level features, which are more specific to the training data, are extracted from the face image. To encode a feature vector; we applied the flatten after the average pooling layer, which is the last layer of the ResNet-50. This architecture encodes 2,048 features and uses them as a feature vector.

1) FaceNet

FaceNet architecture was invented by Schroff et al. [16] to solve the problem of face recognition and clustering. This architecture is invariant to illumination and pose. Firstly, in this technique, the deep CNN architecture, which is inspired by Inception network, is used as a black box. The size of the parameters in FaceNet architecture is 7.5M. The small mini-batch size of around 40 faces per identity (in total, around 1,800 examples) are fed to the deep CNN. These direct to increase convergence while optimizing the network with Stochastic Gradient Descent (SGD).

Secondly, the output from the deep CNN architecture is normalized using L2 normalization and sent to the face embedding process. The embedding process is embeds in a face image into a dimensional space using the Euclidean function. This method guarantees the identity that the face image of person A is closer to other face images of the person A than closer to other face images of other persons.

Finally, the triplet selection is the last process of FaceNet. This process is given the face image of person A to compare other face images from the mini-batch to avoid poor training.

From this process, two parameters are selected, argmax and argmin, which are the hardest positive image of the same person and the hardest negative image of a different person, are selected.

In this paper, we applied FaceNet architecture using Inception network as the core network. This architecture encodes 512 features and used as a feature vector.

III. Face Image Datasets

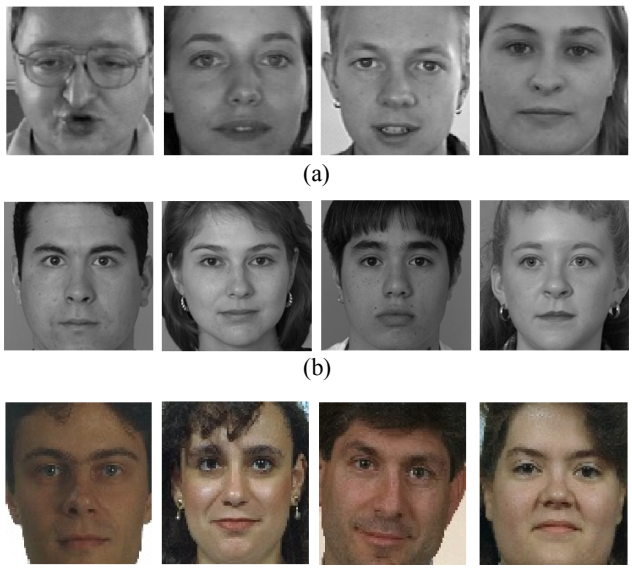
Many face image datasets were invented for face verification systems. In this paper, we select three face image datasets; the BioID, FERET, and ColorFERET dataset for evaluating the face detection and face verification.

A. BioID Face Dataset

The BioID face dataset used in the face detection experiment includes 1,513 frontal view images [27]. In this dataset, the image resolution is 384x286 pixels and stored on the grey level. Additionally, the number of people (subject) used in the face verification experiment is 21 subjects from 1,507 face images. The BioID dataset is shown in Fig. 2(a).

B. FERET and ColorFERET Datasets

The face recognition technology (FERET) dataset and ColorFERET were published in 1993 by J. Phillips and P. Rauss [15-16]. These datasets consist of 1,199 subjects, and the total number of the face images is 14,126 images with an image resolution of 384x256 pixels. In our experiments, we have used the FERET and ColorFERET for face verification. As for the FERET dataset. We selected 1,372 images from 196 subjects from the FERET dataset (See Fig. 2(b)). and 3,553 images from 474 subjects from the ColorFERET dataset (Fig. 2(c)).



(a)

(b)

(c)

Fig. 2. Sample of face images in the (a) BioID, (b) FERET, and (c) ColorFERET datasets.

IV. Experimental Results

A. Evaluation Methods

We have used two methods to evaluate the face verification system. The first evaluation method is face detection accuracy which is given by:

$$Accuracy = Acc - Err \tag{1}$$

where

$$Acc = \frac{c*100}{N} \tag{2}$$

$$Err = \frac{e*100}{N} \tag{3}$$

where c is the number of the face images after applying face detection method, and e is the number of the error face images N is the total number of the face images of the face dataset.

The second method is the accuracy of face verification.

1) We used the cosine similarity function to compare a feature vector extracted from the face image. The most similarity face is given the highest value. Then the correct prediction is that if the label of the highest value is the same as the test image. The cosine similarity function is computed as follows:

$$cos(\theta) = \frac{A.B}{\|A\|\|B\|} \tag{4}$$

where $A.B$ is the dot product of feature vector A and B.

2) To calculate the accuracy, the total number of correct predictions is multiplied by 100 and then divided by the total number of faces in the dataset.

B. Results

In this section, we show the experimental results of face detection techniques and face verification accuracies of CNN face encoding architectures.

1) Face Detection Results

To illustrate the results of face detection, Fig. 3(a) shows face images cropped so as to leave the entire face visible and Fig. 3(b) shows error due to poor cropping that results in the face being only partly visible. In this paper, when calculating the accuracy of the face detection method, we carefully reject the error face images by calculating the error (Err), as shown in Equation 3.

Table I show the experimental results of four different face detection techniques; HOG+SVM, MMOD-CNN, Haar-Cascade, and Faced techniques. Here, the histogram of oriented gradient combined with the linear support vector machine (HOG+SVM) is the only one face detection method that detects face without any error. The performance of HOG+SVM technique obtained on the BioID face dataset is 99.60%. The accuracy obtained from all face detection techniques was over 90%, except for the Faced technique. The face detection results are shown in Fig. 4.

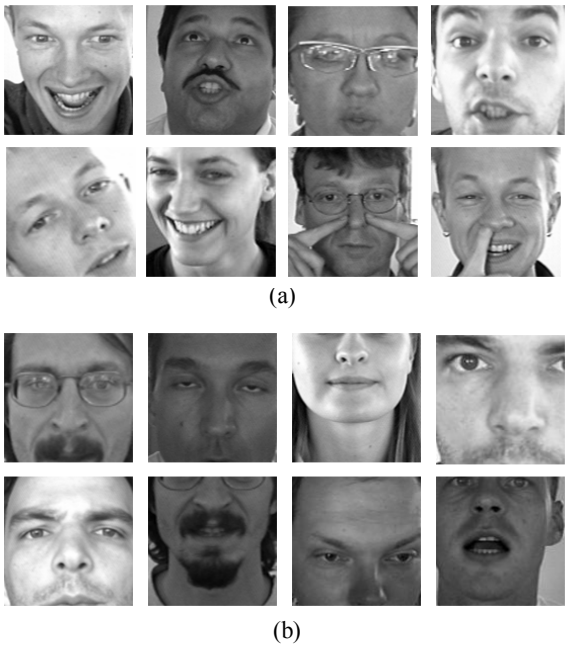


Fig. 3. Sample results of the face images after applying face detection method. (a) entire faces and (b) error faces.

Fig. 4. Face detection results after applying face detection techniques. (a) BioID images, (b) HOG+SVM, (c) MMOD-CNN, (d) Haar-Cascade, and (e) Faced techniques.

TABLE I. PERFORMANCE OF FACE DETECTION TECHNIQUES ON BIOID DATASET

Methods	Number of face detected	Number of error detected	Accuracy (%)
HOG+SVM	1,507	0	**99.60**
MMOD-CNN	1,513	40	97.36
Haar-Cascade	1,459	40	93.79
Faced	1,449	107	88.70

2) Face Verification Results

For the face encoding techniques, we evaluated the performance of three deep convolutional neural networks, including VGG16, FaceNet, and ResNet-50. The image resolution used in the experiments was 224x224 pixels. In the experiments, the VGG16 extracts the highest feature dimension with 25,088 features, followed by ResNet-50 and FaceNet architectures. The image resolution and size of the feature vector are shown in Table II.

In this paper we found that HOG+SVM was the best face detection method based on our experiments on the BioID dataset. We then chose the HOG+SVM method for detecting faces from three face datasets; BioID, FERET, and ColorFERET. As a result, the number of face images detects from the BioID, FERET, and ColorFERET were 1,507, 1,372, 3,553 face images, respectively. This was quite challenging because of the number of subjects in the ColorFERET (474 subjects) was 20 times higher than in the BioID dataset (only 21 subjects). The number of face images and the number of subjects are shown in Table III.

TABLE II. THE RESOLUTION OF FACE IMAGES REQUIRES FOR CNN METHODS AND THE NUMBER OF FEATURES EXTRACTS FROM THREE CNN FACE ENCODING TECHNIQUES

Parameters	Method		
	VGG16	**FaceNet**	**ResNet-50**
Image resolution	224x224	224x224	224x224
Feature vector	25,088	512	2,048

TABLE III. FACE VERIFICATION ACCURACIES (%) AND STANDARD DEVIATIONS OF THREE CNN FEATURE EXTRACTION METHODS. THE EXPERIMENTAL RESULTS ARE COMPUTED USING THREE FACE DATASETS

Dataset	Number of image	Number of subjects	Accuracy (%)		
			Vgg16	**FaceNet**	**ResNet-50**
BioID	1,507	21	99.74±0.38	100	100
FERET	1,372	196	83.93±0.77	100	100
Color FERET	3,553	474	74.96±1.26	99.32±0.32	99.60±0.46

In this paper, five random fold cross-validations are applied to evaluate the performance of the different face encoding methods. In our experiments, the best deep convolutional neural network (CNN) architecture for face encoding was ResNet-50 and FaceNet architectures because these two architectures obtain an accuracy of 100% on BioID and FERET face datasets. We particularly note that ResNet-50 outperforms other deep CNN architectures when experimenting on the ColorFERET dataset which consists of 3,553 face images with 474 subjects. The ResNet-50 and Facenet architectures had highly accuracies of 99.60% and 99.32%, respectively.

II. CONCLUSION

The key factor in achieving the highest accuracy in face verification systems consists of face detection and the face encoding process. In this paper, we have presented an

effective face verification systems. First, the histogram of oriented gradients method combined with the linear support vector machine (HOG+SVM) was applied as the face detection process. The experimental results showed that the HOG+SVM method outperformed other face detection methods; CNN, Haar-Cascade, and Faced methods. There is no error while detecting faces in the BioID dataset with this method. Second, the FaceNet and the Resnet-50 architectures, which are the deep convolutional neural network (CNN), are proposed to use as the face encoding methods. Surprisingly, these two deep CNN architectures obtained an accuracy of 100% on the BioID and FERET datasets. Moreover, ResNet-50 architecture was slightly better than FaceNet architecture. The ResNet-50 and FaceNet architectures obtain very high verification accuracy on ColorFERET dataset, with accuracy of 99.60% and 99.32%, respectively.

REFERENCES

[1] D. Li, H. Zhou, and K. M. Lam, "High-Resolution face verification using pore-scale facial features," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2317–2327, 2015.

[2] A. Singh, M. Singh, and B. Singh, "Face detection and eyes extraction using Sobel edge detection and morphological operations," in *Conference on Advances in Signal Processing (CASP)*, 2016, pp. 295–300.

[3] C. Li, Z. Qi, N. Jia, and J. Wu, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," in *IEEE 13th International Conference on Electronic Measurement and Instruments (ICEMI)*, 2017, pp. 483–487.

[4] E. K. Shimomoto, A. Kimura, and R. Belem, "A faster face detection method combining bayesian and Haar cascade classifiers," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2015, pp. 7–12.

[5] A. Ade-ibijola and K. Aruleba, "Automatic attendance capturing using histogram of oriented gradients on facial images," in *IST-Africa Week Conference (IST-Africa)*, 2018, pp. 1–8.

[6] H. X. Jia and Y. J. Zhang, "Fast human detection by boosting histograms of oriented gradients," in *Proceedings of the Fourth International Conference on Image and Graphics Fast (ICIG)*, 2007, pp. 683–688.

[7] H. ChunYang and X. A. Wang, "Cascade face detection based on histograms of oriented gradients and support vector machine," in *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, 2015, pp. 766–770.

[8] H. Shu, D. Chen, Y. Li, and Shengjin Wang State, "A highly accurate facial region network for unconstrained face detection," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 665–669.

[9] L. Pang, Y. Ming, and L. Chao, "F-DR Net:Face detection and recognition in one net," in *International Conference on Signal Processing (ICSP)*, 2018, pp. 332–337.

[10] A. R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: face and expression recognition," *IEEE Trans. image Process.*, vol. 22, no. 5, pp. 1740–1752, 2013.

[11] F. Juefei-Xu and M. Savvides, "Encoding and decoding local binary patterns for harsh face illumination normalization," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3220–3224.

[12] D. Gong, Z. Li, W. Huang, X. Li, and D. Tao, "Heterogeneous face recognition: a common encoding feature discriminant approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2079–2089, 2017.

[13] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1273–1280, 2017.

[14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015, pp. 1–12.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.

[16] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 815–823, 2015.

[17] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy, "Pose-Robust Face Recognition via Deep Residual Equivariant Mapping," in *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5187–5196.

[18] D. E. King, "Max-margin Object Detection," 2015.

[19] O. Surinta and S. Khruahong, "Tracking people and objects with an autonomous unmanned aerial vehicle using face and color detection," in *International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*, 2019, pp. 206–210.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[21] P. Viola and M. Jones, "Robust real-time object detection," *Vingtieme Siecle Rev. d'Histoire*, vol. 57, pp. 1–25, 2007.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[23] M. Karaaba, O. Surinta, L. Schomaker, and M. A. Wiering, "Robust face recognition by computing distances from multiple histograms of oriented gradients," in *IEEE Symposium Series on Computational Intelligence, (SSCI)*, 2015, pp. 203–209.

[24] V. N. Vapnik, *Statistical Learning Theory*. 1998.

[25] O. Surinta, M. F. Karaaba, L. R. B. Schomaker, and M. A. Wiering, "Recognition of handwritten characters using local gradient feature descriptors," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 405–414, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[27] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *Lecture Notes in Computer Science book series (LNCS)*, 2001, pp. 90–95.

[28] P. J. Phillips, P. J. Rauss, and S. a. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[29] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, pp. 295–306, 1998.

Review Rating Prediction with Gaussian Process Classification

Hidekazu Yanagaimoto
College of Sustainable System Sciences
Osaka Prefecture University
Sakai, Japan
Email: hidekazu@kis.osakafu-u.ac.jp

Kiyota Hashimoto
ESSAND
Price of Songkla University
Phuket, Thailand
Email: kiyota.hashimoto@gmail.com

Many researchers pay attention to training a nonlinear function according to training data because machine learning can represent more flexible function. One of approaches is extend a linear function to a nonlinear function with a kernel method. A typical method is Support Vector Machine (SVM), which is a discriminative model. In SVM, a kernel function is defined previously and you need much knowledge on tasks to choose a good kernel function. Hence, the kernel function selection affect the final performance of SVM are many kernel functions are developed according to each task. Another approach is Gaussian process, which is a generative model. Gaussian process estimate describes a distribution over functions and directly search the optimal function in function space. In SVM, a cost function is defined previously and the optimal parameters are searched with respect to training data. Gaussian process and SVM can construct a nonlinear function but their approaches are different. How are Gaussian process and SVM different in practice? In this paper, we evaluate Gaussian process and SVM with respect to review rating prediction, which is one of natural language tasks. Sentiment analysis research denotes some words contribute to sentiment polarity of a sentence strongly. However, deep learning improves the sentiment analysis adding nonlinear feature construction and it is clear that we have to deal with nonlinearity in sentiment analysis. Moreover, review rating prediction is one of multi-class classifications and the prediction needs more flexibility. Finally, when a review is transformed into a numerical vector with a Bag-of-Words model, the review vector has a very high dimension. Hence, the task have to avoid curse of dimensionality. Kernel trick is a main solution in SVM and a flexible regression function is additional solution in Gaussian process regression.

***Abstract*—**

***Keywords*—Gaussian process, Gaussian process classification, review rating prediction, natural language processing**

I. Introduction

Many researchers pay attention to training a nonlinear function according to training data because machine learning can represent more flexible function. One of approaches is extend a linear function to a nonlinear function with a kernel method[1] . A typical method is Support Vector Machine (SVM)[2], which is a discriminative model. In SVM, a kernel function is defined previously and you need much knowledge on tasks to choose a good kernel function. Hence, the kernel function selection affect the final performance of SVM are many kernel functions are developed according to each task.

Another approach is Gaussian process[3], [4], which is a generative model. Gaussian process estimate describes a distribution over functions and directly search the optimal function in function space. In SVM, a cost function is defined previously and the optimal parameters are searched with respect to training data.

Gaussian process and SVM can construct a nonlinear function but their approaches are different. How are Gaussian process and SVM different in practice? In this paper, we evaluate Gaussian process and SVM with respect to review rating prediction[5], [6], which is one of natural language tasks. Sentiment analysis research denotes some words contribute to sentiment polarity of a sentence strongly. However, deep learning improves the sentiment analysis adding nonlinear feature construction and it is clear that we have to deal with nonlinearity in sentiment analysis. Moreover, review rating prediction is one of multi-class classifications and the prediction needs more flexibility. Finally, when a review is transformed into a numerical vector with a Bag-of-Words model, the review vector has a very high dimension. Hence, the task have to avoid curse of dimensionality. Kernel trick is a main solution in SVM and a flexible regression function is additional solution in Gaussian process regression.

From evaluational experiments, Gaussian process classification[7] is superior to SVM when there are many training data with respect to a review vector dimension. It means that Gaussian process regression can construct a regression function which can capture characteristics of training data and deal with unobserved data. Gaussian process classification generally needs more computation time than SVM because Gaussian process can estimate a latent structure in training data. Hence, making Gaussian process classification learning to be fast is one of important problems.

This paper is constructed below. In Section 2 we describe the proposed method. Especially, we explain Gaussian process, Gaussian process regression, Gaussian process classification, and review rating prediction. So we give you all element techniques of the propose method. In Section 3 we carry out some evaluational experiments with Amazon product review data set. First, we describe characteristics of Amazon review data set. After then, we execute some experiments and discuss the performance of the proposed method. In Section 5 we

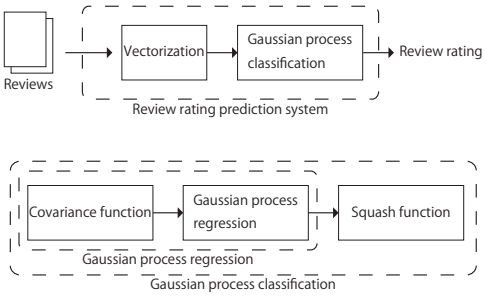


Fig. 1. Architecture of the proposed method

describe conclusions and future works.

II. Review Rating Prediction with Gaussian Process Classification

In this section, we describe a Gaussian process classification, which consists of Gaussian process regression and a prediction probability generator. In Figure 1, an architecture of the proposed method is shown.

Moreover, we explain how to apply Gaussian process classification to review rating prediction.

A. Gaussian Process

A Gaussian process is a collection of random variables which have a joint Gaussian distribution. The Gaussian process is defined with its mean function, $m(\mathbf{x})$ and covariance function, $k(\mathbf{x}^*, \mathbf{x})$. The functions are defined below.

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{2}$$

We can write the Gaussian process with the functions.

$$f(\mathbf{x}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}^*), K(\mathbf{x}^*, \mathbf{x})) \tag{3}$$

$f(\mathbf{x}^*)$ means a random variable at location $\mathbf{x}$. Because a Gaussian process is defined as a collection of random variables, the Gaussian process can generate other values in the function, $f(\mathbf{x}^*)$. It means that Gaussian process can generate a function, $f(\mathbf{x})$.

B. Gaussian Process Regression

In linear regression, parameters, $\mathbf{w}$, are estimated with a cost function, $F(\mathbf{w})$, based on training data.

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}) \tag{4}$$

$$F(\mathbf{w}) = \frac{1}{2}\sum_i \|\mathbf{t}_i - \mathbf{y}_i\|^2 \tag{5}$$

In linear regression, the parameter estimation is a point estimation and the regression function is not enough ability to describe a model for training data set.

One of improved approach is a Bayesian approach, which introduce distributions for the parameters and observations. Now we assume observations are generated by a Gaussian distribution and a prior distribution of the parameters is a Gaussian distribution, too. In this case, we can formalize observation generation below.

$$p(\mathbf{y}|\mathbf{w}) = \mathcal{N}(\bar{\mathbf{y}}, \Sigma_{\mathbf{y}}) \tag{6}$$

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{w}}) \tag{7}$$

In this case, we can can calculate a predictive distribution from the previous settings.

$$p(\mathbf{y}^*|\mathbf{x}^*, X, \mathbf{y}) = \int p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{w})\, p(\mathbf{w}|X, \mathbf{y})\, d\mathbf{w} \tag{8}$$

$p(\mathbf{w}|X, \mathbf{w})$ is defined belwo.

$$p(\mathbf{w}|X, \mathbf{w}) = \frac{p(\mathbf{y}|X, \mathbf{w})\, p(\mathbf{w})}{p(\mathbf{y}|X)} \tag{9}$$

$$\begin{aligned} p(\mathbf{y}|X, \mathbf{w}) &= \prod_i p(y_i|\mathbf{x}_i, \mathbf{w}) \\ &= \prod_i \mathcal{N}\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{w}, \sigma_n^2 I\right) \end{aligned} \tag{10}$$

Hence, the predictive distribution is a Gaussian distribution, $\mathcal{N}\left(\frac{1}{\sigma_n^2}\mathbf{x}^{*\mathrm{T}}A^{-1}X\mathbf{y}, \mathbf{x}^{*\mathrm{T}}Q^{-1}\mathbf{x}^*\right)$, where $A = \frac{1}{\sigma_n^2}XX^{\mathrm{T}} + \Sigma_{\mathbf{y}}^{-1}$.

Another extension of linear regression is a kernel method. Basically, linear regression function is restricted under a linear function but a model of real data is not restricted. To make the regression function to be nonlinear, the kernel method is introduced.

At first, a basis function, $\phi(\mathbf{x})$ is introduced, which maps the input data, $\mathbf{x}$, into a feature space. For example, when we employ $\phi(x) = \left(1, x, x^2, \cdots\right)$, we can construction a polynomial function with an linear regression method. Using the basis function, a predictive distribution is obtained below.

$$\begin{aligned} & p(\mathbf{y}^*|\mathbf{x}^*, X, \mathbf{y}) \\ = \;& \mathcal{N}\left(\phi(\mathbf{x}^*)^{\mathrm{T}} A^{-1}\phi(X)\,\mathbf{y}, \phi(\mathbf{x}^*)^{\mathrm{T}} A^{-1}\phi(\mathbf{x}^*)\right) \\ = \;& \mathcal{N}\Big(\phi(\mathbf{x}^*)^{\mathrm{T}}\Sigma_{\mathbf{y}}\phi(X)\left(K + \sigma_n^2 I\right)\mathbf{y}, \\ & \phi(\mathbf{x}^*)^{\mathrm{T}}\Sigma_{\mathbf{y}}\phi(\mathbf{x}^*) \\ & -\phi(\mathbf{x}^*)^{\mathrm{T}}\Sigma_{\mathbf{y}}\phi(X)\left(K + \sigma_n^2 I\right)^{-1}\phi(X)^{\mathrm{T}}\Sigma_{\mathbf{y}}\phi(\mathbf{x}^*)\Big) \end{aligned} \tag{11}$$

Now, we define $\Psi(\mathbf{x}) = \Sigma_{\mathbf{y}}^{\frac{1}{2}}\phi(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}') = \Psi(\mathbf{x})^{\mathrm{T}}\Psi(\mathbf{x})$. We can rewrite the predictive distribution below.

$$\begin{aligned} p(\mathbf{y}^*|\mathbf{x}^*, X, \mathbf{y}) = \mathcal{N}\Big(&k(\mathbf{x}^*, X)\left(k(X, X) + \sigma_n^2 I\right)\mathbf{y}, \\ &k(\mathbf{x}^*, \mathbf{x}^*) - k(\mathbf{x}^*, X)\left(k(X, X) + \sigma_n^2 I\right)^{-1} k(X, \mathbf{x}^*)\Big) \end{aligned} \tag{12}$$

The result means that you do not need to define the basis function directly to estimate the predictive distribution but we have to define a kernel function, $k(\mathbf{x}, \mathbf{x}')$.

Some problems are remained by these extension of linear regression. The basis function is determined previously and is not the optimal function for training data set. Gaussian process regression can solve the problem because the Gaussian process regression constructs more flexible regression function.

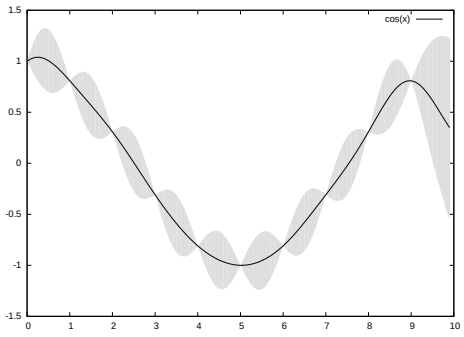


Fig. 2. Example of Gaussian process regression

Now, training data set and test data are $(X, \mathbf{y})$ and test data $\mathbf{x}^*$ respectively. In this case, the training data set and the test data set are generated with Gaussian process regression below.

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{y}^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} f(\mathbf{x}) \\ f(\mathbf{x}^*) \end{pmatrix}, \begin{pmatrix} k(X,X) & k(X,\mathbf{x}^*) \\ k(\mathbf{x}^*,X) & k(\mathbf{x}^*,\mathbf{x}^*) \end{pmatrix} \right) \tag{13}$$

From the joint probability, the predictive distribution is calculates.

$$p(\mathbf{y}^*|X,\mathbf{y},\mathbf{x}^*) = \mathcal{N}\Big(k(\mathbf{x}^*,X)\, k(X,X)^{-1} f(X) \tag{14}$$

$$k(\mathbf{x}^*,\mathbf{x}^*) - k(\mathbf{x}^*,X)\, k(X,X)^{-1} k(X,\mathbf{x}^*) \Big) \tag{15}$$

In Figure 2, example of Gaussian process regression is shown. In this example, we estimate a cosine function with only 10 observations. Gaussian process regression can predict a mean and a variance and we can discuss confidence of the predictions. For example, high variance means prediction includes much ambiguity.

C. Gaussian Process Classification

In classification tasks, an outputs is a discrete value, which denote a class data belongs to. On the other hand, an output from Gaussian process regression is a continuous value and is not restricted within the number of the classes. In this section, we focus on a binary classification task and transform the Gaussian process regression into a binary classifier.

To apply the Gaussian process regression to a classification task, we use a squash function, $\sigma(x)$, which maps a real number into $(0, 1)$.

$$p(c|\mathbf{y}^*) = \sigma(\mathbf{y}^*) \tag{16}$$

A sigmoid function is often employed as the squash function.

Introducing the squash function, a new problem happens, which the predictive distribution is not a Gaussian distribution.

$$p(c|\mathbf{x}^*,X,\mathbf{y},\mathbf{x}^*) = \int \sigma(\mathbf{y}^*)\, p(\mathbf{y}^*|X,\mathbf{y},\mathbf{x}^*)\, d\mathbf{y}^* \tag{17}$$

$p(\mathbf{y}^*|X,\mathbf{y},\mathbf{x}^*)$ is a Gaussian distribution because the distribution is defined with Gaussian process regression but $p(c|\mathbf{x}^*,X,\mathbf{y},\mathbf{x}^*)$ is not a Gaussian distribution and is not tractable. To avoid this problem, we employ a Laplace's approximation method.

TABLE I. CONTENT OF TRAINING DATA AND TEST DATA

Category	Training	Test	Vocabulary
Books	20,000	4,000	12,512
Electronics	20,000	4,000	7,058
Video Games	20,000	4,000	13,079
Clothing, Shoes and Jewelry	20,000	8,000	4,906

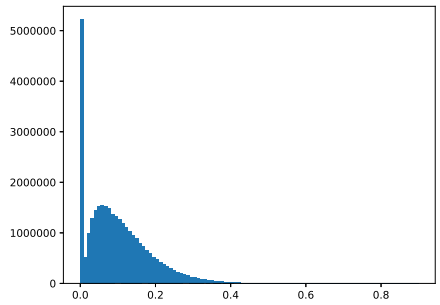


Fig. 3. Cosine similarity among training data

D. Review Rating Prediction

Review rating prediction is one of multi-class classification tasks because we predict a review rating from user's review. The review rating in Amazon is selected from $\{1, 2, 3, 4, 5\}$ and the user review is written with a natural language.

We transform the user review into a vector with a Bag-of-Words model because of constructing an input vector for Gaussian process classification. Of course, when we construct the review vector, we neglect stop words and less frequent words. The dimension of the vector depends on the size of a vocabulary of whole reviews and the dimension is usually very large.

III. EXPERIMENTS

In this paper, we applied the proposed method to some Amazon review datasets, which consists of some product categories because we assume that users' review styles and review rating criteria changes according to product categories.

A. Datasets

In this evaluation experiments, we employ four-category product reviews[8], [9][1]; Books, Electronics, Video Games, and Clothing, Shoes and Jewelry. We pick up 20,000 reviews for training data and 4,000 reviews for test data for each category at random. The content of the training data and the test data is shown in Table I.

The vocabulary size varies according to product category because evaluation criteria are different with respect to its product category. Hence, the review vectors in Books and Video Games are embedded in higher dimensional input spaces. In high dimensional space, each data is uncorrelated each other because of curse of dimensionality. In Fig. 3, cosine similarity among 8,000 training data is shown. Many cosine similarity among training data has 0.0 and training data is uncorrelated each other. The tendency exits in relation between

[1] http://jmcauley.ucsd.edu/data/amazon/

TABLE II. DISTRIBUTION OF REVIEW RATING IN TRAINING DATA

Dataset	Training data				
	5	4	3	2	1
Books	4,964	1,633	673	328	402
Electronics	4,347	1,557	659	476	961
Video Games	4,097	1,951	970	476	506
Clothing, Shoes and Jewelry	4,665	1,548	777	447	563

TABLE III. DISTRIBUTION OF REVIEW RATING IN TEST DATA

Dataset	Test data				
	5	4	3	2	1
Books	1,246	404	154	88	108
Electronics	1,136	385	154	102	223
Video Games	991	481	271	126	131
Clothing, Shoes and Jewelry	1,137	377	216	129	141

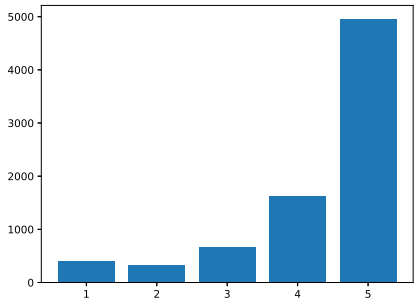


Fig. 4. Review rating distribution in "Book"

training data and test data and it is difficult to predict test data labels with small training data.

A distribution of review ratings are very skew. In Table II, Table III, and Figure 4, the distribution of review rating is shown. Because of skew distributions of training data, it is more difficult to learn a classifier and the classifier causes overfitting easily. For example, because "Books" dataset includes many "5" raging, classifier tends to predict only 5 for all test data. In this case, prediction accuracy achieves about 0.6 and this strategy is better one.

B. Settings

The proposed method includes an important hyper-parameter, a covariance function, and the hyper-parameter is determined previously. In the experiments, the covariance function is defined below.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2\right) \tag{18}$$

We employed a support vector machine (SVM) as a comparative method because the evaluation experiments are one of classification tasks and SVM is the most famous approach. Additionally, both of the methods are similar and we can compare them easily. SVM consists of a kernel method, which maps input data into a feature space nonlinearly, and linear discrimination with a margin maximization criterion. On the other hand, Gaussian process regression consists of nonlinear regression function, which maps input data into a feature space nonlinearly, and logistic discrimination. Performance difference is caused by architecture difference. Especially, it is nonlinear mapping from an input space to a feature space nonlinearly.

In SVM, we use an RBF functions as a kernel function. The RBF kernel is similar to the covariance function in Gaussian process classification(Equation (18)). In this experiments, we employ scikit-learn python module to use SVM.

Both of the methods are a binary classifier and they are not applied to multi-class classification tasks directly. We employed the 1-versus rest strategy to extend the binary classifier for multi-class classification tasks.

TABLE IV. CONTENT OF TRAINING DATA IN EXPERIMENT 1

	Training size	Test size	Vocabulary
Data 1	20,000	4,000	12,515
Data 2	8,000	4,000	11,782
Data 3	5,000	4,000	10,954

TABLE V. PREDICTION ACCURACY WITH RESPECT TO TRAINING DATA SIZE

Dataset	The proposed method	SVM
Data 1	0.639	0.637
Data 2	0.633	0.633
Data 3	0.630	0.630

TABLE VI. CONTENT OF TRAINING DATA IN EXPERIMENT 2

Product category	Vocabulary
Electronics	6,774
Video Games	12,593
Clothing, Shoes and Jewelry	4,772

C. Results

1) Experiment 1: We discuss performance of the proposed method and SVM according to the size of training data. In product category including a large vocabulary, review vectors are embedded in high dimensional input space. In this case, curse of dimensionality tends to happen in high dimensional space. For example, we need a large training data to construct an appropriate classifier because many data is not related each other. In Table IV, the content of training data is shown. In this dataset, the dimension of a review vector is over 10,000 and the dimension of input space is very large.

In Table V, the performances are shown. As the size of training data increases, the proposed method improved the accuracy faster than SVM. When the size of training data is small, the coverage of training data in a input space is small and each training data is uncorrelated. It means that training data does not contribute to estimation of test data and it is difficult to map test data into the feature space with Gaussian process regression. Hence, as the size of training data increases, the Gaussian process regression is improved and the mapping is correct.

2) Experiment 2: We discuss performances according to product categories. Considering the result of Experiment 1 and computational cost, we determine the size of training data as 8,000 and test data as 2,000 respectively. Using 8,000 training data, the proposed method is comparative with SVM. In Table VI, the content of training data is shown. "Electronics" and "Clothing, Shoes and Jewelry" are smaller input spaces but "Video Games" is a larger input space.

In Table VII, performances are shown.

TABLE VII. PREDICTION ACCURACY WITH RESPECT TO PRODUCT CATEGORY

Dataset	The proposed method	SVM
Electronics	0.601	0.583
Video Games	0.516	0.527
Clothing, Shoes and Jewelry	0.595	0.578

TABLE VIII. CONFUSION MATRIX FOR "CLOTHING, SHOES, AND JEWELRY" WITH THE PROPOSE METHOD

21	0	3	5	112
6	0	3	10	110
1	0	6	41	168
0	0	0	47	330
1	0	0	20	1116

TABLE IX. CONFUSION MATRIX FOR "CLOTHING, SHOES, AND JEWELRY" WITH SVM

8	0	1	0	132
2	3	1	1	122
5	1	4	10	196
1	1	6	8	361
1	0	1	2	1133

In Table VIII and IX, confusion matrices for the proposed method and SVM in "Clothing, Shoes, and Jewelry" dataset The confusion matrices denote Gaussian process classification can predict correct review rating for less reviews in training datas set. A classifier tends to predict major raging in training data for test data set because a cost function, which is employed in learning step decreases. In Table IX, SVM can predict many 5 ratings and for 5 rating reviews SVM achieves highest accuracy. However, from the viewpoint of whole test data set, Gaussian process classification is superior to SVM.

When training data can cover an input space, the proposed method is superior to SVM. The results mean that Gaussian process regression can construct a better mapping function from an input space to a feature space depending on training data.

D. Discussions

Both of SVM and Gaussian process classification map input data into high dimensional feature space with a kernel method and construct a linear discriminative function and a nonlinear regression function respectively. Basically, Gaussian process classification is superior to SVM because of nonlinear mapping according to training data distribution. On the other hand, when there are not enough training data, it is difficult to construct appropriate nonlinear regression in Gaussian process classification. It causes that the performance in Gaussian process classification decrease more than in SVM.

Gaussian process classification can construct appropriate classifier regardless of training data set bias. Of course, when there are less training data set with respect to input space dimension, Gaussian process classification does not achieve good accuracy. However, when there are enough training data set, Gaussian process classifier can construct an appropriate nonlinear regression function and improve generalization for unseen data. I think this characteristics is caused by nonlinear regression function construction according to training data. In SVM, a kernel function is predefined one but Gaussian process classification optimizes the regression function according to training data set.

Gaussian process classification need more computational time than SVM because Gaussian process classification optimizes the latent structure and employs Laplace's approximation. We have to discuss relation between computational time and prediction accuracy in future.

IV. CONCLUSIONS

We propose review rating prediction with Gaussian process classification and compare the proposed method with SVM. From evaluational experiments, we confirmed that Gaussian process classification is superior to SVM with respect to prediction accuracy. Especially, Gaussian process classification can consider less rating data in training data is paid attention to. When there are enough training data with respect to input space dimension, Gaussian process classification can construct a better regression function which can obtain characteristics of training data. On the other hand, in all cases, Gaussian process classification needs more learning time than SVM because Gaussian process classification need to model a latent structure in training data, which denotes a regression function. Moreover, Gaussian process classification need Laplace's approximation to describe a Gaussian distribution of the final prediction. Especially, the later approximation is important to evaluate a prediction from a classifier but in simple classification tasks, it is sometimes meaningless process.

In future works, we have to make learning to be fast introducing more approximations. For example, Gaussian process need to calculate a inverted covariance matrix. The computation cost is equal to $O(n^3)$. it means that as training data, n, increases as computational time increase by n^3. To avoid this computational cost, we can approximate the covariance matrix with smaller covariance matrix. However, because the approximation affects the performance of the final prediction directly. Hence, we need more discussion of the approximation. In another future work, we apply the proposed method to other reviews generated from other product categories. Product reviews have some characteristics with respect to product categories. Hence, we have to discuss coverage of the proposed method.

REFERENCES

[1] R. Herbrich, *Learning Kernel Classifiers*, The MIT Press, Cambridge, 2002.

[2] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1999.

[3] C. E. Rasmussen and K. I. Williams, *Gaussian Process for Machine Learning*, the MIT press, 2006.

[4] D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.

[5] L. Qu, G. Ifrim, and G. Weikum, *The bag-of-opinions method for review rating prediction from sparse text patterns*, In Proceedings of the 23rd International Conference on Computatinal Linguistics (COLING2010), pp.913-921, 2010.

[6] C. Leung, S. Chan, and F.Chung, *Integrating collaborative filtering and sentiment analysis: A rating inference approach*, Proceedings of the ECAI 2006 workshop on recommender system, pp.62-66, 2006.

[7] H. Nickisch and C. E. Rasmussen, *Approximation for Binary Gaussian Process Classificatio*, Journal of Machine Learning Research, Vol. 9, pp.2035-2078, 2008.

[8] R. He, and J.McAuley, *Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filetering*, Procs. of WWW2016, 2016.

[9] J. McAuley, C. Targett, J. Shi, A. van den Hengel, *Image-based recommendations on stules and substitutes*, Procs. of SIGR2015, 2015.

Predictive Analytics of Various Factors Influencing Gold Prices in Thailand using ARIMA Model on R

Nuntouchapron Prateepausanont
Management of information Technology, Faculty of Engineering,
Prince of Songkla University
Songkhla, Thailand
nuntouchapron@gmail.com

Chidchanok Choksuchat*
Information and Communication Technology Programme
Faculty of Science,
Prince of Songkla University
cchoksuchat@hotmail.com

Sureena Matayong
Management of information Technology
Faculty of Engineering,
Prince of Songkla University
Songkhla, Thailand
sureena.m@psu.ac.th

***Abstract*— Predictive analytics of numerous factors influencing gold prices in Thailand is a significant topic to study because gold is a precious metal that can grow the world economy. The purpose of this study is to predict gold prices in Thailand based on various factors by means of Autoregressive Integrated Moving Average (ARIMA) models in R programming language. The influencing factors to predict the gold prices are; spot gold, consumer price index, exchange rate, gold prices in Thailand, inflation rate, stock exchange of Thailand index, interest rate and crude oil price. This study uses monthly data are retrieved from several sources; world gold council, bank of Thailand, gold traders' association and U.S. energy information administration during 2009 to 2018. The experimental results through comparing different ARIMAL models provided the most accuracy that returned the validated value via RMS, MAPE, and U as 0.66, 9.12, and 0.1 respectively.**

Keywords*— *predictive analytics, gold price, ARIMA, R language

I. INTRODUCTION

For many decades, gold is always a potential metal that can impact on world economy growth. Changes in gold prices have an effect on the daily global economy until forecast in the future. In 2017, 52.85 percent of gold was used in the jewelry industry, 30.10 percent for investment sector, 9.03 percent is accounted for international reserves of central banks around the world and other industries with 8.02 percent [1]. Stockholders and investors are the very large group who pay more attention to gold investment because the gold is a highly secure asset that they have gained the absolute returns with high liquidity. Although, the gold is considered as a highly secure asset to invest on however, entering into the gold market is not out of risks, since predicting the price of gold requires many influencing factors that is inconstant to analyze before investing. Therefore, this research presents the predictive analytic of various influencing factors that are reviewed and updated from most recent studies found of gold prices in Thailand using ARIMA model in R programing language. The study uses monthly data from many different sources during 2009-2018.

ARIMA model is well known in particular time series analysis in statistics and econometrics. Time series is information that changes with time, which constantly collecting observations that can be used to predict trends. The trends could provide changes information in terms of up and down like the information about the price of gold that is fluctuated. To achieve the purpose of this study, there are 2 objectives perform in this paper; (1) to study and analyze numerous factors that influencing Thailand gold prices and collect them for predictive analytics. (2) To compare the predictive analytics of Thailand gold prices from those factors with various ARIMA models in R programing language. This paper structures as follows. Section 2 mentions related work. Section 3 addresses through software implementation methods. Section 4 presents the analysis and result of this research. The remains are conclusion, future work, and references.

II. RELATED WORK

A. ARIMA Model in Gold Price Predictive Analytics

Gaysornpratoom [2] compared the accuracy in gold price forecasting between three models. These are neural networks, ARIMA and GARCH – M models. The results are shown that the most accurate model of gold price forecasting is the ARIMA model. The other works brought ARIMA model to gold price predicting experiments are as follows. Chotiprapa [3] used Box-Jenkins process to align the same subject that forecast gold price in Thailand. The paper represented the ARIMA model and applied to ARIMAX model comparison. When considered to the RMSE values, they found that ARIMA model and ARIMAX model gave the same accuracy predictions. In particular investigation more inside of ARIMA model [4], they concluded that after stationary process which is contained the data at first difference. Software names E-views which is executed for fitting the model's coefficient, using several graphs, concerned statistics, computing of ACFs (complete auto-correlation function) and PACFs (partial auto-correlation function) of residuals and after several iterations, the model selected is ARIMA (0,1,1). Colwell [5] used five models for predicted gold price in USA are actual, auto ARIMA, manual ARIMA, time series decomposition and exponential smoothing state space model. While they compared the forecasting performance using the Mean Absolute Percentage Error (MAPE), it was found that the auto ARIMA model was the most efficient model for forecasting. Guha and Bandyopadhyay [6] focused on ARIMA model in gold price forecasting domain as well. The content concerns to the performance analysis from data preceding 10 years traded value with ARIMA (1, 1, 1) model in predicting the future values of Gold. It consists of six different model parameters which satisfied all the criteria of fit statistics. Yang [7] applied ARIMA model to predict the international gold price as well. The results are indicated the true gold price trend. More importantly, the output can provide consumers a guide in gold investment. However, due to the gap between the complexity of the real world and the limitations of the model itself, which cannot be fully precisely measured and forecasted.

To summarize about the gold price which has an outstanding characteristic that appropriated to time series and dealt with seasonal component. ARIMA is an effective model to predictive gold price analysis. Moreover, measuring forecasting accuracy has several ways; in our case, we use Root Mean Square Error (RMSE) to validate in the first step, and the other two formulas are Theil's Inequality Coefficient (U), and Mean Absolute Percentage Error (MAPE). Regarding to predictive analytics of gold prices purpose, many factors are potential to evaluate, we explain in the subsection B below.

B. Veroius Factors in Gold Price Predictive Analytics

The researchers first reviewed previous studies regarding factors influencing gold prices and found out 10 relevant factors. There are spot gold, consumer price index, exchange rate, gold prices in Thailand, inflation rate, crude oil price, stock exchange of Thailand index, interest rate, international gold reserve, and silver metal prices. The researchers selected 8 from 10 factors by considering the frequency of studies for each factors at least 3 times from different researchers. The international gold reserve, and silver metal prices are disregarded from this study since the frequency found from different researchers are least than 3. TABLE I shows the literature review result about various factors influencing gold price.

TABLE I. FACTORS INFLUENCING GOLD PRICES IN THAILAND

Ref No.	Factors Influencing Gold Prices in Thailand							
	Spot Gold	CPI	EX	Gold (T)	INF	SET Index	Interest rate	Crude Oil Price
[8]		√	√	√				
[9]			√	√				
[10]	√						√	√
[11]	√	√	√					
[12]						√	√	√
[13]					√	√		√
[14]					√		√	√
[15]		√	√	√				
[16]	√				√	√		
Total	3	3	4	3	3	3	3	4

III. METHODOLOGY

This section addresses the methodology of the research, system architecture, workflow of algorithm, and type of input.

A. System Architecture and Machine Specification

According to system architecture as shown in Fig 1. The overview displays for predicting gold price process in Thailand of this research. The procedure starts from an acquisition of data. The researchers extracted the data from several sources; world gold council, bank of Thailand, gold traders' association and U.S. energy information administration.

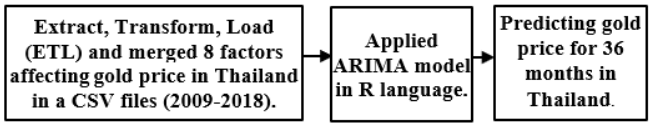


Fig. 1. System Architecture Overview of Predictive Analytics Process

We have separated the total of 10 years data to 7 years for training and 3 years for validation. Next, the data are transformed to numeric for loading into ARIMA model through R Studio. The inputs include 8 factors influence the gold price in Thailand, which is recorded as a comma-separated values (CSV) file format. The output of this process is predicting result of gold price in Thailand for 36 months.

The predictive analytics was performed by machine specification; processer: Intel(R) Core (TM) i7-65000U CPU @ 2.50GHz with 2.60 GHz, system type: 64-bit operating system, x64-based processor, installed memory (RAM): 8.00 GB, and software: R 3.5.3 and R Studio 1.2.1335.

B. Operating of Data Processing using ARIMA model

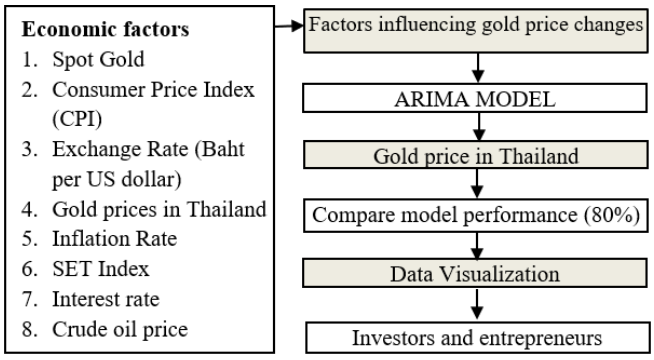


Fig. 2. Research process of predicting gold prices in Thailand by ARIMA

Fig 2 is shown an overview of the research by bringing 8 factors that influence the prices of gold in Thailand using ARIMA model. Based on ARIMA model, all parameters are investigated for inputs to predicting process in R program for each model performance results. Then, compare the performance results with all models. Finally, the most effective model will be selected and applied to develop data visualization for investors and entrepreneurs. The development of data visualization is not included in this paper.

C. Dealing with Various Input Datasources

The predicting data of 8 factors: spot gold, consumer price index, exchange rate, gold prices in Thailand, inflation rate, stock exchange of Thailand index, interest rate and crude oil price as shown in TABLE II. There were total 672 data used for learning models, which is monthly from 1 January 2009 until 31 December 2015 as shown in Fig 3. Another group of data has been used for testing the accuracy of the gold price data in Thailand from 1 January 2016 to 31 December 2018.

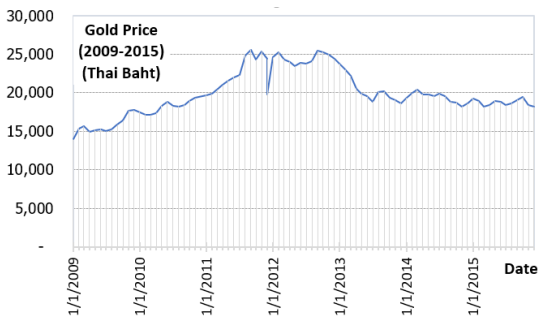


Fig. 3. Monthly gold price from 2009 to 2015

Before we merge 8 input data in example of table 2 which are located in the grey shade header columns. We have

catergirized the date to many types of date-time, e.g., season, order number of interested year that starts from index *0-n*, month, and merged to concerning data set, such as, SET, crude oil price and exchange rate.

TABLE II. EXAMPLE OF 8 INPUTS DATA FOR INFLUENCING FACTORS OF GOLD PRICE IN THAILAND

Instant	Date (MM/DD/YY)	Season	Order of Year	Month	SET	Crude Oil Price	Gold price of Thailand	Spot gold	Interest rate	Exchange rate	Consumer price index	inflation
1	12/1/15	4	0	12	1061.3	37.11	18,182	1,061.30	1	36.03	104.98	0.20
2	11/1/15	4	0	11	1064.42	41.69	18,394	1,064.42	1	35.80	105.46	0.20
3	10/1/15	4	0	10	1142.38	41.69	19,465	1,142.38	1	35.60	105.89	0.20

D. ARIMA Algorithm

There is a procedure in the proposed ARIMA model. In the first phase, this model is suitable for modeling the linear data of time series as shown in Fig 4, which adapted from [17]. The rule is considered the ARIMA models' residuals are homoscedastic. We can explain in detail with the implementation process with R in 6 steps as Fig 5.

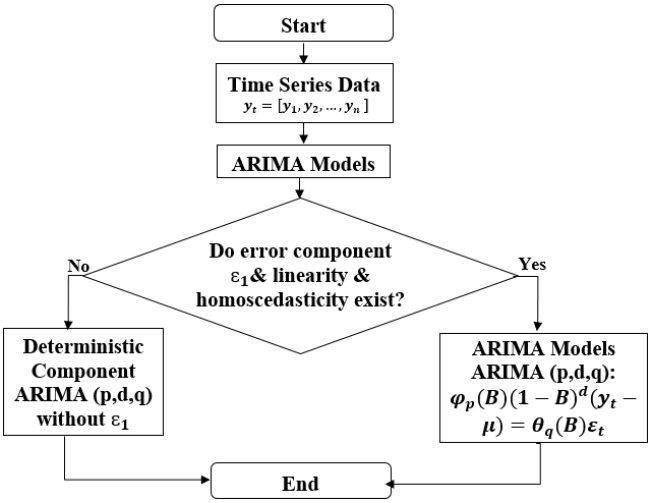


Fig. 4. Flowchart of the procedure for ARIMA models

Step 1: Explore the data by *plot()* and examine its patterns and irregularities and clean up any missing values or outliers with *tsclean()* function that is a convenient method for inputting missing values and outlier removal.

Step 2: Examine and possibly remove components of the series by *decompose()* or *stl()*.

Step 3: Investigate by condition if the data is stationary or non-stationary. ACF, PACF are plotted to determine order of differencing needed by *adf.test()*,.

Step 4: Examine ACF and PACF plots for selecting an order of the ARIMA.

Step 5:Used ARIMA modelling in R through the *arima()* function, or *auto.arima()* that is an automatically generate a set of optimal (p, d, q).

Step 6: Deal with residuals, which should have no patterns and be normally distributed and refit model if desirable. Compare model errors and fit criteria such as AIC or BIC. After that calculate prediction using the chosen model.

After do the ARIMA model in R, we have analyzed the result with the other 4 steps as the four blue arrows in Fig.5.

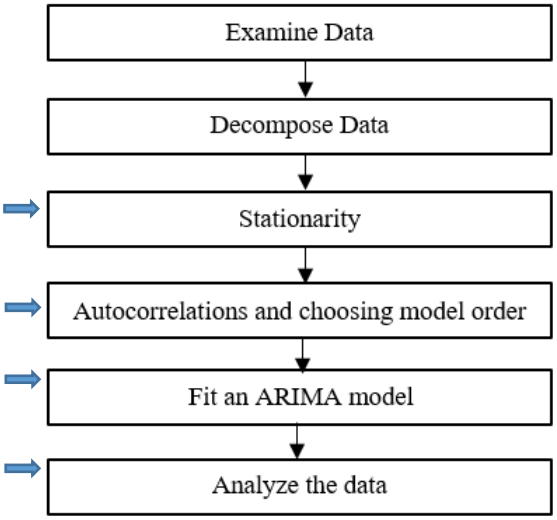


Fig. 5. Implementation process of this paper ARIMA with R

Step 1: Test the properties, a stationary with a Unit Root Test using Lag Length. AIC considers the P-value of the Augmented Dickey-Fuller t-stat test statistical probability more than the set level of significance. It is mean accepted as the main reason for the nonstationary.

If found nonstationary or the instability of data must eliminate the stagnation before creating a time series model which can be done by finding the difference Then test stationary. Using a unit root test using lag length AIC, considering the Augmented Dickey-Fuller test statistic probability value if the Probability value is less than the set level even though the nonstationary is that this information is stationary

Step 2: Determine Autoregressive AR (P) and Moving Average MA (q) which is determined by Autocorrelation (ACF) and Partial Correlation (PACF). Later Estimation coefficient estimation of ARMA (p, q). And Check the significance of the t-test. consider the probability of the t-stat. If below the specified level of significance. We will reject the assumption that the coefficient is 0, ARIMA (p, q) cannot be used to describe the data set.

Step 3: Check the correctness. Diagnostic. Checking the inspection method consists of check the significance of the t-test coefficient.

Step 4: Forecasting Process When the time series is obtained Therefore began to forecast both in time and off-time (Static and Dynamic).

E. Assessing the Fit of Regression Models

This subsection shows three metric methods for ARIMA model assessment as follows; RMSE, MAPE, and U.

- **Root Mean Square Error (RMSE)**

RMSE is a value that shows the tolerance between the estimated value from the model and the actual data value. If the RMSE is near zero, this model has a small discrepancy and can be used to properly represent the actual data.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2} \quad (1)$$

$\hat{Y}_t$ is an estimate from the model.
Y_t is the real value, the real data.

- **Mean Absolute Percentage Error (MAPE)**

MAPE is a precision measurement tool for determining the time series that is statistically appropriate. It will show the percentage value. If the value is low, it shows that it is highly accurate. There are measurement criteria as follows [18]:

Very accurate in forecasting: MAPE less than 10%
High predictive accuracy: MAPE between 10% - 20%
Medium predictive accuracy: MAPE between 20% - 50%
Little predictive accuracy: MAPE over 50%

$$\text{MAPE} = \frac{[\sum|A_t - F_t| / A_t] \times 100}{N} \quad (2)$$

F_t is the forecast for the period of time t
A_t is the actual value during the period t
N is the amount of data

- **Theil's Inequality Coefficient (U)**

U will have a value between 0 and 1. If approaching 0 shows that this model can be used to represent real data appropriately. Let $\hat{Y}_t$ is an estimate from the model and Y_t is the real value, the real data.

$$\text{U} = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t)} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t)^2}} \quad (3)$$

IV. Analysis and Results

In this section, we explain the results of our research. Start from the process of converting data to stationary state which is returned the result by plotting the data, seasonal, trend, and remainder to stationary phase.

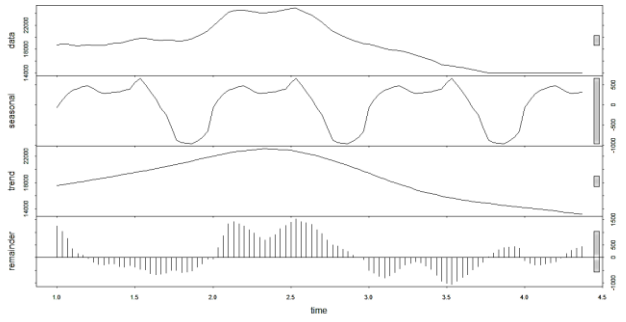


Fig. 6. Preparation process of datasource by converting data to stationary

According to data management, Fig 6 is displayed the first process of data management including gathering data, transforming data, and then load. The result of this state shows 4 stacks of graph from converting data to stationary. From top, 'data' graph is a graph that represents the time series of data to be forecasted. The second part is the 'seasonal' graph which is the graph of data classification into seasonal in order to bring the data to trend. The third part is the trend graph. The final is 'remainder' by adding an overview of the gold price, how the price is moving and to eliminate the abnormal data.

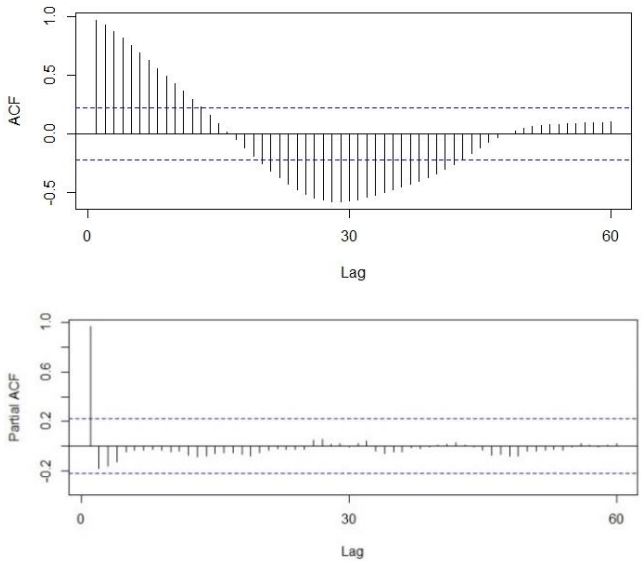


Fig. 7. ACF and PACF

Fig 7 displays that both ACF and PACF. They do not show significant truncation and tailing. Thus, they were refereed after stationary step of ARIMA models. The correlogram of ACF and PACF was plotted to identify the model of ARIMA. Both functions are applied to figure out decisive evidence of flatness state. If the ACF slowly decreases, PACF shows a pause after one lag, then this model is auto regressive model (AR). Alternatively, if ACF tends to be zero after one lag, PACF shows a zigzag decline, then this model is moving average (MA).

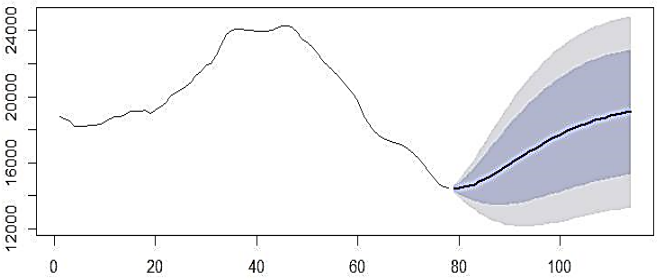


Fig. 8. Thailand gold price prediction

Fig 8 displays the movement of gold price in Thailand from the forecast. The black line represents the actual gold price. A divergent blue line shows the prediction of gold price. We can compare to Fig 9 in detail of gold price prediction in Thai Baht (axis-x) and Month (axis-y).

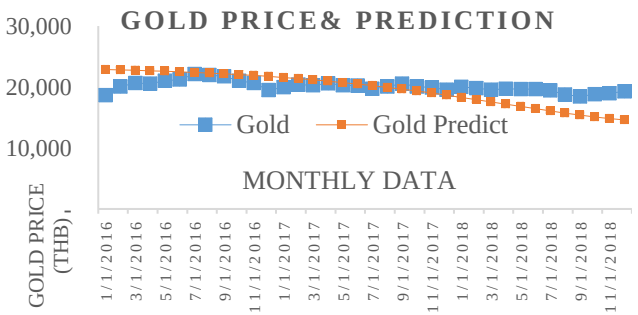


Fig. 9. Thailand gold price and prediction by monthly data between 2016-2018

Based on ARIMA model, TABLE III displays the error of predictions. When considering with RMSE, ARIMA (2,0,1) had the least error. However, considering with MAPE and U, ARIMA (2,0,0) had the least error. And ARIMA (1,0,0), ARIMA (1,0,1) cannot predict.

TABLE III. COMPARISON BETWEEN VARIOUS MODEL INDICATORS

Models	RMSE	MAPE	U
ARIMA (1,0,0)	N/A	N/A	N/A
ARIMA (2,0,0)	2.3333	9.12	0.1045
ARIMA (3,0,0)	0.9093	10.45	0.1148
ARIMA (1,0,1)	N/A	N/A	N/A
ARIMA (2,0,1)	0.6588	10.70	0.1169
ARIMA (3,0,1)	-20.8317	28.84	0.3254
ARIMA (1,1,0)	11.2342	11.95	0.1268
ARIMA (2,1,0)	11.5567	12.30	0.1304
ARIMA (3,1,0)	11.3382	12.09	0.1284
ARIMA (1,1,1)	11.5844	12.33	0.1306
ARIMA (2,1,1)	11.4601	12.21	0.1295
ARIMA (3,1,1)	11.3814	12.13	0.1287
ARIMA (4,1,1)	15.5654	15.88	0.1602
ARIMA (5,1,1)	13.5508	14.07	0.1454
ARIMA (6,1,1)	13.1155	13.69	0.1422
ARIMA (7,1,1)	13.1107	13.68	0.1424
ARIMA (8,1,1)	19.063	19.06	0.1889
ARIMA (9,1,1)	19.3263	19.3263	0.1912

When we compare Fig.8 with the detail of gold price by select the 80% of prediction to do an assessment phase, we found that ARIMA (2,0,0) has the most accuracy compared to other models. We found that ARIMA (2,0,0) has the smallest tolerance measured from MAPE and U. The results of assessment phase can be described as follow:

RMSE is a value that shows the error between the estimated values from the model and the actual data values. If the RMSE value is close to zero, it means that this model has a few mistakes. It can be used to represent real data in the future. The results are in Fig 10 which (2,0,1) is the lowest RMSE as 0.66.

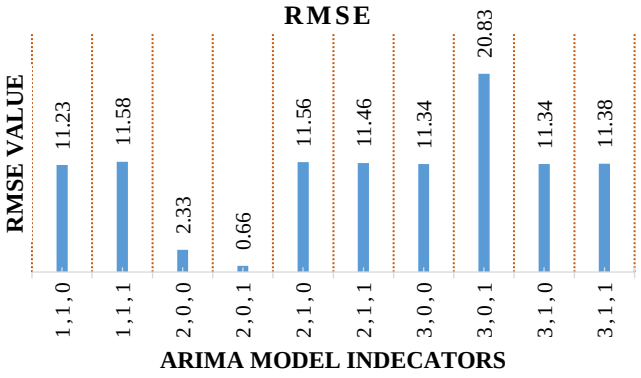


Fig. 10. RMSE

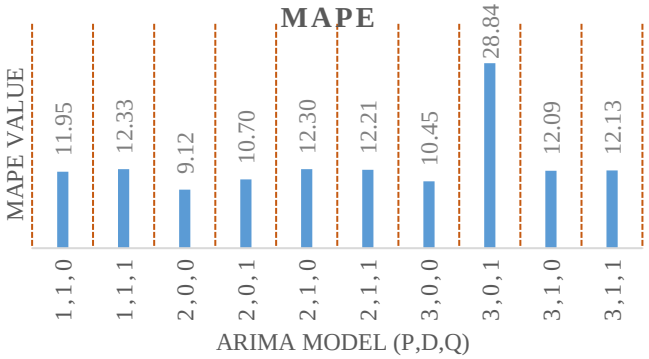


Fig. 11. MAPE comparison chart

Fig 11. is shown the model testing, it was found that the model with the lowest value. The top three are ARIMA (2,0,0) ARIMA (3,0,0) ARIMA (2,0,1) value are 9.12, 10.45 and 10.70 respectively. We checked only these values because from the value ARIMA(4,1,1) to ARIMA(9,1,1) is not accurate as in Table III.

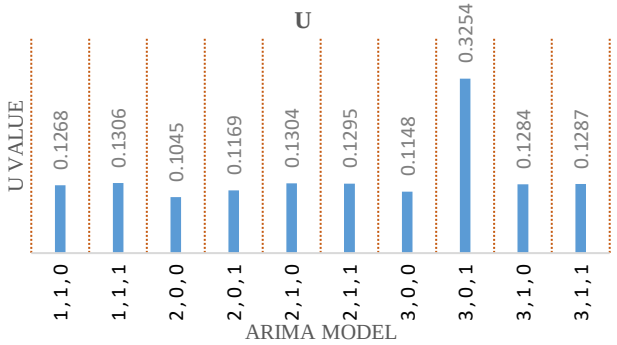


Fig. 12. U comparison chart

Fig 12 From the tests we found that the model with the top three lowest values are ARIMA (2,0,0) ARIMA (3,0,0) ARIMA (2,0,1) which are 0.1045, 0.1148 and 0.1169.

V. Conclusion

This study predicted the gold prices in Thailand based on various factors by ARIMA models in R programming language. The study found the most accurate and appropriate model for predicting with monthly time series as following results; ARIMA (2, 0, 0) appears to be the best model for predicting while ARIMA (3,0,0), ARIMA (2,0,1) has the second and third accuracy values. The models were assessed which, revealed the less error within acceptable criteria. This means that the result is satisfactory and acceptable. Future work is to predict the gold prices with GARCH model to compare ARIMA model performances for implementing of data visualization report.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Potjana Khamjring "Economic factors affecting changes in gold bar price in Thailand," Master of Economics, Ramkhamhaeng University, 2009.

[9] Pongsaton Ruktin, "Factors affecting gold prices in Thailand;" Master of Business Administration, Kasetsart University, 2009

[10] Parmont Innoung, "Factors that affect the price of gold in the Thai market," Master of Economics, Kasetsart University, 2009

[11] Saowarat Apirakdachachai "Factors influencing gold bar price in Thailand" Master of Economics, Ramkhamhaeng University, 2010.

[12] Supakon Ounyangovit "Factor determining the price of gold bar," Master of Economics, Kasetsart University, 2010

[13] Taweepong Saelim "Factors determining changes in price of gold spot," Master of Business Administration, Rangsit University, 2012.

[14] Yothin Komoltrakulwattana "Factors Affecting World Gold Price," Master of Economics, Sukhothai Thammathirat Open University, 2014.

[15] Phibulpanijkarn "The Analysis of Factors Influencing Gold Bar Price In Thailand," Master of Finance, Bangkok University, 2015

[16] Nuchjarin Kaothanthong "Factors Affecting Gold Import and Export of Thailand," Journal of Economics and Management Strategy, 2015

[17] Yaziz, S. R., et al. "The performance of hybrid ARIMA-GARCH modeling in forecasting gold price." 20th International Congress on Modelling and Simulation, Adelaide. 2013.

[18] Lewis, C.D. (1982). Industrial and business forecasting methods. London: Butterworths

The Development of an Alerting System for Spread of Brown planthoppers in Paddy Fields Using Unmanned Aerial Vehicle and Image Processing Technique

Worawut Yimyam
Department of Business Information Management
Phetchaburi Rajabhat University, Thailand
Phetchaburi , Thailand
worawut_yimyam@hotmail.com

Mahasak Ketcham
Department of Information Technology Management
King Mongkut's University of Technology North Bangkok,
Thailand
mahasak.k@it.kmutnb.ac.th

***Abstract*— This research aimed to analyze and design an alerting system for spread of Brown planthoppers in paddy using unmanned aerial vehicle and image processing technique. The research included reviewing and collecting data, designing unmanned aerial vehicle control system, testing flying system, designing color detection and comparing efficiency. To design color detection, image processing technique was employed. The research procedures consisted of receiving data, using 10 fps of an image of 640 x 480 pixel in a form of RGB and converting an image to HSV to adjust color values. The image was delivered to an inspection of images detected using Threshold. It could be summarized that the system could detect rice pests. Of the 50 tests, it was found that the difference of brown color was at 1.40%, of white color was at 0.73 % and of yellow color was at 0.00% in the paddy fields with rice pests. In addition, the error of brown color was at 0.01%, of white color was at 0.05% and of yellow color was at 0.00% in the paddy fields without rice pests.**

Keywords*— *Spread of Brown planthoppers, Unmanned Aerial Vehicle, Image Processing Technique

I. INTRODUCTION

Rice is a very important economic crop in Thailand as it is the main crop for the whole country and is the exported agricultural product earning not less than 1,40 billion baht each year to the country (Office of Agricultural Statistics, 2013). As the price of rice has steadily increased since 2007, it is an incentive for farmers in irrigation areas to have more demand for rice production. Since the area is limited, farmers have a way to increase rice production per year by planting rice several times in a year, resulting in an increase of growing areas for off-season rice from 9 million rai in 2016 to 14 million rai in 2012 [1]

Nowadays, farmers in the central lowlands plant rice three times a year instead of 2-1times a year. Planting rice is monoculture which requires large area. Rice is planted repeatedly and continually on the same area, leading to inevitable pest outbreaks. The pest classified as a serious problem of rice is Brown planthoppers [2]. It is found that there has been an outbreak since 1974 and the outbreak has become intensified and difficult to control, respectively. Especially in the off-season rice period from November 2009 to 2010, there has been a severe outbreak in the central region where more than 2.38 million rai were affected[3]. Brown planthoppers, both larvae and adult larvae, will destroy rice plants by absorbing nutrients in the tube of rice stalk, causing the leaves and stalks burnt, known as Hopper Burn. This causes rice plants to eventually die as shown in Figure 1. Formerly, the use of resistant rice varieties such as Pathum Thani 1 could effectively resist Brown planthoppers. However, planting such resistant rice varieties for a long period of consecutive seasons could cause Brown planthoppers to adapt themselves and eventually could destroy paddy fields [2].

Figure 1. Paddy fields destroyed by Brown planthoppers

The most popular method among farmers is using chemical control. However, there are still many problems due to the use of chemicals such as abamectin and cypermethrin causing more severe outbreaks. This is because these chemicals destroy only adult Brown planthoppers, unable to destroy eggs of Brown planthoppers. In addition, these chemicals can destroy natural enemies of Brown planthoppers so hatched eggs of Brown planthoppers do not have any natural enemies to control. This results in more severe outbreaks[4]. The efficiency control requires integrated management by managing water level by releasing water from paddy fields to reduce humidity, reducing planting rates so that rice plants are not tightly squeezed, selecting types of chemicals to use and selecting appropriate spraying period without spraying on rice younger than 40 days. However, such management will be effective when the destroy level has not reached the economic level and the outbreak has not expanded widely.

According to the problems mentioned, the researcher has developed an alerting system for spread of Brown planthoppers which applies unmanned aircraft technology and image processing technology to explore and monitor the spread of Brown planthoppers in paddy fields.

II. RELATED WORK AND THEORIES

A. *Concept of Digital Image Processing)*

- Basic Knowledge of Digital Image Processing

Digital images are images converted from analog data to digital data using Sampling and Quantization techniques for computer processing. A digital image is defined by two-dimension function f(x,y). The value of x,y and f(x,y) are limited values. Any)x,y(on an image is called Pixel. The value of f(x,y) is color intensity of that pixel. Humans use eyes as one of the important media for perceiving data around themselves. Eye receive data as images before processing them in the brain. Therefore, if a computer is able to perceive data as images via a video camera and process such images at the processing system, it will be another stage of computer's potential development.

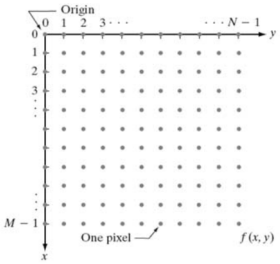


Figure 2. Coordinates of digital image system

Grayscale Image is an image stored by using a format of a 2D array. The value stored is in a certain range where the level of color depends on the size of the bit that stores color value.

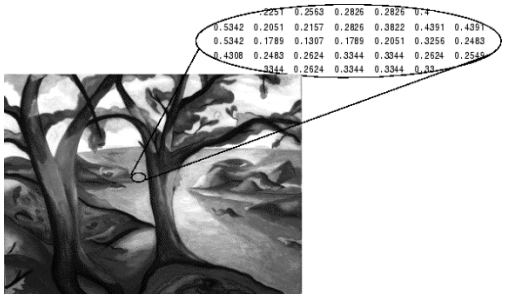


Figure 3. Grayscale image

- Grayscale

Grayscale is a color system in which each image point is represented by a color level from white to black, depending on color intensity from RGB color system. Grayscale system has color values that can be represented by 1-byte or 8-bit data providing 256-level resolution. The equation is as follows.

Grayscale = 0.11 * B + 0.59 * G + 0.30 * R (3)

Figure 4. Sample of converting colored image to grayscale image

- Threshold

Threshold is a value used to divide grayscale system into two levels consisting of white and black. It is used for separating an image part from a background according to color intensity compared to threshold value. If the color value of that image is lower than the threshold value specified, that point will be represented by black level. If the color value of that image is greater than the threshold value, that point is represented by white level as in Figure 2.13. Using the threshold value to find objects within an image is suitable for an image containing objects clearly different from the background. When considering a histogram, there will be a clear difference in height value.

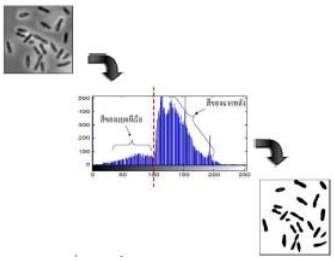


Figure 5. Converting grayscale image to black and white image

III. PROPOSE METHODOLOGY

The procedures are divided into 7 steps as shown in the following diagram.

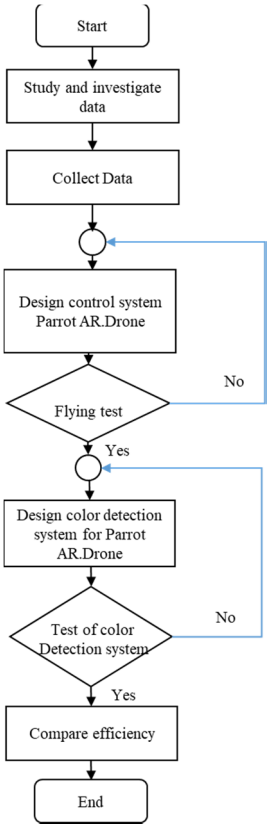


Figure 6. Flowchart

According to Figure 6 Flowchart, the procedures are as follows.

Step 1 Study and investigate data: From the beginning, it is found that there is necessity of study extensive data starting from looking at the important problem in Thailand. Such problem is an agricultural problem, especially rice farming and rice products. After that, the existing and newly-developed solutions to problems as well as technologies that can be beneficially developed are explored.

Step 2 Collect data: After studying the data, all relevant data was collected for designing and controlling unmanned aerial vehicle.

Step 3 Design control system of Parrot AR. Drone: Design and control system are important as it enables AR. Drone to work efficiently. Flying directions are controlled including taking off/landing, left rotating, right rotating, moving forward, moving backward, moving up, moving down and switching a camera, taking photos and checking batteries.

Step 4 Flying test: From Step 3, the researcher tested flying using direction control buttons specified. If there was an error, it would return to solve such error in step 3. If the flight control was normal, it would enter the design phase of color detection system.

Step 5 Design color detection system: In this step, image processing principle was employed to detect colors of different rice leaves. The threshold values were applied to separate different colors to identify status of rice leaves, showing whether they were normal or they had rice pests. If so, what rice pests they had.

Step 6 Test of color detection system: Different rice leaves were tested to detect different color values. If the system could not detect any colors, the color detection system must be re-designed. If the system could detect color, it would proceed to a comparison of efficiency phase.

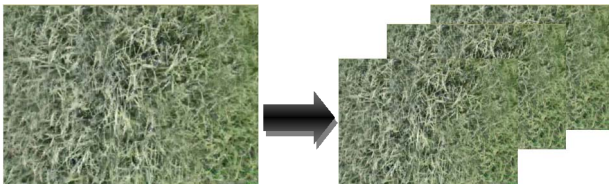
Step 7 Compare efficiency: Image processing was employed to compare efficiency of individual status. The results could identify rice leaves in which areas had rice pests and could reveal rice pests accordingly.

*A. **Design of Color Detection System Display***

Designing color detection system in paddy fields which is a process that processes photos taken by unmanned aerial vehicle in image processing. The procedures are as follows.

- Data reception

In the detection system, a video or an image from an unmanned aerial vehicle camera is received, it is necessary to separate the video into frames in order to detect each frame for rice diseases. It depends on a Frame Rate used to record. In this study, 10 Frames per second (fps) was employed. That is, 10 images can be separated in 1 second. The image size is 640x480 pixel based on the size of video file. The images received from separating frames or from a camera would be in RGB images.



Video File Frames

Figure 7. Separating images in a video file

- Image conversion

It is a conversion of an RGB color image to an HSV image so that it is convenient for the adjustment of color values as shown in (1)(2)(3)

- Hue can be calculated from RGB color system as follows.

$$red_k = red - \min(red, green, blue)$$
$$green_k = green - \min(red, green, blue) \quad (1)$$
$$blue_k = blue - \min(red, green, blue)$$

- Saturation is color purity. If Saturation is equal 0, the color will not have Hue. It can be calculated as follows.

$$Saturation = \frac{\max(red, green, blue) - \min(red, green, blue)}{\max(red, green, blue)} \quad (2)$$

- Value is color brightness which can be measured by the intensity of brightness of each color. It can be calculated from

-

$$value = \max(red, green, blue) \quad (3)$$

- Inspection of detected images

It can be proceeded by setting the threshold values. After receiving the area of detected image, such area is cut off to re-check whether it shows rice diseases in paddy fields or not. For inspection, the cut image is proceeded through the threshold color image with a value of 100. The image in each Color Channel will go through Threshold to adjust colors to have more prominent color characteristics. Unclear colors are deleted.

Then, the image is brought to separate its color values in each Color Channel. The values with bright color including red, green, and blue are adjusted to be black. The image does not contain any white parts. After that, red, green and blue pixel values are counted to calculate the values necessary for rice disease detection system, using various image processing techniques. The following values are applied.

- Percentage of brown, white and yellow pixels per total number of pixels has the following formula:

$$\begin{bmatrix} \text{Brown}[\%] = (\text{Brown\_pixel}/\text{All\_pixel})*100 \\ \text{White}[\%] = (\text{White\_pixel}/\text{All\_pixel})*100 \\ \text{Yello}[\%] = (\text{Yello\_pixel}/\text{All\_pixel})*100 \end{bmatrix} \quad (4)$$

- Percentage of different color pixels per total number of brown, white and yellow pixels has the following formula:

$$\begin{bmatrix} \text{Brown\_c}[\%] = (\text{Red\_pixel}/\text{Color\_pixel})*100 \\ \text{White\_c}[\%] = (\text{White\_pixel}/\text{Color\_pixel})*100 \\ \text{Yello\_c}[\%] = (\text{Yello\_pixel}/\text{Color\_pixel})*100 \end{bmatrix} \quad (5)$$

- The ratio of different colors to the number of pixels of others colors has the following formula:

$$\begin{bmatrix} \text{Brown\_ratio} = (\text{Brown\_pixel}/\text{White\_pixel})*100 \\ \text{Yello\_ratio} = (\text{Yello\_pixel}/\text{White\_pixel})*100 \\ \text{BY\_ratio} = (\text{Brown\_pixel}/\text{Yello\_pixel})*100 \end{bmatrix} \quad (6)$$

The color values of rice diseases in paddy fields can be identified by collecting a large number of images of each pest types in different conditions of light, colors and sizes. All values are calculated as shown above and statistics are collected to find outstanding color characteristics of paddy fields by considering the characteristics of each image type. The details are as follows.

- Brown planthopper is an image consisting of brown, white and yellow parts in the ratio of brown and white to the whole area. The ratio of both brown color and white color must always be greater than yellow color. However, there shall be some white color.

- Rice leaffolder is an image mainly consisting of brown color and white color. Therefore, rice leaffolder has the white color which is more outstanding than other colors, which is about 80% of the total paddy fields. It may consist of some brown color or yellow color.

Different types of pests are grouped and testes with video samples in order to check the data efficiency collected until the right values are retrieved. The error images which are not images of pests in paddy fields are filtered out. Then, the values of position of the filtered images are returned to the display.

- Summary of rice diseases in paddy fields

When positions, sizes and image types are identified, the images are analyzed and displayed in a form of percentage of damage of paddy fields which will be used to find ways to prevent the occurrence of rice pests to prevent further damage.

IV. THE EXPERIMENTAL AND RESULT

The researcher tested the color differences of paddy fields by using image processing to reduce damage to paddy fields caused by pests. The researcher also tested the efficiency of color difference detection program of rice leaves in paddy fields in order to investigate the occurrence of brown planthopper and rice leaffolder problems in paddy fields.

The first step was to test the system with a model. When the desired results were achieved, the actual location was tested with 2 types of paddy fields as follows.

- Test of paddy fields without rice pest problems
- Test of paddy fields with rice pest problems

A. Test of paddy fields without rice pest problems

According to the test of efficiency of the color difference detection program used with paddy fields without rice pest problems, the results are as follows.

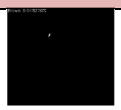
- The results of system when tested in paddy fields without rice pest problems

Based on the test of 50 images of paddy fields without rice pest problems, it was found that brown color value had error at 0.01% of the total area. The error value was caused by the environment where it had brown color, resulting in similar color of brown planthoppers. The brown color included ground, watercourse and noise of image.

The error value of white color was at 0.05% of the total area. The error value was caused by colors similar to the color of rice leaffolder including the reflection of rice leaves and noise of image.

There was no error value of yellow color. The error value was at 0.00%.

TABLE I. RESULTS OF PROGRAM TESTED IN PADDY FIELDS WITHOUT RICE PEST PROBLEMS

Actual image	Brown color value	White color value	Yellow color value
			
	0.01%	0.13%	0.00%
			
	0.03%	0.01%	0.00%
			
	0.01%	0.13%	0.00%
			
	0.01 %	0.03%	0.00%
			
	0.01%	0.01%	0.00%

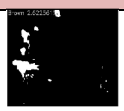
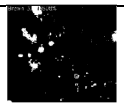
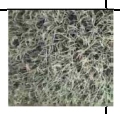
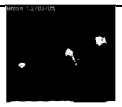
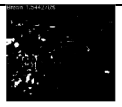
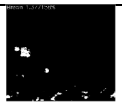
2. Test of paddy fields with rice pest problems

According to the test of efficiency of the color difference detection program used with paddy fields with rice pest problems, the results are as follows.

The results of system when tested in paddy fields with rice pest problems

TABLE II. RESULTS OF PROGRAM TESTED IN PADDY FIELDS WITH RICE PEST PROBLEMS

Based on the test of 50 images, it was found that the

Actual image	Brown color value	White color value	Yellow color value
			
	2.62%	0.11%	0.00%
			
	3.01%	0.00%	0.00%
			
	1.27%	0.20%	0.00%
			
	1.54%	0.05%	0.00%
			
	1.38%	0.07%	0.00%

difference value of brown color was at 1.40%, of white color was 0.73% and of yellow color was at 0.00%. According to these values, it was found that the program could detect values of rice leaves containing rice diseases. That is, the rice leaves in the experimental areas with brown color had the characteristics that matched brown planthopper problem as hypothesized. The rice leaves in the experimental areas with white lines, especially at the tip of the leaves which is different from the rice leaves in the normal condition which had green color and the rice leaves in the condition of brown planthoppers which had brown color. Therefore, such paddy fields had the characteristics corresponding to rice leaffolder problem as hypothesized.

V. CONCLUSION

The development of an alerting system for spread of brown planthoppers in paddy fields using unmanned aerial vehicle and image processing technique aimed to analyze and design an alerting system for spread of brown planthoppers in paddy using image processing technique and unmanned aerial vehicle. The system tested in paddy fields without rice pest problems and paddy fields with rice pest problems. It was found that the system did not detect brown planthopper and rice leaffolder problems in paddy fields without rice pest problems. The error value was at 0.01% with the characteristics that matched brown planthoppers. It also detected brown planthopper problems with the error value at 1.40%. Therefore, the system could evaluate damage caused by rice pests.

REFERENCES

[1] Office of Agricultural Statistics, 2012

[2] W.rattanasak, S.Aruymit and J.Chaiwang. “Situation of the brown planthopper in Thailand” 2011. Academic Conference 2011 at Amari Hotel, Don Mueang Airport.

[3] Department of Rice, 2010

[4] K. Soithong. (2011) Outbreaks of brown planthopper and the use of pesticides in rice production in the central and lower northern regions.

[5] C.Pornpanomchai, S.Rimdusit, P.Tanasap and C.Chaiyod. “Thai Herb Leaf Image Recognition System (THLIRS).” Nat. Sci., pp. 551-562, 2005.

[6] Chumuang N., Ketcham M., Sawatnatee A. (2019) Criminal Background Check Program with Fingerprint. In: Theeramunkong T. et al. (eds) Advances in Intelligent Informatics, Smart Technology and Natural Language Processing. iSAI-NLP 2017. Advances in Intelligent Systems and Computing, vol 807. Springer, Cham.

[7] S. Thaiparnit, N. Khuadthong, N. Chumuang and M. Ketcham, "Tracking Vehicles System Based on License Plate Recognition," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 220-225. doi: 10.1109/ISCIT.2018.8588008

[8] S. Thaiparnit, N. Chumuang and M. Ketcham, "Weapon Detector System by Using X-ray Image Processing Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 214-219. doi: 10.1109/ISCIT.2018.8587853

[9] B. Narin, S. Buntan, N. Chumuang and M. Ketcham, "Crack on Eggshell Detection System Based on Image Processing Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 1-6. doi: 10.1109/ISCIT.2018.8587980

[10] N. Chumuang, P. Chansuek, M. Ketcham, A. Silsanpisut, T. Ganokratanaa and P. Selarat, "Analysis of X-ray for locating the weapon in the vehicle by using scale-invariant features transform," 2017 Fourth Asian Conference on Defence Technology - Japan (ACDT), Tokyo, 2017, pp. 1-6. doi: 10.1109/ACDTJ.2017.8259599

[11] S. Suwannakhun, N. Chumuang and M. Ketcham, "Identification and Retrieval System by Using Face Detection," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 294-298. doi: 10.1109/ISCIT.2018.8587856

[12] Lowe D.G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60(2): 91–110.

[13] Cecilia Di Ruberto and Lorenzo Putzu. “A Fast Leaf Recognition Algorithm based on SVM Classifier and High Dimensional Feature Vector.” 2014 International Conference on Computer Vision Theory and Applications (VISAPP), 2557.

[14] Xinhong Zhang and Fan Zhang. “Images Features Extraction of Tobacco Leaves.” 2008 Congress on Image and Signal Processing, 2551: 773-776

[15] T. Yingthawornsuk, N. Chumuang and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 418-423. doi: 10.1109/SITIS.2017.75

[16] W. Yimyam and M. Ketcham, “The automated parking fee calculation using license plate recognition system.” In 2017 International Conference on Digital Arts, Media and Technology (ICDAMT) (pp. 325-329). IEEE.

[17] M. Ketcham, W. Yimyam, and N. Chumuang, “ Segmentation of overlapping Isan Dhamma character on palm leaf manuscript’s with neural network”. In Recent Advances in Information and Communication Technology 2016 (pp. 55-65). Springer, Cham.

[18] S. Phatchuay, and W. Yimyam, “The System Vehicle of Application Detector for Categorize Type”. In 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 683-687). IEEE.

[19] W. Yimyam and M. Ketcham, “The Electroencephalography Signals Using Artificial Neural Network for Monitoring Fatigue System.” In Pacific Rim International Conference on Artificial Intelligence (pp. 160-169). Springer, Cham.

[20] T. Pinthong, W. Yimyam, N. Chumuang, and M. Ketcham, “License Plate Tracking Based on Template Matching Technique.” In 2018 18th

International Symposium on Communications and Information Technologies (ISCIT) (pp. 299-303). IEEE.

[21] W. Yimyam and M. Ketcham, "Video Surveillance System Using IP Camera for Target Person Detection." In 2018 18th International Symposium on Communications and Information Technologies (ISCIT) (pp. 176-179). IEEE.

[22] W. Yimyam and M. Ketcham, "The Grading Multiple Choice Tests System via Mobile Phone using Image Processing Technique." International Journal of Emerging Technologies in Learning (iJET), 13(10), 260-269.

[23] W. Yimyam and M. Ketcham, "The System for Driver Fatigue Monitoring Using Decision Tree via Wireless Sensor Network for Intelligent Transport System." International Journal of Online Engineering (iJOE), 14(10), 21-39.

[24] T. Pinthong and W. Yimyam, "The Model of Teenager's Internet Usage Behavior Analysis Using Data Mining ". In The Joint International Symposium on Artificial Intelligence and Natural Language Processing (pp. 196-203). Springer, Cham.

[25] W. Yimyam and M. Ketcham, "Eye Region Detection in Fatigue Monitoring for the Military Using AdaBoost Algorithm". In International Symposium on Natural Language Processing (pp. 151-161). Springer, Cham.

Bandit Multiclass Linear Classification for the Group Linear Separable Case

***Abstract*—We consider the online multiclass linear classification under the bandit feedback setting. Beygelzimer, Pal, Szorenyi, Thiruvenkatachari, Wei, and Zhang [ICML'19] considered two notions of linear separability, weak and strong linear separability. When examples are strongly linearly separable with margin γ, they presented an algorithm based on MULTICLASS PERCEPTRON with mistake bound $O(K/\gamma^2)$, where K is the number of classes. They employed rational kernel to deal with examples under the weakly linearly separable condition, and obtained the mistake bound of $\min(K \cdot 2^{\tilde{O}(K \log^2(1/\gamma))}, K \cdot 2^{\tilde{O}(\sqrt{1/\gamma} \log K)})$. In this paper, we refine the notion of weak linear separability to support the notion of class grouping, called group weak linear separable condition. This situation may arise from the fact that class structures contain inherent grouping. We show that under this condition, we can also use the rational kernel and obtain the mistake bound of $K \cdot 2^{\tilde{O}(\sqrt{1/\gamma} \log L)})$, where $L \leq K$ represents the number of groups.**

***Index Terms*—multiclass, bandit, linear separable, kernel, group linear separable**

I. INTRODUCTION

In an online-learning paradigm, at each time step t, the learner receives, a feature vector x_t, makes a prediction $\hat{y}_t$, and obtains a feedback. Note that the learner is playing against an adversary who picks the vector x_t and the correct class y_t from a set of K classes. In the standard *full-information feedback setting*, the feedback is the correct class y_t, while in the *bandit feedback setting*, the only feedback is a binary indicator specifying if the learner makes the correct prediction, i.e., $\mathbb{1}[\hat{y}_t = y_t]$. The performance of the learner is measured by the total number of mistakes over all the steps.

Typically, the theoretical analysis is carried out under particular linear separability with margin assumptions. Beygelzimer, Pal, Szorenyi, Thiruvenkatachari, Wei, and Zhang [1] introduced two definitions of linear separability, called *strong* and *weak* linear separability. We give a brief summary here (see formal definitions in Section II-A). For both definitions, there are K vectors w_i defining K hyperplanes. The weak linear separable condition which is similar to standard multiclass linear separability defined in Crammer and Singer [2] ensures that examples from each class lie in the intersection of K halfspaces induced by these hyperplanes. The strong linear separable condition requires that each class is separated by a single hyperplane.

In the full-information feedback setting, Crammer and Singer [2] showed that if all examples are weakly linear separable with margin γ and have norm at most R, the MULTICLASS PERCEPTRON algorithm makes at most $\lfloor 2(R/\gamma)^2 \rfloor$ mistakes. This is tight (up to a constant) since any algorithms must make at least $\frac{1}{2}\lfloor (R/\gamma)^2 \rfloor$ mistakes in the worst case.

For the bandit feedback setting [3], Beygelzimer *et al.* [1] presented an algorithm that make at most $O(K(R/\gamma)^2)$ if the examples are strongly linear separable with margin γ, paying the price of a factor of K for the bandit feedback setting. They also showed how to extend the algorithm to work with weakly linear separable case using the kernel approach. More specifically, they (non-linearly) transform the examples to higher dimensional space so that the examples are strongly linear separable with margin γ' (which depends only on γ and K).

In this paper, we introduce a more refined linear separability condition. Intuitively, the set of weight vectors w_i represents the "directions" of the examples. In this paper, we are interested in the cases where these directions collapsed, i.e., while there are K classes of examples, the number of distinct weight vectors required to linearly separate them is less than K. This situation may arise from the fact that class structures contain inherent grouping where intra-group classes can be separated with a single weight vector (or direction). (See Fig. 1, for example.)

More specifically, we consider the case where the classes can be partitioned into L groups, where $L \leq K$, such that (1) examples from any two classes in the same group are linearly separable with a margin with a single weight vector, and (2) examples from two classes under different groups are weakly linear separable with a margin. We refer to this condition as the *group weakly linear separable condition*.

We show that under this refined condition, the same kernel as in [1] can also be used so that the algorithm works in the space where there is (strong) margin γ' that depends on L. Our proofs, as well as that of [1], use the ideas from Klivans and Servedio [4] (which is also based on Beigel *et al.* [5]).

We note that our key contribution is the mathematical analysis of the margin for group weakly linearly separable examples for the kernelized algorithm in Beygelzimer *et al.*. This means that everything in their paper works under this group condition (with a better margin bound that depends on L not K).

Section II gives definitions and problem settings. Our main result is in Section III. In particular, Section III-B contains our technical theorem that establish the margin under the transformed inner product space. We provide small examples in Section IV.

II. Definitions and problem settings

A. Linear separability

We restate the definitions for strong and weak linear separability by Beygelzimer *et. al.* [1] here. We use the common notation that $[K] = \{1, 2, \ldots, K\}$.

The examples lie in an inner product space $(V, \langle \cdot, \cdot \rangle)$. Let K be the number of classes and let γ be a positive real number. Labeled examples

$$(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T) \in V \times [K]$$

are *strongly linear separable with margin* γ if there exist vectors $w_1, w_2, \ldots, w_K \in V$ such that for all $t \in [T]$,

$$\langle x_t, w_{y_t} \rangle \geq \gamma/2,$$

and

$$\langle x_t, w_i \rangle \leq -\gamma/2,$$

for $i \in [K] \setminus \{y_t\}$, and $\sum_{i=1}^{K} \|w_i\|^2 \leq 1$.

On the other hand, the labeled examples are *weakly linear separable with margin* γ if there exist vectors $w_1, w_2, \ldots, w_K \in V$ such that for all $t \in [T]$,

$$\langle x_t, w_{y_t} \rangle \geq \langle x_t, w_i \rangle + \gamma,$$

for $i \in [K] \setminus \{y_t\}$, and $\sum_{i=1}^{K} \|w_i\|^2 \leq 1$.

The strong linear separability also appears in Chen *et al.* [6]. The weak linear separable condition appears in Crammer and Singer [2].

We now define group weakly linear separability. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_L\}$ be a partition of $[K]$, i.e., $G_i \subseteq [K]$ for all i, $G_i \cap G_j = \emptyset$ for $i \neq j$, and $\bigcup G_i = [K]$. Let $g : [K] \to [L]$ be a mapping function such that $g(i) \mapsto j$ iff $i \in G_j$. We say that the labeled examples

$$(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T) \in V \times [K]$$

are *group weakly linear separable with margin* γ *under* $\mathcal{G}$ if

1) there exist vectors $u_1, u_2, \ldots, u_L \in V$ such that $\sum_{i=1}^{L} \|u_i\|^2 \leq 1$, and, for all $t \in [T]$,

$$\langle x_t, u_{g(y_t)} \rangle \geq \langle x_t, u_p \rangle + \gamma,$$

for all $p \in [L] \setminus \{g(y_t)\}$,

2) there exist vectors $u'_1, u'_2, \ldots, u'_L \in V$ such that $\sum_{i=1}^{L} \|u'_i\|^2 \leq 1$, and, for all $t \in [T], t' \in [T]$ such that $y_t \neq y_{t'}$ and $g(y_t) = g(y_{t'})$, either

$$\langle x_t, u'_{g(y_t)} \rangle \geq \langle x_{t'}, u'_{g(y_t)} \rangle + \gamma,$$

or

$$\langle x_t, u'_{g(y_t)} \rangle \leq \langle x_{t'}, u'_{g(y_t)} \rangle - \gamma.$$

Note that vectors u_i's define inter-group hyperplanes, while each u'_i defines intra-group boundaries.

To illustrate the idea, Fig. 1 shows 3 sets of examples.

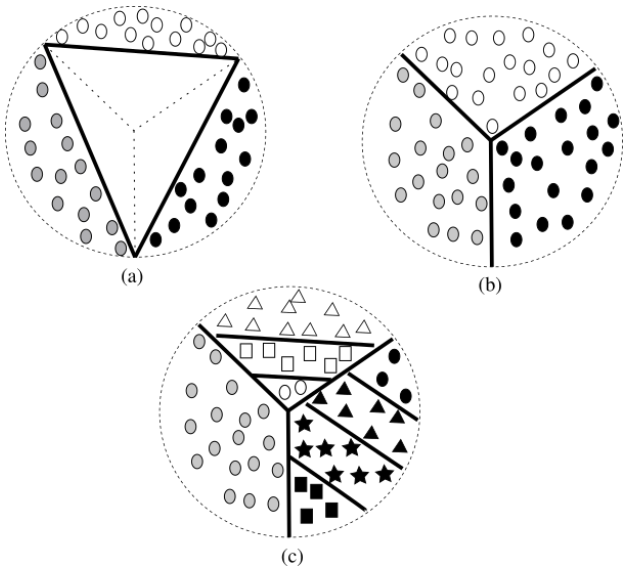


Fig. 1. Three set of examples in $\mathbb{R}^2$ showing different linear separable conditions. Thick lines represent class boundaries. (a) Strongly linear separable examples with 3 classes (linearly separable in $\mathbb{R}^3$). (b) Weakly linear separable examples with 3 classes. (c) Group weakly linear separable examples with 3 groups; group 1 (white) contains 3 classes, group 2 (black) contains 4 classes, and group 3 (gray) contains 1 class.

B. Kernel methods

We give an overview of the kernel methods (see [7] for expositions) and the rational kernel [8].

The kernel method is a standard approach to extend linear classification algorithms that use only inner products to handle the notions of "distance" between pairs of examples to nonlinear classification. A *positive definite kernel* (or *kernel*) is a function of the form $k : X \times X \to \mathbb{R}$ for some set X such that the matrix $[k(x_i, x_j)]_{i,j=1}^{m}$ is symmetric positive definite for any set of m examples $x_1, x_2, \ldots, x_m \in X$. It is known that for every kernel k, there exists some inner product space $(V, \langle \cdot, \cdot \rangle)$ and a feature map $\phi : X \to V$ such that $k(x, x') = \langle \phi(x), \phi(x') \rangle$. Therefore, a linear learning algorithm can essentially non-linearly map every example into V and work in V instead of the original space without explicitly working with ϕ using k. This can be very helpful when the dimension of V is infinite.

As in Beygelzimer *et al.* [1], we use the rational kernel. Assume that examples are in $\mathbb{R}^d$. Denote by $\mathrm{B}(0, 1)$ a unit ball centered at 0 in $\mathbb{R}^d$. The *rational kernel* $k : B(0, 1) \times B(0, 1) \to \mathbb{R}$ is defined as

$$k(x, x') = \frac{1}{1 - \frac{1}{2}\langle x, x' \rangle_{\mathbb{R}^d}}.$$

Given $x, x' \in \mathbb{R}^d$, $k(x, x')$ can be computed in $O(d)$ time.

Let $\ell_2 = \{x \in \mathbb{R}^\infty : \sum_{i=1}^{\infty} x_i^2 < +\infty\}$ be the classical real separable Hilbert space equipped with the standard inner product $\langle x, x' \rangle_{\ell_2} = \sum_{i=1}^{\infty} x_i x'_i$. We can index the coordinates of ℓ_2 by d-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_d)$ of non-negative integers,

the associated feature map $\phi : \mathrm{B}(0,1) \to \ell_2$ to k is defined as

$$(\phi(x_1, x_2, \ldots, x_d))_{(\alpha_1,\alpha_2,\ldots,\alpha_d)} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d} \cdot \sqrt{2^{-(\alpha_1+\alpha_2+\cdots+\alpha_d)} \binom{\alpha_1+\alpha_2+\cdots+\alpha_d}{\alpha_1,\alpha_2,\ldots,\alpha_d}}, \tag{1}$$

where $\binom{\alpha_1+\alpha_2+\cdots+\alpha_d}{\alpha_1,\alpha_2,\ldots,\alpha_d} = \frac{(\alpha_1+\alpha_2+\cdots+\alpha_d)!}{\alpha_1!\alpha_2!\cdots\alpha_d!}$ is the multinomial coefficient. It can be verified that k is the kernel with its feature map ϕ to ℓ_2 and for any $x \in \mathrm{B}(0,1)$, $\phi(x) \in \ell_2$.

C. Multiclass Linear Classification

Beygelzimer *et al.* [1] presented a learning algorithm for the strongly linearly separable examples based using K copies of the BINARY PERCEPTRON. They obtained a mistake bound of $O(K(R/\gamma)^2)$ when the examples are from $\mathbb{R}^d$ with maximum norm R with margin γ.

Their approach for dealing the weakly linear separable case is to use the kernel method. They introduced the KERNELIZED BANDIT ALGORITHM (Algorithm 1) and proved the following theorem.

Theorem 1 (Theorem 4 from [1])**.** *Let X be a non-empty set, let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Let $\phi : X \to V$ be a feature map and let $k : X \times X \to \mathbb{R}$, where $k(x,x') = \langle\phi(x), \phi(x')\rangle$, be the kernel. If $(x_1,y_1),(x_2,y_2),\ldots,(x_T,y_T) \in X \times \{1,2,\ldots,K\}$ are labeled examples such that*

1) *the mapped examples $(\phi(x_1),y_1),\ldots,(\phi(x_T),y_T)$ are strongly linearly separable with margin γ,*
2) *$k(x_1,x_1), k(x_2,x_2),\ldots,k(x_T,x_T) \le R^2$*

then the expected number of mistakes that the KERNELIZED BANDIT ALGORITHM *makes is at most $(K-1)\lfloor 4(R/\gamma)^2 \rfloor$.*

The key theorem for establishing the mistake bound is the following margin transformation theorem based on the rational kernel.

Theorem 2 (Theorem 5 from [1])**.** *(Margin transformation from [1]). Let $(x_1,y_1),(x_2,y_2),\ldots,(x_T,y_T) \in \mathrm{B}(0,1) \times [K]$ be a sequence of labeled examples that is weakly linear separable with margin $\gamma > 0$. Let ϕ defined as in (1) let*

$$\gamma_1 = \frac{\left[376\lceil\log_2(2K-2)\rceil \cdot \left\lceil\sqrt{\frac{2}{\gamma}}\right\rceil\right]^{\frac{-\lceil\log_2(2K-2)\rceil \cdot \left\lceil\sqrt{2/\gamma}\right\rceil}{2}}}{2\sqrt{K}},$$

$$\gamma_2 = \frac{\left(2^{s+1} r(K-1)(4s+2)\right)^{-(s+1/2)r(K-1)}}{4\sqrt{K}(4K-5)2^{K-1}}$$

where $r = 2\lceil\frac{1}{4}\log_2(4K-3)\rceil + 1$ and $s = \lceil\log_2(2/\gamma)\rceil$. Then the feature map ϕ makes the sequence $(\phi(x_1),y_1),(\phi(x_2),y_2),\ldots,(\phi(x_T),y_T)$ strongly linearly separable with margin $\gamma' = \max\{\gamma_1,\gamma_2\}$. Also for all t, $k(x_t,x_t) \le 2$.

This implies the following mistake bound.

Data: Number of classes K, number of rounds T
Data: Kernel function $k(\cdot,\cdot)$
begin
  Initialize $J_1^{(1)} = J_2^{(2)} = \cdots = J_k^{(k)} = \emptyset$
  for $t = 1,2,\ldots,T$ **do**
    Observe feature vector x_t
    Compute $S_t = \left\{i : 1 \le i \le K, \sum_{(x,y)\in J_i^{(t)}} yk(x,x_t) \ge 0\right\}$
    if $S_t = \emptyset$ **then**
      Predict $\hat{y}_t \sim$ Uniform$(\{1,2,\ldots,K\})$
      Observe feedback $z_t = \mathbb{1}[\hat{y}_t \ne y_t]$
      if $z_t = 1$ **then**
        Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1,2,\ldots,K\}$
      else
        Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1,2,\ldots,K\} \setminus \{\hat{y}_t\}$
        Update $J_{\hat{y}_t}^{(t+1)} = J_{\hat{y}_t}^{(t)} \cup \{(x_t,+1)\}$
      end
    else
      Predict $\hat{y}_t \in S_t$ chosen arbitrarily
      Observe feedback $z_t = \mathbb{1}[\hat{y}_t \ne y_t]$
      if $z_t = 1$ **then**
        Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1,2,\ldots,K\} \setminus \{\hat{y}_t\}$
        Update $J_{\hat{y}_t}^{(t+1)} = J_{\hat{y}_t}^{(t)} \cup \{(x_t,-1)\}$
      else
        Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1,2,\ldots,K\}$
      end
    end
  end
end

Algorithm 1: KERNELIZED BANDIT ALGORITHM [1]

Corollary 1 (Corollary 6 from [1])**.** *(Mistake upper bound from [1]). The mistake bound made by Algorithm 1 when the examples are weakly linearly separable with margin γ is at most* $\min(2^{\tilde{O}(K\log^2(1/\gamma))}, 2^{\tilde{O}(\sqrt{1/\gamma}\log K)})$.

D. Our contribution

We consider labeled examples with group weakly linearly separable with margin γ and show that in this case, the rational kernel also transforms the margin and the new margin depends on the number of groups L instead of the number of classes K. More specifically we prove the margin transformation in Theorem 3 and show the mistake bound of $K \cdot 2^{\tilde{O}(\sqrt{1/\gamma}\log L)}$ in Corollary 2. This can be compared to one of the mistake bound of $K \cdot 2^{\tilde{O}(\sqrt{1/\gamma}\log K)}$ in [1].

The proofs are fairly technical. We follow the idea in [1] and construct a "good" polynomial that separate examples from one class to the other (strong separation) based on the Chebyshev polynomials [9].

III. Main result

Our main technical result is the following margin transformation using the rational kernel.

Theorem 3. *(Margin transformation). Let* $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T) \in \mathrm{B}(0,1) \times [K]$ *be a sequence of labeled examples that is group weakly linear separable with margin* $\gamma > 0$*. Let* L *be number of group weakly separable such that* $L \leq K$*. Let* ϕ *defined as in (1) let*

$$\gamma = \frac{\frac{7}{2}\left[188\lceil\log_2(2L+1)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{-\frac{\lceil\log_2(2L+1)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}}{2\sqrt{L}},$$

The feature map ϕ *makes the sequence* $(\phi(x_1), y_1), (\phi(x_2), y_2), \ldots, (\phi(x_T), y_T)$ *strongly linearly separable with margin* γ*.*

We note that the margin depends on L, the number of groups, instead of K, the number of classes. Using Theorem 3 with Theorem 1 we obtain the following mistake bound for our algorithm.

Corollary 2. *(Mistake bound for group weakly linearly separable case) Let* K *be positive integer,* $L \leq K$ *and* γ *be positive real number. The mistake bound made by Algorithm 1 when the examples are group weakly linearly separable with margin* γ *with* L *groups is at most* $K \cdot 2^{\tilde{O}(\sqrt{1/\gamma \log L})}$*.*

Note that multiplicative factor of K is hidden from the second bound of [1] because of the $\tilde{O}$ notation on the exponent. We cannot do that because in our exponent we have only $\log L$ which can be much smaller than K. Their actual bound (showing K), which can be compared to ours, is $K \cdot 2^{\tilde{O}(\sqrt{1/\gamma \log K})}$.

A. Margin transformation

This section is devoted to the proofs of Theorem 3. To proof Theorem 3, we need to establish the existence of polynomials that separate one class from the other as shown in the Theorem 4 below. This theorem is proved in Section III-B.

Theorem 4. *(Polynomial approximation of intersection of halfspaces) Let* $v_1, v_2, \ldots, v_m \in V$ *such that* $\|v_1\|, \|v_2\|, \ldots, \|v_m\| \leq 1$*. Let* $v_a, v_b, u \in V$ *such that* $\|v_a\|, \|v_b\|, \|u\| \leq 1$*. Let* $\gamma \in (0,1)$*. Let* $x \in \mathrm{B}(0,1)$*. There exists a multivariate polynomial* $p : \mathbb{R}^d \to \mathbb{R}$ *such that*

1) $p(x) \geq \frac{1}{2}$ *for all* $x \in \bigcap_{i=1}^{m} \{\langle v_i, x\rangle \geq \gamma\} \cap \{\langle x, u'\rangle \geq \langle v_a, u'\rangle + \gamma\} \cap \{\langle x, u'\rangle \leq \langle v_b, u'\rangle - \gamma\}$,
2) $p(x) \leq -\frac{1}{2}$ *for all* $x \in \bigcup_{i=1}^{m} \{\langle v_i, x\rangle \leq -\gamma\} \cup \{\langle x, u'\rangle \leq \langle v_a, u'\rangle + \gamma\} \cup \{\langle x, u'\rangle \geq \langle v_b, u'\rangle - \gamma\}$,
3) $deg(p) = \lceil\log_2(2m+3)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil$,
4) $\|p\| \leq \frac{7}{2}\left[188\lceil\log_2(2m+3)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{\frac{\lceil\log_2(2m+3)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}$

Proof of Theorem 3. From Theorem 4 ,there exists a multivariate polynomial $p_i : \mathbb{R}^d \to \mathbb{R}$ such that for all $t \in [T]$ and the sequence $(x_1, y_1), (x_2, y_2), (x_t, y_t), \ldots, (x_T, y_T)$. If $y_t = i$; $p_i(x_t) \geq \frac{1}{2}$ and otherwise $y_t \neq i$; $p_i(x_t) \leq -\frac{1}{2}$. Theorem 2 implies that

$$\|p\| \leq \frac{7}{2}\left[188\lceil\log_2(2L+1)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{\frac{\lceil\log_2(2L+1)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}$$

By lemma 3, there exists $c_i \in \ell_2$ such that $\langle c_i, \phi(x)\rangle = p_i(x)$, and

$$\|c_i\|_{\ell_2} \leq 7\left[188\lceil\log_2(2L+1)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{\frac{\lceil\log_2(2L+1)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}.$$

We are ready to construct vectors for group weakly separable in ℓ_2 such that $\|z_1\|^2 + \|z_2\|^2 + \ldots + \|z_L\|^2 \leq 1$ and for all $t \in [T]$ $\langle z_{y_t}, x_t\rangle \geq \gamma$, and for all $j \neq y_t \langle z_j, x_t\rangle \leq -\gamma$,

$$z_i = \frac{c_i}{\sqrt{L} \cdot 7\left[188\lceil\log_2(2L+1)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{\frac{\lceil\log_2(2L+1)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}},$$

and

$$\gamma = \frac{\frac{7}{2}\left[188\lceil\log_2(2L+1)\rceil \cdot \left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil\right]^{-\frac{\lceil\log_2(2L+1)\rceil\cdot\left\lceil\sqrt{\frac{1}{\gamma}}\right\rceil}{2}}}{2\sqrt{L}},$$

□

B. Separating polynomials

As in [1] and [4], we use the Chebyshev polynomials [9] $T_n(\cdot)$ defined as follows.

$$\begin{aligned} T_0(z) &= 1, \\ T_1(z) &= z, \\ T_{n+1} &= 2zT_n(z) - T_{n-1}(z) \text{ for } n \geq 1 \end{aligned}$$

The following three lemmas are from [1].

Lemma 1 (from Lemma 15 in [1])**.** *(Properties of Chebyshev polynomials) Chebyshev polynomials satisfy*

1) $deg(T_n) = n$ *for all* $n \geq 0$*.*
2) *If* $n \geq 1$*, the leading coefficient of* $T_n(z)$ *is* $2n-1$*.*
3) $T_n(\cos(\theta)) = \cos(n\theta)$ *for all* $\theta \in \mathbb{R}$ *and all* $n \geq 0$*.*
4) $T_n(\cosh(\theta)) = \cosh(n\theta)$ *for all* $\theta \in \mathbb{R}$ *and all* $n \geq 0$*.*
5) $|T_n(z)| \leq 1$ *for all* $z \in [-1,1]$ *and all* $n \geq 0$*.*
6) $T_n(z) \geq 1 + n^2(z-1)$ *for all* $z \geq 1$ *and all* $n \geq 0$*.*
7) $\|T_n\| \leq (1+\sqrt{2})^n$ *for all* $n \geq 0$*.*

Lemma 2 (from Lemma 14 in [1])**.** *(Properties of norm of polynomials)*

1) *Let* $p_1, p_2, \ldots, p_n$ *be multivariate polynomials and let* $p(x) = \prod_{j=1}^{n} p_j(x)$ *be their product. Then,* $\|p\|^2 \leq n^{\sum_{j=1}^{n} deg(p_j)} \prod_{j=1}^{n} \|p_j\|^2$*.*
2) *Let* q *be a multivariate polynomial of degree at most* s *and let* $p(x) = (q(x))^n$*. Then,* $\|p\|^2 \leq n^{ns}\|q\|^{2n}$*.*
3) *Let* $p_1, p_2, \ldots, p_n$ *be multivariate polynomials. Then,* $\left\|\sum_{j=1}^{n} p_j\right\|^2 \leq n\sum_{j=1}^{n} \|p_j\|^2$*.*

Lemma 3 (from Lemma 9 in [1]). *(Norm bound) Let $p : \mathbb{R}^d \to \mathbb{R}$ be a multivariate polynomial. There exists $c \in \ell_2$ such that $p(x) = \langle c, \phi(x)\rangle_{\ell_2}$ and $\|c\|_{\ell_2} \leq 2^{deg(p)/2}\|p\|$.*

Our proof follows the general approach in [1].

Proof of Theorem 4. We need to construct a multivariate polynomial p_i such that for all $t \in [T]$ and for all $i \in [K]$,

- $y_t = i \Rightarrow p_i(x_t) \geq \frac{\gamma'}{2}$,
- $y_t \neq i \Rightarrow p_i(x_t) \geq -\frac{\gamma'}{2}$.

Let $v_1, v_2, \ldots, v_m \in V$ such that $\|v_1\|, \|v_2\|, \ldots, \|v_m\| \leq 1$. We say that x_t is in $G_{g(y_t)}$ if

$$\forall_{i\in m}\langle x_t, v_i\rangle \geq \gamma.$$

There exist $v_a, v_b \in V$ such that $\|v_a\|, \|v_b\| \leq 1$, for all $t \in [T]$ and $g(y_t) = j$

$$\langle x_t, u_j\rangle \geq \langle v_a, u_j\rangle + \gamma$$

and

$$\langle x_t, u_j\rangle \leq \langle v_b, u_j\rangle - \gamma.$$

We are ready to modify polynomial in theorem 7 of [1]. Let $r = \lceil \log_2(2m+3)\rceil$ and $s = \left\lceil \sqrt{\frac{1}{\gamma}} \right\rceil$. Define the polynomial $p : \mathbb{R}^d \to \mathbb{R}$ as

$$\begin{aligned} p(x) = m + \frac{3}{2} - \sum_{i=1}^{m}(T_s(1 - \langle v_i, x\rangle))^r \\ - (T_s(\langle x - v_a, u'\rangle/2))^r \\ - (T_s(\langle v_b - x, u'\rangle/2))^r. \end{aligned}$$

Notice that x in group G then $\langle v_i, x\rangle \geq \gamma$ for all $i \in [m]$. Since $\|x\| \leq 1$ and $\|v_i\| \leq 1$ we have $\langle v_i, x\rangle \in [0, 1]$ then $(T_s(1 - \langle v_i, x\rangle))^r \in [-1, 1]$, cause x lie in the group, $\langle x - v_a, u'\rangle \in [\frac{-1}{2}, \frac{1}{2}]$ and $\langle v_b - x, u'\rangle \in [\frac{-1}{2}, \frac{1}{2}]$ then $(T_s(\langle x - v_a, u'\rangle))^r \in [0, \frac{1}{2}]$ and $(T_s(\langle v_b - x, u'\rangle))^r \in [0, \frac{1}{2}]$. Therefore,

$$p(x) \geq m + \frac{3}{2} - m - 1 \geq \frac{1}{2}.$$

For x not in group, there exists at least one $i \in [m]$ such that $\langle v_i, x\rangle \leq -\gamma$. Therefore, $1 - \langle v_i, x\rangle \geq 1 + \gamma$ and 1 (part 6) imply that

$$T_s(1 - \langle v_i, x\rangle) \geq 1 + s^2\gamma \geq 2$$

ans thus

$$(T_s(1 - \langle v_i, x\rangle))^r \geq 2^r \geq 2m + 3,$$

for any $j \in [m]$, we have $\langle v_j, x\rangle \in [-1, 1]$ and thus $1 - \langle v_j, x\rangle \in [0, 2]$. According to Proposition 1 (part 5 and 6), $T_s(1 - \langle v_i, x\rangle) \geq -1$. And also $\|v_b - x\| \in [0, 2]$ and thus $\langle v_b - x, u'\rangle \in [-2, 1]$, and same for $\langle v_b - x, u'\rangle$, then $(T_s(\langle x - v_a, u'\rangle/2))^r \in [-1, 1]$ and $(T_s(\langle v_b - x, u'\rangle/2))^r \in [-1, 1]$. Therefore,

$$\begin{aligned} p(x) &= m + \frac{3}{2} - (T_s(1 - \langle v_i, x\rangle))^r - \sum_{j\in[m]j\neq i}(T_s(1 - \langle v_j, x\rangle))^r \\ &\quad - (T_s(\langle x - v_a, u'\rangle))^r - (T_s(\langle v_b - x, u'\rangle))^r \\ &\leq m + \frac{3}{2} - (2m + 3) + (m - 1) + 2. \end{aligned}$$

The degree of p is the same for all terms $(T_s(1-\langle v_i, x\rangle))^r$ that is $r \cdot s$. We prove the upper bound of norm of p. Let $f_i(x) = 1 - \langle v_i, x\rangle$, let $k_a(x) = \langle x - v_a, u'\rangle/2$, $k_b(x) = \langle v_b - x, u'\rangle/2$.

$$\|f_i\|^2 = 1 + \|v_i\|^2 \leq 1 + 1 = 2,$$

$$\|k_a\|^2 = \frac{\|x - v_a\|^2 \cdot \|u'\|^2}{2} \leq \frac{4 \cdot 1}{2} = 2$$

and

$$\|k_b\|^2 = \frac{\|v_b - x\|^2 \cdot \|u'\|^2}{2} \leq \frac{4 \cdot 1}{2} = 2.$$

Let $T_s(z) = \sum_{j=0}^{2} c_j z^j$ be the expansion of s-th Chebyshev polynomial. Then,

$$\begin{aligned} \|T_s(1 - \langle v_i, x\rangle)\|^2 &= \|T_s(f_i)\|^2 \\ &= \left\|\sum_{j=0}^{s} c_j (f_i)^j\right\|^2 \\ &\leq (s+1)\sum_{j=0}^{s}\left\|c_j (f_i)^j\right\|^2 \\ &= (s+1)\sum_{j=0}^{s} c_j^2\left\|(f_i)^j\right\|^2 \\ &\leq (s+1)\sum_{j=0}^{s} c_j^2 j^j \|f_i\|^{2j} \\ &\leq (s+1)\sum_{j=0}^{s} c_j^2 j^j 2^{2j} \\ &\leq (s+1)s^s 2^{2s}\sum_{j=0}^{s} c_j^2 \\ &= (s+1)s^s 2^{2s}\|T_s\|^2 \\ &= (s+1)s^s 2^{2s}(1+\sqrt{2})^{2s} \\ &= (s+1)\left(4(1+\sqrt{2})^2 s\right)^2 \\ &\leq (8(1+\sqrt{2})^2 s)^s \\ &\leq (47s)^s, \end{aligned}$$

$$\begin{aligned} \|T_s(\langle x - v_a, u'\rangle/2)\|^2 &= \|T_s(k_a)\|^2 \\ &= \left\|\sum_{j=0}^{s} c_j (k_a)^j\right\|^2 \\ &\leq (47s)^s \end{aligned}$$

and also

$$\begin{aligned} \|T_s(\langle v_b - x, u'\rangle/2)\|^2 &= \|T_s(k_b)\|^2 \\ &= \left\|\sum_{j=0}^{s} c_j (k_b)^j\right\|^2 \\ &\leq (47s)^s. \end{aligned}$$

Finally,

$$\begin{aligned}
\|p\| &\le m + \frac{3}{2} + \sum_{i=1}^{m} \|T_s(f_i)^r\| + \|T_s(k_a)^r\| + \|T_s(k_b)^r\| \\
&= m + \frac{3}{2} + \sum_{i=1}^{m} \sqrt{\|T_s(f_i)^r\|^2} + \sqrt{\|T_s(k_a)^r\|^2} + \sqrt{\|T_s(k_b)} \\
&\le m + \frac{3}{2} + \sum_{i=1}^{m} \sqrt{r^{rs} \|T_s(f_i)^r\|^{2r}} \\
&+ \sqrt{r^{rs} \|T_s(k_a)^r\|^{2r}} + \sqrt{r^{rs} \|T_s(k_b)^r\|^{2r}} \\
&\le m + \frac{3}{2} + mr^{rs/2}(47s)^{rs/2} + r^{rs/2}(47s)^{rs/2} + r^{rs/2}(47s) \\
&= m + \frac{3}{2} + (m+2)(47rs)^{rs/2}.
\end{aligned}$$

using $m \le \frac{1}{2}2^r$, since $r, s \ge 1$,

$$\begin{aligned}
\|p\| &\le m + \frac{3}{2} + (m+2)(47rs)^{rs/2} \\
&\le \frac{1}{2}2^r + \frac{3}{2} + \left(\frac{1}{2}2^r + 2\right)(47rs)^{rs/2} \\
&\le 2 \cdot 2^r + \frac{3}{2} \cdot 2^r (47rs)^{rs/2} \\
&= 2^r \left(2 + \frac{3}{2}\right)(47rs)^{rs/2} \\
&\le 4^{rs/2} \cdot \frac{7}{2}(47rs)^{rs/2} \\
&= \frac{7}{2}(188rs)^{rs/2}.
\end{aligned}$$

A simple substituting finishes the proof. □

IV. Experiments

While we focus mostly on the theoretical aspect of the problem, we perform some experiment to visualize the algorithm.

We generate a dataset in $\mathbb{R}^2$ under the weakly linear separable condition, with $K = 9$ classes and $L = 3$ groups with margin γ, shown in Fig 2.

We run Algorithm 1 for $T = 13866$ steps. The algorithm makes 4586 mistakes. To see the decision boundary, we consider all points in $\mathbb{R}^2$ and plot the algorithm's predictions in Fig 3. The boundaries of the decisions can be seen in it.

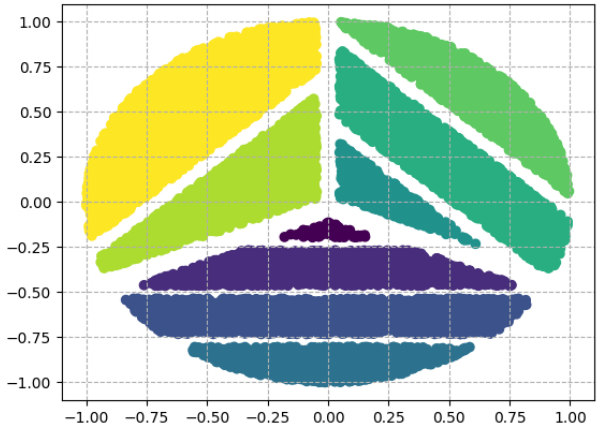


Fig. 2. Weakly separable with group leveled linear separable

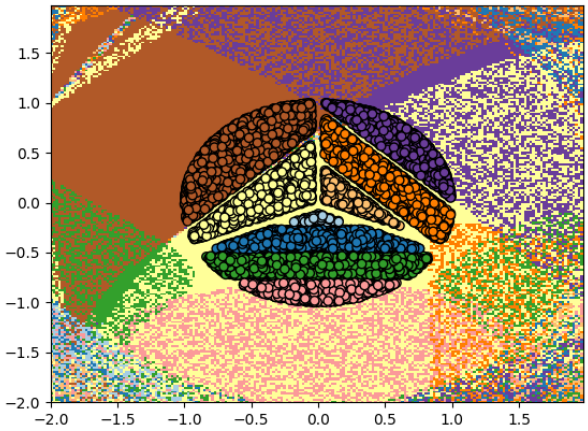


Fig. 3. Algorithm with rational kernel final decision boundaries

References

[1] A. Beygelzimer, D. Pal, B. Szorenyi, D. Thiruvenkatachari, C.-Y. Wei, and C. Zhang, "Bandit multiclass linear classification: Efficient algorithms for the separable case," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. Long Beach, California, USA: PMLR, 09–15 Jun 2019, pp. 624–633. [Online]. Available: http://proceedings.mlr.press/v97/beygelzimer19a.html

[2] K. Crammer and Y. Singer, "Ultraconservative online algorithms for multiclass problems," *J. Mach. Learn. Res.*, vol. 3, pp. 951–991, Mar. 2003.

[3] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, "Efficient bandit algorithms for online multiclass prediction," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 440–447.

[4] A. R. Klivans and R. A. Servedio, "Learning intersections of halfspaces with a margin," *Journal of Computer and System Sciences*, vol. 74, no. 1, pp. 35 – 48, 2008, learning Theory 2004.

[5] R. Beigel, N. Reingold, and D. Spielman, "Pp is closed under intersection," *Journal of Computer and System Sciences*, vol. 50, no. 2, pp. 191 – 202, 1995.

[6] G. Chen, G. Chen, J. Zhang, S. Chen, and C. Zhang, "Beyond banditron: A conservative and efficient reduction for online multiclass prediction with bandit setting model," in *ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009*, 2009, pp. 71–80.

[7] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[8] S. Shalev-Shwartz, O. Shamir, and K. Sridharan, "Learning kernel-based halfspaces with the 0-1 loss," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1623–1646, 2011. [Online]. Available: https://doi.org/10.1137/100806126

[9] J. Mason and D. Handscomb, *Chebyshev Polynomials*. CRC Press, 2002. [Online]. Available: https://books.google.co.th/books?id=8FHf0P3to0UC

Parsing Thai Social Data: A New Challenge for Thai NLP

Sattaya Singkul
King Mongkut's Institute of Technology Ladkrabang
Bankok, Thailand
joeysattaya@gmail.com

Borirat Khampingyot
Chiang Mai University
Chiang Mai, Thailand
borirat.khampingyot@gmail.com

Nattasit Maharattamalai

Supawat Taerungruang
Kasikorn Labs
Nonthaburi, Thailand
{nattasit.m}{supawat.tae}{tawunrat.c}@kbtg.tech

Tawunrat Chalothorn

***Abstract*—Dependency parsing (DP) is a task that analyzes text for syntactic structure and relationship between words. DP is widely used to improve natural language processing (NLP) applications in many languages such as English. Previous works on DP are generally applicable to formally written languages. However, they do not apply to informal languages such as the ones used in social networks. Therefore, DP has to be researched and explored with such social network data. In this paper, we explore and identify a DP model that is suitable for Thai social network data. After that, we will identify the appropriate linguistic unit as an input. The result showed that, the transition based model called, improve Elkared dependency parser outperform the others at UAS of 81.42%.**

Keywords—natural language processing, dependency parsing, social data

I. INTRODUCTION

Dependency parsing (DP) is a task that analyzes text for syntactic structure and relationship between words. DP could be used for improving NLP tasks such as information extraction [1], question answering [2, 3], and semantic parsing [4]. Social media are platforms that people use for communication, especially in the context of customer service support (i.e., customers reporting problems or feedback to a company's social network page). In fact, customer service support via social networks is increasingly popular among business companies. Consequently, the amount of textual data a company receives from a social network channel has also increased substantially. This also poses a challenge for the company's customer service department to analyze such massive amount of data in order to identify problems and improve their service quality. To do so, each piece of text must be extracted for customer intention as well as products or services mentioned. However, social media texts are more difficult to process than traditional texts [5] and, sometimes, they can be more difficult to understand. Moreover, there is also a challenge of syntax ambiguity because it is harder to identify sentence boundaries and grammars in Thai social language.

As shown in Table I, the first and second sentence clearly indicate that a customer wants a Chopper card. The third sentence, however, consists of two intentions from the customer: 1) he/she wants to apply for a Chopper card, and 2) he/she is looking for a Chopper card that is cuter than Rilakkuma card. Finally, the fourth sentence has the most complex structure, indicating that 1) the customer wants a Chopper card, 2) the Chopper card must come with an installment plan, 3) the Chopper card must be cuter than Rilakkuma card, and 4) he/she recalls that a Rilakkuma card can withdraw cash from cash machines. The fourth sentence consists of two intentions, three services, and two brands. Such a complex sentence requires non-trivial effort from human operators to analyze. Moreover, human operators must also aggregate all the analyzed results and generate a report regularly within limited time. Such manual process is cumbersome and yet the results might be inaccurate. Therefore, there is a need for an automated system that can interpret intentions and sentiments from such complex sentences at scale.

TABLE I. SENTENCES IN SOCIAL DATA

Sentence Type	Sentence
Normal	ขอบัตรซ็อปเปอร์ได้มะ 'Can I have a Chopper card?'
Normal	ขอสมัครบัตรซ็อปเปอร์ได้มะ 'Can I apply for a Chopper card?'
Long	ขอสมัครบัตรซ็อปเปอร์ของกสิกรที่น่ารักกว่าริระคุดั้ยมะอ่ะ 'Can I apply for Kasikorn's Chopper card that is cuter than Rilakkuma?'
Complex	ขอสมัครบัตรกสิกรอันที่ผ่อนได้มีลายซ็อปเปอร์ไหมอ่ะที่น่ารักกว่าที่กดเงินสดที่เป็นลายริระคุมะได้ป่าว 'I would like to apply for Kasikorn's card that can be used to pay by installments, is there a Chopper pattern, which is lovelier than Rilakkuma that can withdraw cash?'

In order to perform automated intention classification and sentiment analysis of complex sentences (like the ones shown in Table I), Dependency Parsing (DP) must be achieved. According to [6-8], if we do not understand relationships between words, relationships between entities in a sentence cannot be extracted. In particular, if the text consists of multiple entities (as shown in Table I: complex sentence), relationships between entities can help identify which entity should be focused (e.g., entity of "Kasikorn's card" should be focused in the complex sentence example from Table I). Therefore, the lack of Thai language DP could lead to misunderstanding in the meaning of the sentence. In fact, the lack of Thai language DP is one of the reasons why high-level Thai NLP tasks (e.g., sentiment analysis, question answering) cannot be implemented. Previous research works on DP are based on English text corpus [9-11] and hence cannot be used with Thai social network text.

Nonetheless, to solve such problem, two challenges are explored and addressed in this paper. The first challenge is to identify a suitable model for parsing Thai social data. The second challenge is to identify an appropriate linguistic unit as an input for DP. The paper addresses the first challenge by analyzing the characteristics of Thai social language. To address the second challenge, the paper proposes to use Elementary Discourse Units (EDUs) as input to conform to those linguistic characteristics. Ultimately, the experiment demonstrates interesting performance resulting from the selection of suitable models and input units.

The remainder of this paper is structured as follows; the theoretical background of the related works is reviewed in Section II. Section III describes the characteristics of Thai social data. The experiment and its results are discussed in Section IV and V, respectively. Finally, Section VI concludes the paper.

II. Related Works

Dependency parser (DP) is a task of natural language processing (NLP) that is widely used for extracting and analyzing grammatical structure of a sentence [12, 13]. Dependency links are close to the semantic relationships needed for text interpretation [14] (e.g., dependency relation can clearly show the relationship between words.) In addition, there are two approaches normally used in the tasks of dependency parsing: transition-based and graph-based.

Transition-based DP is a process of parsing a sequence of actions (transitions) for building a dependency graph and constructing a dependency tree by scanning left-to-right (or right-to-left) through words along the sentence. There are many research works that explore this method. For example, Zhang and Nirve [15] proposed new features that achieved the Unlabeled Attachment Score (UAS) of 92.9% on Penn Treebank and 86.0% on Chinese Treebank. Those features are composed of distance, valency, unigrams, third-order and label set. Moreover, stacked LSTM is proposed for transition-based DP by Dyer et al. [16]. Their model achieved better performance in both Stanford Dependency Treebank and Penn Chinese Treebank 5.1 with the UAS of 93.1% and 87.2%, respectively. Stenetorp [17] suggested to use recursive neural networks in transition-based parsing and achieved UAS of 86.25% on CoNLL 2008 dataset.

On the other hand, graph-based dependency parsing uses a concept of node to represent each word in a sentence. A search process then starts by constructing a dependency graph to adjust the weight of each edge in the connected graph such that 1) all nodes are covered, and 2) the sum of highest scoring edges is maximized. Flanigan et al. [18] used the inspiration of graph-based parsing techniques for abstract meaning representation (AMR). Their concept achieved an F-score of 84% on the testing data of LDC2013E117 corpus [19]. Moreover, Wang and Chang [20] proposed to used Bidirectional LSTM for graph-based parsing with English Penn-YM Treebank [21], English Penn-SD Treebank [22] and Chinese Penn Treebank (CTB5) [23]. They claimed that their results achieved better performance on Penn-SD dataset (UAS of 94.08%) where the data size is four times larger than Penn-YM (UAS of 93.51%) and CTB5 (UAS of 87.55%) datasets.

Furthermore, universal dependency (UD) [24] is a framework that aims to create treebank across different human languages. Also, UD is an open community producing more than 100 treebanks in over 70 languages. UD dataset is typically used in the research works such as [25], [26], and [27]. Parallel Universal Dependencies (PUD) treebanks, which were created for the CoNLL 2017 shared task on Multilingual Parsing from Raw Text to Universal Dependencies, are multilingual treebanks taken from news domains and Wikipedia. Moreover, there is a Thai PUD that consists of 1,000 lines of sentences or 22,322 tokens of word. Because of the lacking of labeled dataset, the Thai PUD is used as one of the corpus in this work.

III. Characteristics of Thai Social Data

It is generally known that communication channel is one of the factors that affects language usage patterns. On social media, texts have characteristics that reflect the social conversations. For this reason, the language used on social media is diverse and constantly changes according to people, topics, and situations. In this section, we discuss the key language characteristics of Thai social data that drive us to build parser for social domain.

A. Word

Word is a linguistic unit that represents concepts [28]. In general, the concepts represent through words are meaning or grammatical functions. However, words in social domain have behaviors that are different from those in formal domain because of the rapid variation of online communication.

In terms of word form, the same word may appear in a variety of forms. A variation of word form is usually made by sound variant [29]. For example, “จัง” /caŋ1/ is changed to “จุง” /cuŋ1/, “จรุง” /cruŋ1/, or “ชรุง” /chruŋ1/.

In terms of the meaning, a number of words that appear in social domain have different meanings to the same word form that appears in formal domain. For example, in formal domain, “กาก” /kaak2/ denotes ‘the rest after the good part is removed’. But in social domain, it means ‘bad’. In addition, the meanings of the words that appear in the social domain are also varied according to the number of new words being added according to the behavior of language users. These words, for example “ตะมุตะมิ” /ta1.mu4.ta1.mi4/, “สายเปย์” /saaj5.pe1/, “ป๊ะ” /pua?4/, are all not found in the dictionary.

In terms of function, grammatical functions of some words are extended beyond those appeared in formal domain. For instance, a word “แบบ” /bɛɛp2/ ‘form, model’ normally functions as a subject “*แบบ*อยู่ในลิ้นชัก” ‘A form is in the drawer’, an object “พนักงานยื่น*แบบ*ทางอินเทอร์เน็ต” ‘Employee submits forms on the internet’, or a classifier “เจ้าหน้าที่เสนอทางเลือก 2 *แบบ*” ‘Officer offers 2 options’. Conversely, in social domain, “แบบ” has more grammatical functions, e.g. an adverb marker “เราก็ขึ้นรถ*แบบ*งงๆ” ‘I got in a car confusedly’, a subordinate conjunction for adverbial clause “นางก็เดินไป*แบบ*ไม่หันกลับเลยจ้า” ‘She walked without turning back’, a relativizer “เขาเป็นคน*แบบ*ไม่สนโลก” ‘He is a person who doesn’t care about anything’, a discourse marker “*แบบ*จะไปทำงานสายแล้วไง” ‘being go to work late’.

With the aforementioned characteristics of the words, a language processing tool built on formal domain data may return unsatisfactory results. Because there are words that do

not appear in formal domain. Moreover, the new words and the extended grammatical functions of the words that cannot be found in formal domain will directly affect the part of speech tagging task. If the function of a word changes, the POS of word changes accordingly. Especially, in a syntactic task like this work, POS plays a very important role in expressing the relationship between words in the text. For such reasons, a parser model in this work uses Thai social data for training and testing.

B. *Sentence*

Sentence structures in the social language exhibit various complexity levels. For example, each sentence may consist of a small amount of words or may contain complex clauses that modify each other. However, the complexity of the social language is different from that of the formal language. This represents a challenge for social language processing. However, the language used in online media is similar to the spoken language. In addition to the characteristics of the words mentioned in the previous section, the characteristics of sentences in the social domain are also influenced by the spoken language. For this reason, the sentence structure is not strict. Consequently, many sentences cannot be communicated clearly.

For example, a sentence "แม่ชอบไปพารากอนมีหลายชั้น" 'Mom likes to go to Paragon has many floors'. This sentence informs two ideas: "Mom likes to go to Paragon" and "Paragon has many floors". As usual, in formal style, this sample sentence should be written separately into 2 sentences. It is not known how this phenomenon occurs, but this could be assumed that ellipses are the mechanism behind them. The ellipsis is a linguistic mechanism that is often used in spoken language [30], which includes languages in social domain.

Considering the example sentence above, there is possible that a relativize "ซึ่ง" was removed form "แม่ชอบไปพารากอน (ซึ่ง) มีหลายชั้น" 'Mom likes to go to Paragon (which) has many floors'. Based on this assumption, the other types of grammatical units that can be removed in social languages are found. For example, a verb "ผม(คิด)ว่าเขาไม่ไปหรอก" 'I (think) that he doesn't go', or a complementizer "แบงก์ชาติคาด(ว่า)อัตราเงินเฟ้อของไทยมีแนวโน้มต่ำลง" 'Bank of Thailand expects (that) Thailand's inflation rate will be lower'

The complexity of the sentence is another characteristic that needs to be discussed. Since sentences in social media are not formal and are similar to spoken language. The length of sentences is vary. Most of sentences are very complex because the speaker typed the sentence immediately without proper screening for clear communication. Therefore, they may be complex and difficult to understand. For example, a sentence "ช่วยโพสรูปที่ถ่ายเมื่อวานตอนเย็นที่เราไปกินข้าวกันที่สยามสแควร์ที่เรานั่งข้างๆ เธอให้ทีได้มะ" 'Can you post a photo taken yesterday evening that we went to have dinner together at Siam Square that I sat next to you?'. This sentence consists of at least 4 main information: "Can you post a photo?", "A photo taken yesterday evening", "we went to have dinner together at Siam Square on yesterday evening", and "a photo that I sat next to you".

However, because of the complex modification of this sentence, some people may receive more information, i.e. "yesterday evening that I sat next to you". Such ambiguity is common in social language, and it has a significant effect on processing. Actually, a clause "ที่เรานั่งข้างๆ เธอ" 'that I sat next to you' can modify many nouns in the sentence, including "รูป" 'picture', "เมื่อวาน" 'yesterday', "ตอนเย็น" 'evening', "เมื่อวานตอนเย็น" 'yesterday evening', and "สยามสแควร์" 'Siam Square'. Therefore, the information received will depend on the noun that the hearer selects to modify.

With problems of Thai sentence mentioned above, together with the characteristics of the Thai language in which the sentence has no clear boundary [31], EDUs [32] is chosen to be a processing unit in this work. Due to, in semantic perspective, EDUs can convey single piece of information clearly. On the other hand, in syntactic perspective, EDUs are in the form of clauses or phrases with a strong marker [33] and hence can be clearly identified the boundaries. Furthermore, because of a characteristic of clauseness, structure of EDUs is also less complicated than sentences.

IV. Experiment

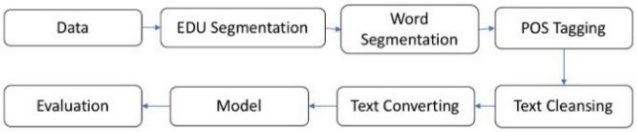


Fig. 1. Experiment process

A. *Data*

There are 2 Thai datasets used in the experiment: public UD data and social data in financial domain. Both datasets contain 1,000 sentences are grouped into 10 folds for cross-validation. Each fold consists of 800 sentences, 100 sentences and 100 sentences, respectively.

1) *Thai Social data*

This dataset is collected from social media, such as Facebook, Twitter, and Pantip by focusing on financial domain. The data is analysed and segmented into EDU by applying the principles proposed by Intasaw and Aroonmanakun [33]. The size of dataset is 219,585 EDUs.

In term of the length of EDUs, as shown in Fig. 2, the distribution of word per EDU varies. The length of word per EDU is between 2-24 words and has uniform distribution.

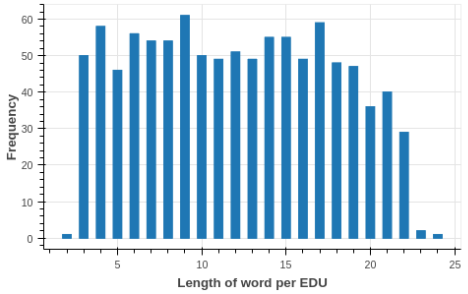


Fig. 2. distribution of word length

Fig. 3 shows the distribution of POS and the number of POS tag sets used in the data. The tag set is adapted from a universal POS tag set [34].

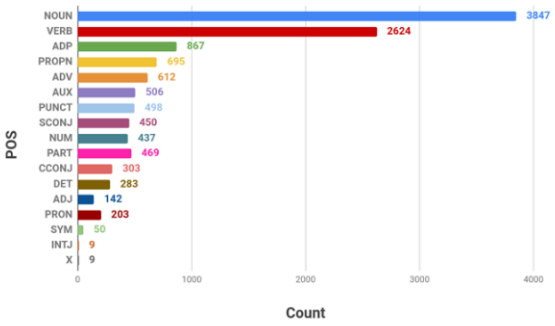


Fig. 3. Distribution of POS

2) UD Thai Tree Bank

This dataset consists of 22,322 words, which is a standard Thai language dataset normally used in supervised learning. The label of each sample is the relationship between words.

B. Evaluation

There are many different evaluation metrics used in the dependency parsing task. The commonly used metrics are unlabeled attachment score (UAS) and labeled attachment score (LAS). However, due to the lack of dependency labels, UAS is used to evaluate the quality of dependency parsers. UAS focuses on the percentage of words that get the correct predictions. Equation (1) defines UAS as the number of correctly predicted heads divided by all heads in ground truth.

$$UAS = \frac{\#\ of\ correct\ heads}{\#\ of\ heads} \quad (1)$$

C. Preprocessing

Before the training step, the dataset is passed through data preprocessing step and turned into an appropriate format. There are 5 processes : EDU segmentation, word segmentation, part of speech tagging, text cleansing, and text converting. The sentences are segmented in EDU segmentation process and then word segmentation process. After that, each word is marked into a category of words such as subject, verb, noun and objective. Next, special characters and excessive space are removed. Then, the text and number that have been separated by the error of word segmentation are combined in the text cleansing process. Furthermore, text converting process will convert the data to a universal format (CONLL-U format).

D. Model

Transition-based and graph-based methods explored and used in the experiment are discussed in this section.

1) Transition-based

Transition-based models identify relationship between words by considering the transition of words through oracle parsing in order to see the change in each transition as shift reduction. They then use mapping features for extracting feature and converting the data into a suitable format for model training.

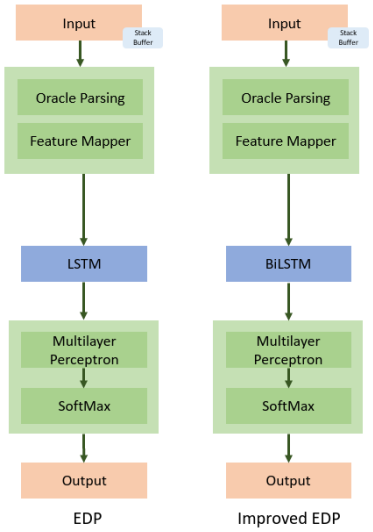


Fig. 4. A structure of transition based models consist of Elkaref Dependency Parser (EDP) and Improved Elkaref Dependency Parser.

a) Elkaref Dependency Parser (EDP)

Elkaref Dependency Parser [35] is composed of a single LSTM hidden layer replacing the hidden layer in the usual feed-forward network architecture. It also proposes a new initialization method that uses the pre-trained weights from a feed-forward neural network to initialize the LSTM-based model.

b) Improved Elkaref Dependency Parser

The concept of EDP, which has only one direction of word sequence relation, may not be enough. The concept of Kiperwasser Dependency Parsing [36] is developed using Bi-LSTM instead of LSTM to extract bi-directional features word sequence relation. Each sentence token is associated with a Bi-LSTM vector representing the token in its sentential context. Feature vectors are then constructed by concatenating a few Bi-LSTM vectors.

2) Graph-based

Graph-based models identify relationship between words by considering the characteristics of the graph. They focus on each pair of words and check if the pair correlate through the matrix scoring process using LSTM Encoder and Decoder.

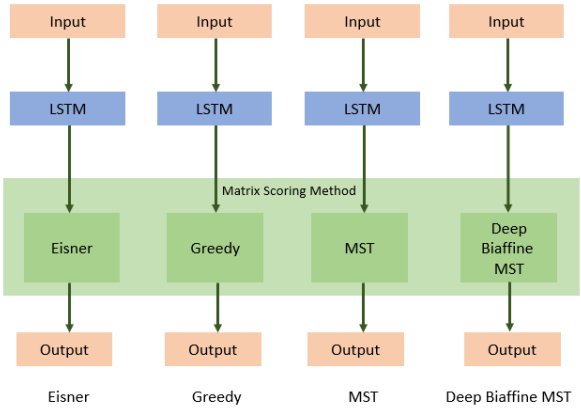


Fig. 5. A structure of graph-based models consists of Eisner, Greedy, Maximum Spanning Tree (MST) and Deep Biaffine MST.

a) Eisner

Eisner is a bottom-up dependency parsing algorithm. It is a projective dependency parsing and focuses on subgraph process. Adding one link at a time making it easy to multiply the model's probability factor similar to CKY method.

b) Greedy

Greedy is an algorithm which always selects the highest weighted edges. It is non-projective dependency parsing and compares on next-to-edge by memory-based parser [37].

c) Maximum Spanning Tree (MST)

Maximum Spanning Tree finds a dependency tree with higher score on a directed graph. Scores are independent from other dependencies. It is a non-projective dependency parsing and applied Chu-Liu-Edmonds algorithm [37] to find MST from directed graphs. There are composed of 3 steps: greedy, contract and recursive. Greedy step finds edges with the highest weight. Contract step detects cycles and breaks them by removing the edge with the smallest value in the cycle. Recursive step repeats the process until a spanning tree is obtained.

d) Deep Biaffine MST

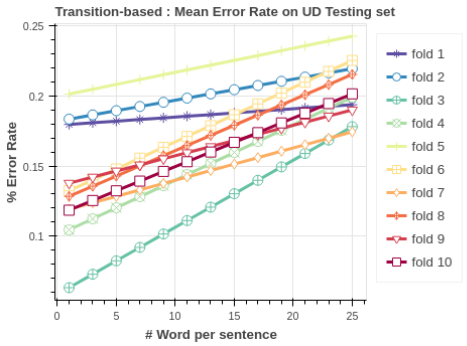
Deep Biaffine MST [38] is a deep learning model. By adding the Bi-LSTM and a Biaffine classifier, the model performs comparably to the state-of-the-art model. The model utilizes Bi-LSTM, which gives a long-term dependency, and Biaffine classifier, which improves parsing speed.

V. Results and Discussions

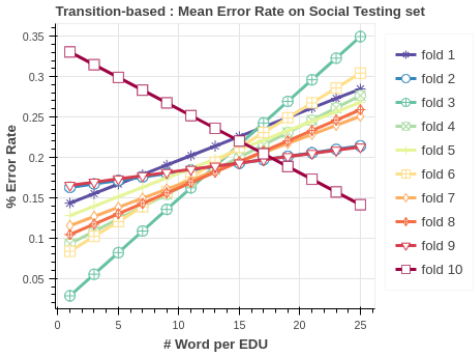
There are two experiments in this paper. The first experiment focuses on the correlation between word length and error rate. The second experiment focuses on Thai social language model.

1) The correlation between word lengths and error rates

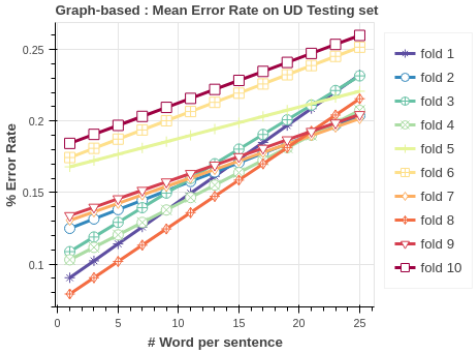
This experiment was conducted on both UD dataset and Social Banking domain dataset. Two different types of methods, transition-based and graph based, are used to analyze how word lengths affect the model performance. The Improved Elkaref Dependency Parser [35] is used for training a transition-based model and the Deep Biaffine Attention [38] is used for training a graph-based model. The mean error rate is evaluated by counting the frequency of the wrong predictions in each sentences / EDUs and calculating mean error rate in each word length. Fig. 6 (a) and (c), representing training and testing on UD dataset with transition-based and graph-based, show that the more number of words in the sentences or EDUs are contained, the more error rates are found for both transition-based and graph-based model. In addition, Fig. 6 (b) and (d), which is training and testing on Social Banking domain dataset, show that nine out of ten folds yield the same direction of correlation between word lengths and error rates. Due to Social Banking domain dataset separated into EDUs with short words, this might cause a different correlation result in another fold. To simplify the problem, EDU segmentation is suggested to be used in Thai dependency parsing instead of sentence segmentation.



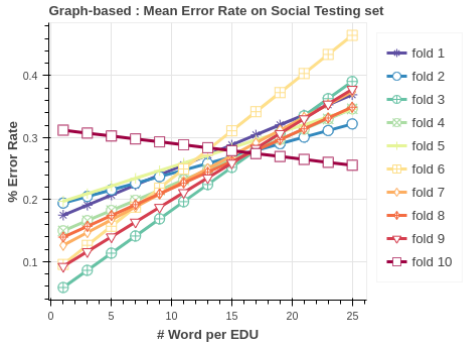
(a)



(b)



(c)



(d)

Fig. 6. Each line represents a direction of error rate occurred while the number of word per EDU increase using linear regression in (a) UD Testing set with transition-based, (b) Social Testing set with transition-based, (c) UD Testing set with graph-based and (d) Social Testing set with graph-based. X axis represents the number of words. Y axis represents percent error rates. Transition-based and graph-based use the same data distribution of UD and social dataset in training, validation and testing.

2) Thai Social Model

This experiment was conducted to find the best model for Thai social dependency parsing. As shown in Table 2, there are six models, including two transition-based models and four graph-based models, tested in this experiment. The results show that transition-based models perform better than graph-based models. The improved Elkaref dependency parser achieves the average 10-fold UAS of 78.62% on UD dataset and the average 10-fold UAS of 79.84% on social dataset. Moreover, the transition-based UAS (79.84%) outperformed the graph-based UAS (73.27%) on social dataset. Because of the concept of EDUs, the word length in sentence is always longer than or equal to the word length in EDU. This finding is consistent with the research work in [39], which found that "transition-based models performed better than graph-based models at short length sentences". The best model for Thai social model is improved Elkaref dependency parser.

TABLE II. RESULT ON THE UD DATASET AND SOCIAL DATASET

Type	**Model**	**UD Dataset**	**Social Dataset**
		UAS	***UAS***
Transition	EDP	55.01	73.92
	Improved EDP	**78.62**	**79.84**
Graph	Eisner	56.93	58.37
	Greedy	55.24	53.80
	MST	57.12	60.99
	Deep Biaffine MST	76.95	73.27

VI. CONCLUSION

In this paper, we have shown that length is one of the error factors in the dependency parsing problem. We suggested the use of EDU segmentation to simplify sentences instead of using sentence segmentation or long raw text. Our experimental results also show that transition-based DP models outperform the graph-based DP models in Thai social data when segmented by EDUs. Moreover, improved Elkaref dependency parser yielded the best performance among various DP models. For future works, exploration of error factors is a promising area to explore in order to improve the model performance.

VII. ACKNOWLEDGEMENT

This work was supported by Kasikorn Business-Technology Group (KBTG).

REFERENCES

[1] Mausam, M. Schmitz, S. Soderland, R. Bart, and O. Etzioni, "Open Language Learning for Information Extraction," presented at the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL '12), Jeju Island, Korea, jul, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1048.

[2] G. Attardi, A. Cisternino, F. Formica, M. Simi, and R. Tommasi, "PiQASso: Pisa Question Answering System," presented at the Tenth Text REtrieval Conference (TREC 2001), Gaithersburg, Maryland, USA, 2001.

[3] H. Li and F. Xu, "Question Answering with DBpedia Based on the Dependency Parser and Entity-centric Index," in *2016 International Conference on Computational Intelligence and Applications (ICCIA)*, Jeju, 2016, pp. 41-45, doi: 10.1109/ICCIA.2016.10.

[4] S. Reddy *et al.*, "Transforming Dependency Structures to Logical Forms for Semantic Parsing," *Transactions of the Association for Computational Linguistics,* vol. 4, pp. 127-140, 2016, doi: 10.1162/tacl_a_00088.

[5] F. Benamara, D. Inkpen, and M. Taboada, "Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information," *Computational Linguistics,* vol. 44, no. 4, pp. 663-681, 2018, doi: 10.1162/coli_a_00333.

[6] A. Akbik and J. Broß, "Wanderlust: Extracting Semantic Relations from Natural Language Text Using Dependency Grammar Patterns," presented at the 2009 Semantic Search Workshop at the 18th International World Wide Web Conference, Madrid, Spain, 2009.

[7] T. Wang, Y. Li, K. Bontcheva, H. Cunningham, and J. Wang, "Automatic Extraction of Hierarchical Relations from Text," Berlin, Heidelberg, 2006: Springer Berlin Heidelberg, in The Semantic Web: Research and Applications, pp. 215-229.

[8] K. Fundel, R. Küffner, and R. Zimmer, "RelEx—Relation extraction using dependency parse trees," *Bioinformatics,* vol. 23, no. 3, pp. 365-371, 2006, doi: 10.1093/bioinformatics/btl616.

[9] A. Ivanova, S. Oepen, and L. Øvrelid, "Survey on parsing three dependency representations for English," presented at the ACL Student Research Workshop, Sofia, Bulgaria, 2013.

[10] J. Nivre and M. Scholz, "Deterministic Dependency Parsing of English Text," Geneva, Switzerland, aug 23–aug 27 2004: COLING, in COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, pp. 64-70.

[11] L. Kong, N. Schneider, S. Swayamdipta, A. Bhatia, C. Dyer, and N. A. Smith, "A Dependency Parser for Tweets," Doha, Qatar, oct 2014: Association for Computational Linguistics, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1001-1012, doi: 10.3115/v1/D14-1108.

[12] D. Jurafsky and J. H. Martin, *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd ed. (Prentice Hall series in artificial intelligence). Upper Saddle River, N.J.: Pearson Prentice Hall, 2009.

[13] F. T. Martins, N. A. Smith, and E. P. Xing, "Concise integer linear programming formulations for dependency parsing," presented at the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1, Suntec, Singapore, 2009.

[14] M. A. Covington, "A fundamental algorithm for dependency parsing," presented at the 39th Annual ACM Southeast Conference, Athens, Georgia, 2001.

[15] Y. Zhang and J. Nivre, "Transition-based Dependency Parsing with Rich Non-local Features," presented at the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, 2011.

[16] C. Dyer, M. Ballesteros, W. Ling, A. Matthews, and N. A. Smith, "Transition-Based Dependency Parsing with Stack Long Short-Term Memory," presented at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015.

[17] P. Stenetorp, "Transition-based dependency parsing using recursive neural networks," presented at the NIPS Workshop on Deep Learning, Lake Tahoe, USA, 2013.

[18] J. Flanigan, S. Thomson, J. Carbonell, C. Dyer, and N. A. Smith, "A Discriminative Graph-Based Parser for the Abstract Meaning Representation," Baltimore, Maryland, jun 2014: Association for Computational Linguistics, in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1426-1436, doi: 10.3115/v1/P14-1134.

[19] L. Banarescu *et al.*, "Abstract Meaning Representation for Sembanking," presented at the 7th Linguistic Annotation Workshop and Interoperability with Discourse, Sofia, Bulgaria, 2013.

[20] W. Wang and B. Chang, "Graph-based Dependency Parsing with Bidirectional LSTM," Berlin, Germany, aug, 2016.

[21] H. Yamada and Y. Matsumoto, "Statistical dependency analysis with support vector machines," in *International Conference on Parsing Technologies (IWPT)*, 2003, pp. 195-206.

[22] M.-C. de Marneffe, B. MacCartney, and C. D. Manning, "Generating Typed Dependency Parses from Phrase Structure Parses," presented at the 5th International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, may, 2006. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2006/pdf/440_pdf.pdf.

[23] Y. Zhang and S. Clark, "A Tale of Two Parsers: Investigating and Combining Graph-based and Transition-based Dependency Parsing," Honolulu, Hawaii, oct, 2008. [Online]. Available: https://www.aclweb.org/anthology/D08-1059.

[24] J. Nivre *et al.*, "Universal Dependencies v1: A Multilingual Treebank Collection," Portorož, Slovenia, may 2016: European Language Resources Association (ELRA), in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), pp. 1659-1666.

[25] J. Bjerva, B. Plank, and J. Bos, "Semantic Tagging with Deep Residual Networks," presented at the 26th International Conference on Computational Linguistics, Osaka, Japan, 2016.

[26] M. Zampieri *et al.*, "Findings of the VarDial Evaluation Campaign 2017," Valencia, Spain, apr 2017: Association for Computational Linguistics, in Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial), pp. 1-15, doi: 10.18653/v1/W17-1201.

[27] H. M. Alonso and B. Plank, "When is multitask learning effective? Semantic sequence prediction under varying data conditions," presented at the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 2017.

[28] T. Givón, *Syntax : An introduction*. Amsterdam: John Benjamins, 2001.

[29] W. Aroonmanakun, N. Nupairoj, V. Muangsing, and S. Choemprayong, "Thai Monitor Corpus: Challenges and Contribution to Thai NLP," *Vacana,* vol. 6, no. 2, pp. 1-14, 2018.

[30] S. Nariyama, "Pragmatic information extraction from subject ellipsis in informal English," presented at the 3rd Workshop on Scalable Natural Language Understanding, New York City, New York, 2006.

[31] A. Lertpiya *et al.*, "A Preliminary Study on Fundamental Thai NLP Tasks for User-generated Web Content," presented at the 13th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2018), Pattaya, Thailand, 2018.

[32] L. Carlson, D. Marcu, and M. E. Okurowski, "Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory," presented at the Second SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark, September 1-2, 2001.

[33] N. Intasaw and W. Aroonmanakun, "Basic principles for segmenting Thai EDUs," presented at the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27), Taipei, Taiwan, November 22-24, 2013.

[34] S. Petrov, D. Das, and R. McDonald, "A Universal Part-of-Speech Tagset," presented at the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, 2012.

[35] M. Elkaref and B. Bohnet, "A Simple LSTM model for Transition-based Dependency Parsing," *arXiv e-prints*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2017arXiv170808959E

[36] E. Kiperwasser and Y. Goldberg, "Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations," *arXiv e-prints*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160304351K

[37] R. McDonald, F. Pereira, K. Ribarov, and J. Hajič, "Non-Projective Dependency Parsing using Spanning Tree Algorithms," Vancouver, British Columbia, Canada, oct 2005: Association for Computational Linguistics, in Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 523-530. [Online]. Available: https://www.aclweb.org/anthology/H05-1066.

[38] T. Dozat and C. D. Manning, "Deep Biaffine Attention for Neural Dependency Parsing," *arXiv e-prints*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv161101734D

[39] R. McDonald and J. Nivre, "Analyzing and integrating dependency parsers," *Comput. Linguist.,* vol. 37, no. 1, pp. 197-230, 2011, doi: 10.1162/coli_a_00039.

Using Label Noise Filtering and Ensemble Method for Sentiment Analysis on Thai Social Data

Chayanont Eamwiwat*, Pongpisit Thanasutives*, Chanatip Saetia, Tawunrat Chalothorn
Kasikorn Labs
Nonthaburi, Thailand
{chayanont.eam, pongpisit.tha}@gmail.com, {chanatip.sae,tawunrat.c}@kbtg.tech
* *equally contributed authors*

***Abstract*—Sentiment analysis is an essential task for social listening, especially in service and product analysis. Prior works on sentiment analysis, especially in Thai language, mostly focus on the improvement of model architecture without considering error propagation from word tokenizers or noisy text from social media. In this paper, three contributions are proposed for implementing social analysis model. First, text pre-processing is used to mitigate noise from input texts. Second, robustness towards word segmentation is enhanced by using an ensemble process with two tokenizers. Lastly, the training process inspired by Co-training method is proposed in order to filter label noise within the data. In the experiments, the model achieves 2.56% improvement on the average macro f-1 score when compared with the baseline models in social media data.**

***Keywords*—Sentiment analysis, Text classification, Text preprocessing, Word segmentation, Noise-cancelling algorithm**

I. Introduction

Recently, social listening has become a crucial method to gather information of customers from public channels such as Twitter and Facebook. Social listening, in its usual form, retrieves information and context from social channel. It then performs sentiment analysis to inspect each message and label its sentiment into "neutral", "positive", or "negative".

Prior studies on sentiment analysis proposed deep learning models, which outperforms traditional approaches such as Support Vector Machine and Naive Bayes [1], [2]. However, most of the studies do not focus on the impact of error propagation from word segmentation and noisy text. Moreover, the labels are inconsistent due to the subjectivity of the task, which focuses on detecting positive and negative instances. As a result, there are some instances that are annotated as neutral because they do not relate to the target domain and expected to be ignored.

To handle the mentioned problems, three contributions to improve sentiment analysis are proposed.

First, text processing is designed to reduce noises in Thai social text. Unnecessary parts of the input message, such as URLs and usernames, are removed. Duplicated spaces are replaced with one space. Moreover, a space is inserted between Thai and English characters to help word tokenizers identify word boundaries between them.

The second contribution is the ensemble model, which combines the output from different tokenizers. The ensemble model adds more perspectives on input tokens and therefore gains the benefit from both input tokens while tolerating false word segments caused by each tokenizer.

Finally, a label-noise filtering algorithm inspired by Cotraining [3] is adopted to explore label errors and imbalanced data set. This process is performed to avoid confusion caused by subjectivity from the labeling process. In the observed dataset, there are labeled instances which are annotated as neutral because they are not related to the target domain (although they contain positive or negative keywords). To clean up this ambiguous labeling process, training instances are re-labelled via probabilistic prediction from a trained model. Then, the re-labelled instances are used to train the new model. This process is our third contribution.

The remainder of this paper is structured as follows; the theoretical background of the related works is reviewed in Section II. Section III describes the proposed algorithms and architectures. The datasets, baseline models and evaluation metrics are explained in Section IV. After that, the results are analysed and discussed in Section V. Finally, the conclusion is in Section VI.

II. Related works

In this section, related works are presented in three subsections. They consist of sentiment analysis, the models for Thai word segmentation and label noise, and semi-supervised learning methods.

A. Sentiment analysis

Sentiment analysis is a method to retrieve emotional, opinion-oriented information from input texts [3]. Thai text classification and sentiment analysis using deep learning and statistical learning techniques have been studied in the recent years [1], [2]. For example, Kim Y. [4] proposed multiple branches of CNN with different kernel sizes to find grams (multiple sizes) for capturing the sentiment of messages. Meanwhile, Vateekul P. et al. [1] showed that deep neural network outperforms most of the traditional approaches such as Naïve Bayes and Logistic regression in social media data. Besides that, Yang Z. et al. [5] proposed the model called Hierarchical

Attention Network (HAN), which outperforms LSTM [6] in document classification. HAN utilizes a hierarchical structure

of Gated Recurrent Unit (GRU) with self-attention mechanism to understand the sentiment of messages by considering words in order and focusing on the important context related to the message.

In this paper, CNN and HAN are considered as the baseline models. CNN is widely considered as the state-of-the-art method for Thai sentiment analysis model [1]. HAN is also a competitive model for English document classification [5].

B. Word segmentation

Word segmentation is the process of dividing the written text into meaningful words. Most of the prior word segmentation models, such as Deepcut [7] and Sertis [8], are based on deep learning modules, including CNN and GRU. These models achieved 0.992 F1-score on NECTEC's BEST corpus [9]. However, employing these word segmentation models causes a significant problem. Since the models were trained on the NECTEC's BEST corpus containing formal Thai language, they are not suitable for texts from social media. On the contrary, KBTGTK [10] deep learning tokenizer was specially designed and trained with the UGWC dataset from social media. Due to the similarity between UGWC and our dataset, KBTGTK is chosen to be used in the experiment in this paper.

Besides deep learning word segmentation models, another approach of word segmentation called NEWMM [11] tokenizes the text to achieve the fewest number of words found in the dictionary.

NEWMM is a dictionary-based maximal matching algorithm, which is fast and only produces words appearing in the specified dictionary. Hence, NEWMM accurately segments known words and ignores noisy texts around the known words. While NEWMM does not face the same scalability problem other deep learning-based tokenizers, it cannot handle unknown words, such as name entities, that are not included in the specified dictionary.

Since KBTGTK and NEWMM have different advantages and drawbacks, both algorithms are explored in the experiments in Section V-B.

C. Label noise and semi-supervised learning

Large human-annotated dataset suffers from labelling errors due to various problems such as subjectivity, human errors, and notably the presence of label noise [12]. Label noise becomes a common problem when performing classification and causes many potential negative consequences. [12] shows that there are various approaches that could be used to handle label noise such as applying algorithms to avoid overfitting, improving quality of data using filtering approaches, and incorporating label noise as embedding in the model. The filtering approach is simple but efficient since it removes noisy data explicitly whereas the other approaches do not [12]. Hence, this paper uses a filtering approach to determine and relabel noisy instances to improve the quality of data.

Specifically, this paper applies a noise-filtering algorithm similar to Co-training [13], which is a semi-supervised learning method to utilize unlabelled data. The first step of Co-training is training two models with labelled data. After that, unlabelled data, which is predicted by the first-step models with high confidence, is labelled and used for training the models again. The method is performed repeatedly until there is no more instances of unlabelled data with prediction confidence level higher than the specified threshold. In this work, the method is used to correct noisy labelled data rather than using it to label unlabelled data.

III. Methodology

For this section, the methods that are used in the experiment are described, and the workflow of the process is shown in Figure 1.

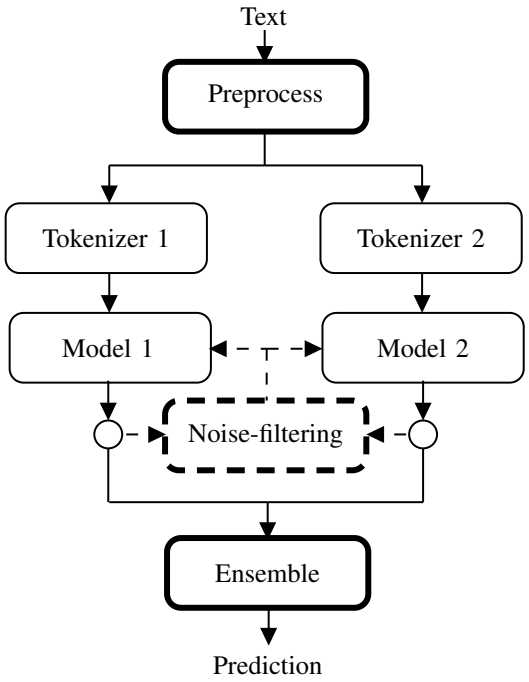


Figure 1. Work procedure of purposed methods. Bold line represents the proposed methods of this paper. Dash line represents the method applied only over training phase.

First, an input text goes through the text preprocessing as described in Section III-A. Second, for the ensemble process, the cleaned text is used for training and inference with two tokenizers and models. After that, in the training process, the output from both models are utilized to filter label noise as explained in Section III-B. In the inference process, two output from both models are combined using the ensemble method mentioned in Section III-C.

A. Text preprocessing

Text preprocessing is performed before word segmentation in order to clean inputs before tokenization and remove unnecessary parts of the text. Five steps of text preprocessing are performed as follows.

1) Convert all uppercase characters to be lowercase.
2) Remove characters that are not in Thai and English language.

3) Remove URL patterns and usernames, which are usually meaningless for sentiment analysis.
4) Add a space (“ ”) between Thai and English character, which is usually a word boundary.
5) Replace a sequence of duplicated spaces with one space.

B. Ensemble with Different Word Segmentation Algorithms

In this section, trained models are applied with ensemble methods, as shown below. The ensemble model is constructed from different tokenizers and models, as shown in Figure 1. The weighted average of the two probability is calculated from the output probability of each ensemble model.

$$P = \alpha * p_1 + (1 - \alpha) * p_2 \quad (1)$$

In Eq. 1, P denotes the output probability of the ensemble model. Meanwhile, p_1 and p_2 are the prediction probabilities given from different combinations of tokenizers and models. α is determined for weighting between two probabilities where $0 \leq \alpha \leq 1$.

C. Noise-filtering algorithm for imbalanced dataset

This section describes the noise-filtering algorithm used in the paper. The applied algorithm is inspired by Co-training [13]. However, instead of using the algorithm to label the unlabelled data, the algorithm is used to re-label the labelled data.

The algorithm is illustrated in Algorithm 1. m_1 and m_2 are the models trained from different aspects such as architectures or tokenizers. τ is the optimal threshold for relabeling the neutral instance. The threshold is tuned to obtain the best macro-F1 score whereas $\tau \in \{0.05, 0.10, 0.15, ..., 1.00\}$. p_{class} is the probability for identical class given from averaging of probabilities of m_1 and m_2.

Algorithm 1: Noise-filtering algorithm for sentiment analysis dataset

Input: m_1: 1^{st} model to be used as filter, m_2: 2^{nd} model to be used as filter, U: train dataset
Output: m: model after being applied noise-filtering algorithm

```
/* train m1 and m2 on U                                   */
m1 ⟵ train_model(U);
m2 ⟵ train_model(U);
/* Loop to relabel an instance                            */
for x ∈ U do
    if x is "neutral" then
        /* average probabilities from models              */
        p1 ⟵ m1.predict(x);
        p2 ⟵ m2.predict(x);
        p_neutral, p_positive, p_negative = (p1 + p2)/2;
        /* relabel an instance with probability
           lower than the threshold                       */
        if p_neutral ≤ τ then
            x ⟵ max(P_positive, P_negative)
        end
    end
end
/* train model with the relabeled dataset                 */
m ⟵ train_model(U);
```

IV. Experimental Setup

The processes of setting and conducting experiment are explained in this section.

A. Model

In the experiments, Hierarchical attention neural networks (HAN) [5], and the convolutional neural networks (CNN) [4] are used with their weights initialized randomly. Early stopping is also used to avoid overfitting problem based on macro-F1 of validation dataset. Moreover, Adam optimizer [14] with 0.001 initial learning rate is used to optimize the categorical cross-entropy loss. In our experiments, AllenNLP framework is used for implementing the models [15].

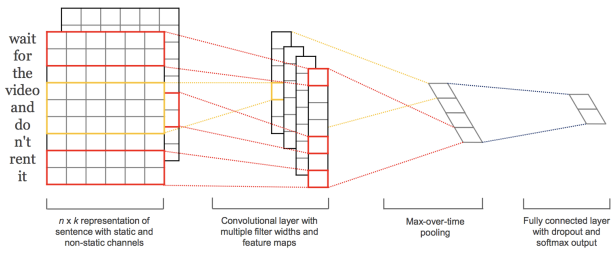


Figure 2. Convolutional neural networks for text classification [4]

Moreover, for both CNN and HAN models, the input is a sequence of 200 embedded words. Each word is embedded into a vector with 256 dimensions without pre-trained word embedding before being dispatched into the models. Each model maps input text to sentiment class probabilities.

The architecture of CNN, which is proposed for text classification [4], is shown in Figure 2. In this work, CNN has three branches of convolutional layers with max pooling. Before applying max-pooling operation, ReLU [16] have been applied to the feature maps output of every convolutional layer. The convolutional layer in each branch has 50 filters with different kernel sizes; 3x256, 4x256 and 5x256. The filters outputs from different branches are concatenated and flattened into a single vector. Then the vector is passed through a softmax layer to get sentiment class probabilities.

Meanwhile, Figure 3 illustrates the architecture of HAN for text classification. In the model, the sequence of word vectors is dispatched into bidirectional GRU with 128 nodes. Then, the attention mechanism (at word level) is applied to create a sequence of 200 vectors with 256 dimensions. After that, the vectors are forwarded into bidirectional GRU and attention mechanism (at the sentence level) with the same procedure and finally projected into a vector used for classification via a softmax layer to retrieve the class probabilities.

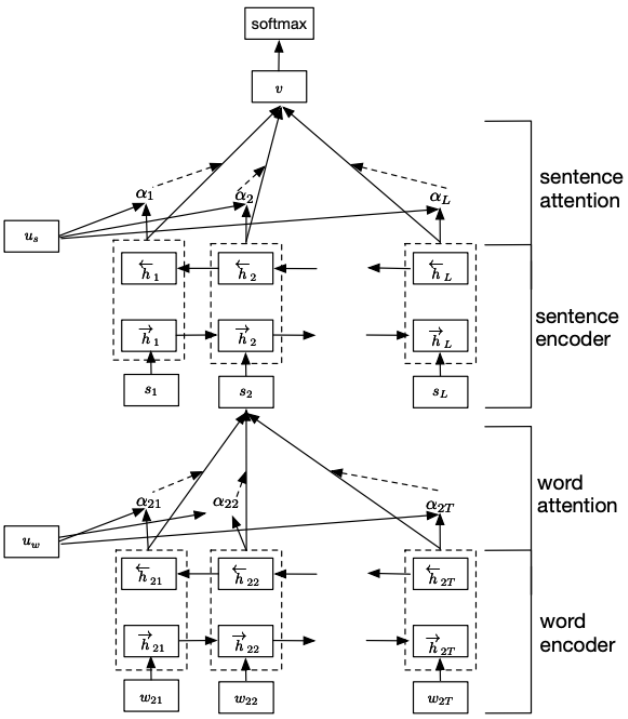


Figure 3. The architecture of Hierarchical Attention Networks (HAN) [5]

B. Dataset

Table I
THE LABEL DISTRIBUTION OF A SENTIMENT DATASET

Class	*Number of Sample*	*Percentage*
Neutral	267,397	82.73
Positive	14,769	4.56
Negative	41,030	12.69

The dataset is collected from social media platforms such as Facebook and Twitter. The 323,196 samples of them are conducted in this experiment and manually labelled by human annotators. The labels are "neutral", "positive" and "negative" of which distributions are shown in Table I.

This dataset contains neutral instances for more than 80%, which causes imbalanced data issue. Most instances are annotated as neutral due to two reasons. First, the data set contains a large portion of advertisement, blogs, and news (which are normally written with unbiased contents). Second, some instances that are not related to the target domain are considered neutral regardless of any positive and negative meaning.

In the experiments, the dataset is divided into train set, validation set, and test set with proportion 80%, 10%, and 10% respectively. The validation set are used for hyperparameter tuning and performing early stopping.

C. Evaluation metric

$$precision_{\text{class}} = \frac{\text{True positive}}{\text{True positive + False positive}}$$

$$recall_{\text{class}} = \frac{\text{True positive}}{\text{True positive + False negative}}$$

$$F1_{\text{class}} = \frac{2 \cdot precision_{\text{class}} \cdot recall_{\text{class}}}{precision_{\text{class}} + recall_{\text{class}}}$$

$$F1_{\text{macro}} = \frac{F1_{\text{neutral}} + F1_{\text{positive}} + F1_{\text{negative}}}{3} \quad (2)$$

For all experiments, the models are evaluated based on macro-F1 as each class is treated equally, as shown in Eq. 2. The reported score is the average macro-F1 calculated from three models, which are initialized with different random seeds.

V. RESULTS AND DISCUSSIONS

This section presents the experimental results and discusses the effect of the three proposed ideas on the results.

A. Effect of our text preprocessing on metrics

Text preprocessing is applied toward different combinations of models and tokenizers. The results are shown in Table II which text preprocessing enhances the model macro-F1 in all combinations of models and tokenizers. The models with text preprocessing improve slightly on average macro-F1 score (0.60%). Especially, HAN with KBTGTK gains 1.37% macro-F1 improvement from text preprocessing.

B. Ensemble with different tokenizers

In this experiment, the models that are used for assembling are selected with three different aspects. First, the models, which are assembled, use the same architecture and tokenizer, but they are trained with different random seeds. Second, the ensemble model is constructed from the models with different tokenizers but with the same architecture. Finally, the models with different architectures and tokenizers are assembled.

From Table II, the ensemble model with identical word segmentation algorithms (NEWMM + NEWMM and KBTGTK + KBTGTK) slightly improves the macro-F1 in both models. When NEWMM + NEWMM is applied, the macro-F1 scores are improved from 71.95% to 72.91% (+ 0.96%) in CNN and 70.85% to 71.25% (+ 0.40%) in HAN. Meanwhile, the macro-F1 scores of the model with KBTGTK + KBTGTk is raised from 70.63% to 71.45% (+ 0.82%) in CNN and 71.25% to 71.75% (+ 0.50%) in HAN.

Moreover, two different word segmentation algorithms are also assembled. The ensemble model with NEWMM + KBTGTK achieves better accuracy in term of the averaged macro-F1, increased from 71.75% to 72.89% (+1.14%) in HAN and 72.91% to 73.06% (+0.15%) in CNN. This result shows that the model exploits the benefit of both tokenizers (NEWMM and KBTGTK). The input, which is tokenized by NEWMM, is ensured that the given words are found in the dictionary.

Table II
THE RESULTS OF TEXT PROCESSING AND ENSEMBLE USED IN EXPERIMENTS

Model 1	**Tokenizer 1**	**Model 1**	**Tokenizer 2**	**F1 macro**	
				- Preprocessing	***+ Preprocessing***
HAN	NEWMM	-	-	70.71	70.85
	KBTGTK	-	-	69.88	71.25
	NEWMM	HAN	NEWMM	71.56	71.25
	KBTGTK		KBTGTK	70.30	71.75
	NEWMM		KBTGTK	71.95	72.89
CNN	NEWMM	-	-	71.55	71.95
	KBTGTK	-	-	70.01	70.63
	NEWMM	CNN	NEWMM	72.47	72.91
	KBTGTK		KBTGTK	70.97	71.45
	NEWMM		KBTGTK	72.78	73.06
HAN	KBTGTK	CNN	NEWMM	**72.90**	**73.67**

Table III
THE RESULTS OF APPLYING NOISE-FILTERING ALGORITHM FOR IMBALANCED DATASET USED IN EXPERIMENTS

Model	**F1 macro**		
	- Pre - NF	***+ Pre - NF***	***+ Pre + NF***
HAN + KBTGTK	70.71	70.85	71.51
CNN + NEWMM	71.55	71.95	72.44
Ensemble Model	72.90	73.67	73.69

Note: **Pre** denotes preprocessing process. Meanwhile, **NF** denotes noise filtering process.

Meanwhile, the Out-of-Vocabulary (OOV) words are handled by KBTGTK. Thus, utilizing NEWMM and KBTGTK could enhance the robustness and diminish errors caused by each tokenizer.

In addition, CNN and HAN are assembled together as the ensemble model to gain a higher macro-F1 score. For word segmentation, the tokenizer that is combined with each model and achieves the higher macro-F1 score is selected. NEWMM is selected for CNN. Meanwhile, HAN is used with KBTGTK. The ensemble model achieves the highest averaged macro-F1 at 73.67%.

C. Noise-filtering algorithm for imbalanced dataset

In this experiment, the best ensemble model given the best results (HAN + KBTGTK and CNN + NEWMM) is used as the model for reducing noise in the proposed noise filtering algorithm. The improvements of various models are noticeable after applying our noise-filtering algorithm. In Table III, the results show that the models with noise filtering improve slightly on macro-F1 from 72.67% to 72.69% in the ensemble model. Also, the improvement of the individual models (HAN + KBTG and CNN + NEWMM) are measured, after the noise-filtering algorithm is applied. The macro-F1 of HAN with KBTGTK tokenizer improves from 70.85% to 71.51%. Meanwhile, the macro-F1 of CNN with NEWMM tokenizer increases from 71.95% to 72.44%.

Furthermore, the label distribution after relabelling by the noise filtering is examined and shown in Table IV. The label distributions of both models with noise-filtering are changed in a similar way. The number of neutral instances decreases while the numbers of other label instances increase. These results demonstrate subjectivity in sentiment analysis labelling tasks.

In the data set, some instances not related to the target domain are annotated as neutral regardless of positive and negative keywords. For instance, "สัด กุดูตั้ง 4 นาทีกว่า คิดว่าโฆษณาหนังใหม่เหอะ" is annotated as neutral because the instance is about a released video and is not related to the target domain (i.e., finance). However, the word "สัด" is obviously a negative keyword, Therefore, the trained model is easily confused with these conflicting neutral instances. After integrating label noise-filtering, these neutral instances are relabelled as positive and negative. This leads to a decrease in the number of neutral instances as shown in Table IV.

By performing label-noise filtering, the model can focus on sentiment keywords and ignore the bias from the annotators.

Table IV
DISTRIBUTION OF TRAINING DATASET (BEFORE AND AFTER APPLYING NOISE-FILTERING)

Model	**Tokenizer**	τ	***Class***	**Percentage (before $\rightarrow$ after)**
HAN	KBTGTK	0.60	neutral	82.73 $\rightarrow$ 78.92
			positive	4.56 $\rightarrow$ 5.89
			negative	12.70 $\rightarrow$ 15.17
CNN	NEWMM	0.50	neutral	82.73 $\rightarrow$ 80.38
			positive	4.56 $\rightarrow$ 5.38
			negative	12.70 $\rightarrow$ 14.23

Moreover, the imbalanced data issue is also cured by this process. As a result, the model improves marginally.

D. Overall improvement of the model

In Table III, the model with the combination of all proposed contributions reaches 73.69% macro-F1, which is the highest performance among all models in the experiment. To evaluate the model (HAN + KBTGTK and CNN + NEWMM) and models with none of the three proposed contributions are considered as baseline models. The macro-F1 of the model improves by 2.98% and 2.14% compared to HAN + KBTGTK and CNN + NEWMM baseline models, respectively.

VI. Conclusion

In this paper, three contributions were conducted to improve sentiment analysis. First, five steps of text preprocessing were applied to reduce input noise. After getting rid of the noise, the macro-F1 score of each model improved 0.60% on average. Second, the ensemble model with different tokenizers was also proposed to tolerate the error produced by a single tokenizer. The result showed that assembling two tokenizers (NEWMM and KBTGTK) could improve the macro-F1 of HAN and CNN by 1.14% and 0.15%, respectively. Moreover, by combining two models (HAN and CNN), the macro-F1 increased from 73.06% to 73.67%. Finally, to address the subjectivity of sentiment analysis task, the labelled noise was handled by a noise-filtering algorithm. The model gains the improvement of macro-F1 at 0.58%. The model that contains all three contributions yielded 2.98% and 2.14% improvement on macro-F1 for HAN and CNN, respectively.

However, the model described in this paper could be further integrated with intention classifiers. Therefore, intention and sentiment could be an automated system for social listening that can evaluate the branding image of products and companies.

Acknowledgment

We would like to thank you the linguist team: Dr. Supawat Taerunruang and Dr. Nutcha Tirasaroj for their invaluable data analysis. Also, this work was supported by Kasikorn Business-Technology Group (KBTG).

References

[1] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on thai twitter data," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016, pp. 1–6.

[2] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment analysis using deep learning techniques: a review," *Int J Adv Comput Sci Appl*, vol. 8, no. 6, p. 424, 2017.

[3] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[4] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[5] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.

[6] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1422–1432.

[7] R. Kittinaradorn, "A thai word tokenization library using deep neural network." [Online]. Available: https://github.com/rkcosmos/deepcut

[8] J. Jousimo, "Thai word segmentation with bi-directional rnn." [Online]. Available: https://sertiscorp.com/thai-word-segmentation-with-bi-directional_rnn

[9] K. Kosawat, M. Boriboon, P. Chootrakool, A. Chotimongkol, S. Klaithin, S. Kongyoung, K. Kriengket, S. Phaholphinyo, S. Purodakananda, T. Thanakulwarapas *et al.*, "Best 2009: Thai word segmentation software contest," in *2009 Eighth International Symposium on Natural Language Processing*. IEEE, 2009, pp. 83–88.

[10] A. Lertpiya, T. Chaiwachirasak, N. Maharattanamalai, T. Lapjaturapit, T. Chalothorn, N. Tirasaroj, and E. Chuangsuwanich, "A preliminary study on fundamental thai nlp tasks for user-generated web content," in *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*. IEEE, 2018, pp. 1–8.

[11] K. Chaovavanich, "Dictionary-based thai word segmentation using maximal matching algorithm and thai character cluster." [Online]. Available: https://github.com/PyThaiNLP/pythainlp

[12] B. Frénay and M. Verleysen, "Classification in the presence of label noise: a survey," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 5, pp. 845–869, 2013.

[13] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*. Citeseer, 1998, pp. 92–100.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. F. Liu, M. Peters, M. Schmitz, and L. S. Zettlemoyer, "Allennlp: A deep semantic natural language processing platform," 2017.

[16] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

A Classification Model for Thai Statement Sentiments by deep learning techniques

line 1: 1st Given Name Surname
line 2: *dept. name of organization*
(of Affiliation)
line 3: *name of organization*
(of Affiliation)
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization*
(of Affiliation)
line 3: *name of organization*
(of Affiliation)
line 4: City, Country
line 5: email address or ORCID

***Abstract*— At present, many organizations realized the importance of sentiment analysis for consumer reviews. The positive and negative comments can help to evaluate the user satisfaction of products and services to control and improve their qualities. In addition, the deep learning techniques are very interesting methods for current researches in the data mining field. Therefore, this research studied on the deep learning techniques to analyzed user reviews and comments from the TripAdvisor website. To begin with, Thai user comments in four categories: hotels, restaurants, tourist attractions, and airlines were collected and tested on the combination of two basic deep learning technique that are convolutional neural network and long-short term memory. All user comments were divided into individual statements to classify into three groups: positive feelings, negative feelings, non-expressed feelings or neutrality. The research results found that the best classification model is the combination of three convolutional neural networks with 32, 64, and 128 filters, respectively, and the kernel size of 2 equal to the three components. Moreover, the performance of the proposed classification model was evaluated by the accuracy and the precision values were higher than 80% in all groups.**

Keywords—sentiment analysis, sentiment classification model, convolutional neural network

I. INTRODUCTION

Information is available in a variety form with different contents at present. One of them is the subjective statement, which has a lot of content on the internet. That information can be used in many ways, such as creating recommendation systems, sentiment analysis, and other data mining systems about consumer satisfactory. In addition, many organizations perceive the importance of analyzing feelings from messages, because if user comments or opinion can be separated into positive or negative messages, it can be the benefit of controlling the product quality or maintaining the service quality, including improving the quality of products or services.

One way of the implementation of sentiment analysis system from texts is using the machine learning techniques with the sentiment corpus, that contains a word list expressing positive and negative feelings. However, it is difficult to identify every sense of the whole words in a sentence. Then, building the classification model for determining the sense of the messages will be benefit for generating useful information from positive or negative feeling automatically. A statement or a message can be classified into three categories: messages expressing positive feelings, messages displaying negative feelings, and messages that do not show feelings or neutral messages.

There are some researches about the sentiment analysis in Thai. The article [1] used the deep learning techniques: Dynamic Convolutional Neural Network (DCNN) and Long-Short Term Memory (LSTM) to classify the Thai Twitter messages (tweets). In addition, the accuracy of result classification by these two deep learning techniques is compared to the classical machine learning techniques: Naïve Bayes, Support Vector Machines (SVMs), and Maximum Entropy. The results were shown that the accuracy of both deep learning techniques is higher than that of NBs and SVMs, but less than ME. On the other hand, the study [2] collected and analyzed a set of tweets [3] in Thai using three machine learning techniques to classify the sentiments of messages. This research discovered tweets based on time duration and used Latent Dirichlet Allocation (LDA) to mention the messages topics. The performance of sentiment classification from all these machine learning techniques which are Naïve Bayes, SVMs, and Maximum Entropy, are very high. All evaluated values: the accuracy, the precision of positive and negative messages, and the recall of both classes are more than 95%, including the F-measure score.

The next related paper [4] about Thai sentiments analyzed Thai user reviews of two mobile applications from Google Play Store which are a virtual keyboard app called "แม่น แม่น (Maen Maen)" and an online TV app named "H-TV". To analyze aspects (the topics of messages) with sentiment words, this work applied the natural language processing technique and the lexicon-based approach that are SentiWordNet [5] a Thai-English dictionary called LEXiTRON [6]. Additionally, Latent Dirichlet Allocation (LDA) were used to discover the aspect. However, one limitation of this sentiment analysis was depending on words in SentiWordNet and LEXiTRON. To solve this problem, another research [7] implemented the sentiment analysis application on English tweets about skin care products using combining the machine learning algorithms with the word information in SentiWordNet. Naïve Bayes and SVMs are two machine learning techniques for identifying the positive or negative tweets. Then, the levels of positive and negative emotion in messages were separated into five levels: very positive, positive, neutral, negative, very negative. The performance of result classification is calculated by the accuracy, the precision, and the recall rate, and all evaluated values are more than 75%.

The last interesting paper about sentiment analysis is the article [8], which is a review of sentiment analysis using the deep learning. There are various artificial neural networks that were deployed to predict the sentiments on different datasets, i.e., images from Twitter, tweets, micro-blog comments, movie reviews, hotel reviews, reviews from

Amazon, news, and political articles. The examples of these deep learning techniques are a CNN, a Recursive Neural Network (RNN), a Deep Neural Network (DNN), LSTM, a Probabilistic Neural Network (PNN), and the combination of them. The paper concluded that the deep learning techniques can apply to solve the variety of problems. Moreover, some results of the sentiment analysis had the high accuracy.

For all previous reasons, our research developed the classification model by two basic deep machine learnings for Thai sentiment analysis. The artificial neural networks in the form of CNN and LSTM were tried and tested on Thai statements of reviews for identifying the sentiments. All statements were divided into three categories: messages that express positive feelings, messages that present negative feelings, and messages that do not display feelings or neutral statements.

II. Background Knowledge

A. Thai Sentiments

In every language, there are words that can express emotions and feelings in positive and negative sentiments. The examples of these words are กลัว/klua (fear) โกรธ/krot (anger) รำคาญ/ram-kha (annoyance) ยุ่งยากลำบาก/yungyak-lambak (difficulty) เศร้า/saw (sadness) ไม่ชอบ/mai-chop (dislike) ความไม่พอใจ/khwam-mai-phocai (displeasure) สดชื่น/sotchuen (freshness) เอาใจใส่/awcaisai (caring) เชื่อมั่น/chueaman (confident) มีความสุข/mi-khwamsuk (happy) สนุกสนาน/sanuksanan (fun) and ร่าเริง/raroeng (cheerful). The consumers' opinions on products or services can be analyzed to positive or negative comments by word sentiments. The positive words can mean users feel satisfied or agree with that, while the negative words can show the feeling that they are not satisfied or do not agree with that.

B. Machine Learning and Deep Learning

The machine learning is a system or a program to learn and solve problems or make decisions by itself automatically. Most of applications have to use a lot of information as training data to train systems or programs to make decisions and give answers correctly using creating models for solving problems. Therefore, unlike traditional systems or programs, there are input data which will be proceeded to generate output data or to get answers. The machine learning techniques use input with answers for training to create models and serve models to find answers [9]. One of the machine learning techniques, which is the current famous learning, is called the deep learning.

The deep learning is the artificial neural network, which simulates the process of the neural network in the human brain (each neuron is connected by a nerve). There are three main parts that are 1. the input layer 2. the hidden layer 3. The output layer (the neurons that send the data after processing). Each part is linked by artificial nerves with individual weight connected by multiple layers of the nervous system as shown in Fig. 1. There are many nodes in hidden layer which calculate the values according to the specified activation function. Nodes in different layers are connected by signal lines with their weight to control the flow of data between the connected nodes. Therefore, the learning of the artificial neural network is caused by adjusting the weight of each signal line to be suitable for the transmitted information and the answer output [10].

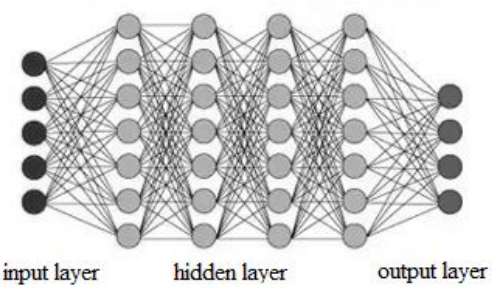


Fig. 1. Overview of deep learning neural network

The weight of each signal will be adjusted according to each loss of an output or a result. The derivatives of the loss values are compared to the derivatives of the results to calculate values for adjusting the weight using the non-linear equation/function. In addition, there is a variety of optimization algorithms that helps to calculate and adjust each weight of signal lines in the artificial neural network. In our research, there are two interesting types of the artificial neural network, which are CNNs and LSTM neural networks, used in the fields of sentiment analysis. CNNs can classify words or statements in the form of text format [11] and LSTM neural networks are able to capable of supporting time series data in which text-based information consists of a sequence of words, e.g. resulting in a sentence, a clause or phrase [12].

CNNs can distinguish features of data into smaller features as sub-characteristics by using comparative calculations. Filters and the kernel are implemented to pull out the interesting features. Normally, one filter with the kernel will pull out one of the focused features, so multiple filters work together for extracting characteristics of data. LSTM is also one of the recurrent neural networks, which is created to simulate the patterns of peoples' memory with limited memory capacity. When new events come into memory, the brain will choose to remember or not, thus LSTM can support sequential data.

C. Evaluation of the deep learning in Sentiments

The deep learning is often used for problems with large data sets. Therefore, an effective test is needed to evaluate the performance of the model in invisible data and compare the performance with other configurations reliably. In general, all data will be divided into training dataset, validation dataset and test dataset as shown in Fig. 2 [13].

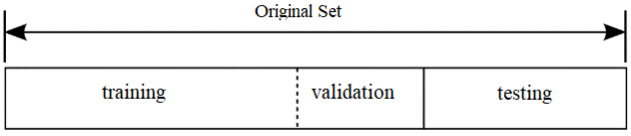


Fig. 2. Splited data for the performance evaluation.

Moreover, sentiments can be separated into three categories that are positive, negative, and neutral. The predictive data may be different from the actual data, and the result of classification may be correct or incorrect. A confusion matrix of three class, which displays the comparison of actual classes and predicted classes among all classes, is represented in Table I. A, E, and I data values are correctly predicted sentiments for positive, negative, and neutral classes, respectively. The performance evaluation values that are accuracy, precision and recall will be calculated by equations in Table II.

TABLE I. A CONFUSION MATRIX

Actual class	Predicted class		
	Positive	***Negative***	***Neutral***
Positive	A	B	C
Negative	D	E	F
Neutral	G	H	I

TABLE II. THE PERFORMANCE EVALUATION VALUES

Table Head	Table Column Head		
	Accuracy	***Precision***	***Recall***
Positive	A+E+I)/ (A+B+C+D+E +F+G+H+I)	A/(A+D+G)	A/(A+B+C)
Negative		E/(E+B+H)	E/(E+D+F)
Neutral		I/(I+C+F)	I/(I+G+H)

Furthermore, the overall of classification performance will be denoted by *F1 score* and calculated by (1).

$$F1\ Score = 2*((Precision*Recall)/(Precision + Recall)) \quad (1)$$

III. A PROPOSED CLASSIFICATION MODEL

A. Data Collection

Messages from the web system of TripAdvisor (https://th.tripadvisor.com [14]) were collected to develop the sentiment classification model. This we system is popular for Thai people to find accommodation, restaurants, tourist attractions, and airline services. The system allows users to freely review about various places and services. In addition, all comment messages were categories into different groups. Our research focused on four categories that are hotels, restaurants, and airlines. A total of collected statements are 12,596 messages. These statements are divided into 3 sentiment labels which are positive (label 0), negative (label 1), and neutral (label 2) as shown in Table III. The examples of positive comments are displayed in Fig 3.

B. Data Analysis

To test on the various combination of CNN and LSTM neural networks, 7,200 statements of four categories (1,800 statements per each category: hotels, restaurants, tourist attractions and airlines) were randomly selected in the same ratio of sentiment labels (600 messages per each sentiment and each group) as the balanced training data for generating classification model. After that, the word segmentation was performed by a Thai tokenization library called deepcut [15]. This library extracts words in the form of a token array as expressed in Fig 4. Then, all unique tokens were converted to specific numbers that are the number 1 to the number 7,240, so there are 7,240 unique words from 7,200 statements and Fig. 5 shows the examples of unique words with their specific numbers.

TABLE III. ALL COLLECTED DATA

Category	Class Label			
	Positive	***Negative***	***Neutral***	***Total***
Hotels	1,544	1,183	713	3,440
Restaurants	1,401	713	920	3,034
Tourist Attractions	1,286	723	1,027	3,036
Airlines	1,470	636	980	3,086
Total	5,701	3,255	3,640	12,596

```
อาหารซีฟู้ดสดมากเหมือนขึ้นมาจากทะเลใหม่ ไม่ต้องไปกินไกลถึงทะเล,0
ราคาไม่แรงมากพอรับได้ เหมาะสมกับราคาค่ะ เดินทางสะดวก,0
ร้านตกแต่งน่ารัก บรรยากาศสงบ,0
บรรยากาศร้านดีมาก อาหารก็ตกแต่งสวยและอร่อย ,0
 บรรยากาศตกแต่งดูดีมากเน้นเป็นสีขาว,0
แต่รสชาดก็ถือว่าใช้ได้ อาหารเสริฟค่อนข้างเร็วที่เดียวไม่ต้องเสียเวลานาน,0
```

Fig. 3. Examples of positive comments

C. Experiemnts for Selecting Deep Learning Techniques

In our research, there are many experiments to build the classification model by the combination of CNN and LSTM neural networks. All 7,200 data were randomly separated into 5,400 training data and 1,800 validation data. The efficiency by calculated accuracy in Table II with 4-fold cross validation and the average running time of learning for predicting validation data are applied to select the suitable deep learning techniques. The combination of CNN and LSTM with the higher percentage of accuracy mean the more effective classification. If there are similar percentages of accuracy, the one that takes less time to learn is more effective.

```
อยู่ ห่าง จาก ปาก ซอย พอสมควร โรง แรม ใหม่ สะอาด ที่ นอน สบาย มี ม่าน ปิด ห้องน่า ดี มาก
เหม็น ดัง เตียง อึดอัด ห้องน้ำ สกปรก ห้อง พัก กล่อง
แนะนำ หมู ย อคะ ห่อ ละ 20 บาท ได้ เยอะ ไม่ มี แป้ง คะ
ห้อง พัก สวยงาม สะดวกสบาย แอร์เย็นฉ่ำ
พนักงาน ไม่ เป็น มิตร เลย ตั้งแต่ พนักงาน ส่วนหน้า ไป จนถึง ตาม ชั้น
ห้องน้ำ สะอาด มาก น้ำ ร้อน ไหล แรง แอร์เย็น ฉ่ำ มี ระเบียง ด้วย สำหรับ คน ดูด บุหรี่
ร้าน อยู่ ริม แม่น้ำน่าน เลย คับ ทาน อาหาร ไป ชม วิว แม่น้ำน่าน ไป ดื่ม ด่ำ บรรยากาศ รับประทาน อาหาร อร่อย ๆ ๆ ไป พนักงาน บริการดี
```

Fig. 4. Example of word segmentation results

```
{'ที่': 1, 'ไม่': 2, 'มี': 3, 'มาก': 4, 'ดี': 5, 'ได้': 6, 'และ': 7, 'ไป': 8, 'ๆ': 9, 'เป็น': 10, 'การ': 11, 'อาหาร': 12, 'ก็': 13, 'ให
': 14, 'มา': 15, 'จะ': 16, 'ใน': 17, 'แค่': 18, 'บริการ': 19, 'ราคา': 20, 'ร้าน': 21, 'บิน': 22, 'ของ': 23, 'กับ': 24, 'ห้อง': 25, 'เค
รื่อง': 26, 'พนักงาน': 27, 'นี่': 28, 'ว่า': 29, 'พัก': 30, 'เลย': 31, 'อยู่': 32, 'อร่อย': 33, 'นั่ง': 34, 'จาก': 35, 'เดินทาง': 36, 'เรา':
37, 'คน': 38, 'โรง': 39, 'ทำ': 40, 'ต้อง': 41, 'เวลา': 42, 'ความ': 43, 'คะ': 44, 'แรม': 45, 'แล้ว': 46, 'อย่าง': 47, 'ด้วย': 48, 'ไข
': 49, 'สาย': 50, 'รถ': 51, 'เพราะ': 52, 'นี่': 53, 'สะอาด': 54, 'สะดวก': 55, 'น่า': 56, 'น้ำ': 57, 'เลือก': 58, 'อีก': 59, 'ครั้ง': 60,
'แพง': 61, 'ถึง': 62, 'สำหรับ': 63, 'โดย': 64, 'ดู': 65, 'ทุก': 66, 'เข้า': 67, 'ชอบ': 68, 'กว่า': 69, 'แบบ': 70, 'กัน': 71, 'ยัง': 7
2, 'บน': 73, 'ถ้า': 74, 'ทั้ง': 75, 'ขึ้น': 76, 'วัน': 77, 'ครับ': 78, 'ค่อนข้าง': 79, 'ตัว': 80, 'ถูก': 81, 'ฉัน': 82, 'คือ': 83, 'เยอะ':
84, 'ตรง': 85, 'สามารถ': 86, 'ส่วน': 87, 'ทาง': 88, 'ใกล้': 89, 'ทาน': 90, 'ใหญ่': 91, 'ออก': 92, 'ค่า': 93, 'ดอน': 94, 'เหมือน': 9
5, 'หรือ': 96, 'หลาย': 97, 'กลับ': 98, 'เดิน': 99, 'กิน': 100, 'สบาย': 101, 'ถือ': 102, 'รสชาติ': 103, 'ประทับใจ': 104, 'เที่ยว': 105,
'บรรยากาศ': 106, 'จอง': 107, 'รับ': 108, 'รวม': 109, 'ไทย': 110, 'นั้น': 111, 'ค่อย': 112, 'เล็ก': 113, 'แนะนำ': 114, 'คุ้ม': 115, 'มั
น': 116, 'ต่อ': 117, 'แย่': 118, 'เข้า': 119, 'เมนู': 120, 'กระเป๋า': 121, 'เช็ค': 122, 'นะ': 123, 'ใหม่': 124, 'หน้า': 125, 'ประหยัด': 1
26, 'ผู้': 127, 'สั่ง': 128, 'ตั้ง': 129, 'ก่อน': 130, 'เคย': 131, 'เท่า': 132, 'ผม': 133, 'แรก': 134, 'จอด': 135, 'คุณ': 136, 'คิด': 13
```

Fig. 5. Example of positive statements

Initially, a single artificial neural network was build using CNN that there are two important values to adjust for more accurate predictions:

- The number of filters is the number of considered features to be desired from the data.
- The size of the kernel is the width of the kernel that will be used to find the feature.

The symbol CNN(the number of filters, the size of the kernel) is represented to the convolutional neural network with specific filter and kernel values, e.g. CNN(32,2) means the one-dimensional CNN with 32 filters and a kernel size of 3. The deep learning needs to enter the same amount of input features, and there is the suggestion of the deep learning that the number of input data should be greater than or equal to the maximum number of words contained in statements for model training. Therefore, our experiments of selecting deep learning techniques tried on the input data at 100 words and 150 words for each sentence of 7,200 experimental statement data. The results of testing on the number of input nodes are shown in Table IV.

TABLE IV. RESULTS OF TESTING ON THE NUMBER OF INPUT NODES

Models	Percentage of Accuracy / Time (sec)
CNN(32,1) 100 input	80.68% / 237 s
CNN(32,1) 150 input	80.29% / 278 s
CNN(32,2) 100 input	81.17% / 277 s
CNN(32,2) 150 input	81.31% / 329 s

According to TABLE IV, the number of input data nodes was specified as the number 100 because of shorter processing time with the similar percentage of accuracy. Therefore, messages with fewer words than the number of input nodes were filled with the number 0 in front of the unique number of word tokens equal to the amount of data imported to the model as shown in Fig 6. On the other hand, only the first 100 word -tokens of the statement with the length more than 100 words were imported to be input data.

```
[   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
    0,    0,    0,    0,    0,    0,    0,  164,   54,    4,   57,
  253,  760,  422,  896, 1446,    3, 1162,   48,   63,   38, 4471,
  575],
```

Fig. 6. Example of input data of one statement.

Most of the CNN applications used a kernel size of 3 [12], so the different number of filters were run with a fixed kernel size to find the suitable number of filters. The best filter value was 32 as the displayed result in TABLE V. After getting the right number of filters, then there were experiments to find the right kernel size again as shown in TABLE VI.

TABLE V. RESULTS OF TESTING ON ADJUSTING FILTER VALUES

Models	Percentage of Accuracy / time (sec)
CNN(16,3)	78.95/334
CNN(32,3)	80.57/358
CNN(64,3)	80.45/396
CNN(128,3)	80.50/442

TABLE VI. RESULTS OF TESTING ON ADJUSTING KERNEL SIZE

Models	Percentage of Accuracy / time (sec)
CNN(32,1)	80.68/237
CNN(32,2)	81.17/277
CNN(32,3)	80.57/358

According to TABLE VI, the CNN (32,2) was the best single layered neural network to classify the sentiments. Next, there were experiments to test on double and triple layered CNNs that may identify the sentiments of statements precisely. The results of these experiments were demonstrated in TABLE VII.

TABLE VII. RESULTS OF TESTING ON DOUBLE AND TRIPLE LAYERED CNNS

Models	Percentage of Accuracy / time (sec)
CNN(32,2)+ CNN(16,2)	81.47/316
CNN(32,2)+ CNN(32,2)	81.69/326
CNN(32,2)+ CNN(64,2)	81.69/302
CNN(32,2)+ CNN(128,2)	81.43/357
CNN(32,2)+ CNN(64,1)	81.11/305
CNN(32,2)+ CNN(64,3)	81.69/336
CNN(32,2)+CNN(64,2)+ CNN(128,2)	82.02/403

Referring to TABLE VII, there were tests for changing both filter and kernel size values of the double layered CNNs. The double layered CNNs with the highest accuracy and the lowest processing time was the first layer CNN(32,2) connected to the second layer CNN(64,2). There was one observation that the increasing filter values and changing in kernel sizes for the second layer did not get the better performance. Consequently, the third layer of CNN with an increase in filters were implemented. The result found that the triple layered CNNs with the number of filters 32, 64, and 128, respectively using a kernel size of 2 can classify the sentiments most effectively. In addition, the 4 layered CNNs with the number of filters equal to 32, 64, 128, and 256 were tested, but the percentage of accuracy was less than 82% with longer processing time. Therefore, the experiments on multiple layered CNNs were stopped to test on the LSTM neural network.

For the LSTM neural network, the number of memory cells is only one value that affects the accuracy. The symbol LSTM(the number of memory cells) is represented to the LSTM neural network with specific memory cell values, e.g. LSTM(1) means the LSTM neural network that has one memory cell. The result of testing on adjusting the number of memory cell for the LSTM neural network was shown in TABLE VIII. The single LSTM neural network with two memory cells got the best performance.

TABLE VIII. RESULTS OF TESTING ON ADJUSTING MEMORY CELLS

Models	Percentage of Accuracy / time (sec)
LSTM(1)	59.50/619
LSTM(2)	73.92/617
LSTM(3)	69.80/627

TABLE IX displayed the result of testing on the different number of memory cells in the second layer of double LSTM neural networks.

TABLE IX. RESULTS OF TESTING ON DOUBLE LAYERED LSTM

Models	Percentage of Accuracy / time (sec)
LSTM(2)+ LSTM(1)	61.99/872
LSTM(2)+ LSTM(2)	69.25/876
LSTM(2)+ LSTM(3)	72.75/917
LSTM(2)+ LSTM(4)	68.37/961

According to TABLE VIII and IX, the highest accuracy of multiple LSTM neural network was about 70% with taking long processing time. Therefore, the experiments on the combination of CNN and LSTM neural networks were run to find the suitable deep learning technique for sentiment classification. The results of some experiments on various combination of CNN and LSTM were demonstrated on TABLE X.

TABLE X. RESULTS OF TESTING ON THE COMBINATION OF CNNS AND LSTM

Models	Percentage of Accuracy / time (sec)
CNN(32,2)+LSTM(1)	62.03/538
CNN(32,2)+LSTM(2)	76.05/574
CNN(32,2)+LSTM(3)	77.07/546
CNN(32,2)+LSTM(10)	78.99/553
CNN(32,2)+LSTM(11)	80.01/542
CNN(32,2)+LSTM(12)	78.61/559
CNN(32,2)+LSTM(15)	77.68/567
CNN(32,2)+CNN(64,2)+LSTM(11)	78.70/608
CNN(32,2)+CNN(64,2)+LSTM(13)	78.90/616
CNN(32,2)+CNN(64,2)+LSTM(15)	79.55/689
CNN(32,2)+CNN(64,2)+LSTM(20)	79.28/668
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(1)	32.27/711
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(2)	68.55/722
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(3)	68.69/711
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(5)	78.05/701
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(10)	81.02/752
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(12)	82.20/805
CNN(32,2)+CNN(64,2)+CNN(128,2)+LSTM(13)	80.88/771
LSTM(2)+CNN(32,2)+CNN(64,2)+CNN(128,2)	80.01/784
LSTM(2)+CNN(32,2)	79.47/713
LSTM(2)+CNN(64,2)	79.74/641
LSTM(2)+CNN(128,2)	74.36/671
LSTM(2)+CNN(32,1)	75.11/602
LSTM(2)+CNN(32,3)	78.01/672
LSTM(2)+CNN(32,2)+CNN(64,2)	76.39/698

Although the combination of triple layered CNNs and the LSTM with 12 memory cells had the highest performance (82.20% of accuracy), this technique for generating classification model took a long time for processing. So, the triple layered CNNs with the number of filters equal to 32, 64, and 128, respectively using a kernel size of 2 were selected to create the sentiment classification model.

D. Model Developmemt

The implementation of the proposed deep learning technique uses the Jupyter Notebook[16] (python editor) and the Keras library [17] (deep learning library) to generate Thai sentiment classification model.

IV. EXPERIMENTAL RESULTS

The evaluation of proposed deep learning technique was separated into two processes. The first process was the performance of sentiment classification on the validation data, and the second process was on the testing data. Therefore, all collected data were divided into 3 parts which are training data for generating the model, validation data for checking the efficiency of the model and testing data for evaluating the performance of the model on unseen data. All collected 12,596 messages were separated into 15% of testing data (1,890 messages), and 85% of training data with validation data (10,706 messages). In addition, the ratio of training data and validation data was 80:20, so there were 8,564 training statements and 2,142 validation statements. The categories of data are shown in TABLE XI.

TABLE XI. RESULTS OF TESTING ON THE NUMBER OF INPUT NODES

Category	Class Label			Total
	Positive	*Negative*	*Neutral*	
Training	3,898	2,220	2,446	8,564
Validation	947	550	645	2,142
Testing	856	485	549	1,890
Total	5,701	3,255	3,640	1,2596

The results of classifying the sentiments of statements on validation data and testing data in the form of a confusion matrix were represented in TABLE XII and TABLE XIII, respectively. The performance of classification model on validation data and testing data were demonstrated on TABLE XIV and TABLE XV, respectively.

TABLE XII. A CONFUSION MATRIX OF EVALUATING ON VALIDATION DATA

True Label	Predicted Label			Total
	Positive	*Negative*	*Neutral*	
Positive	856	37	54	947
Negative	50	453	47	550
Neutral	76	68	501	645

TABLE XIII. A CONFUSION MATRIX OF EVALUATION ON TESTING DATA (UNSEEN DATA)

True Label	Predicted Label			Total
	Positive	*Negative*	*Neutral*	
Positive	765	947	54	856
Negative	26	550	35	485
Neutral	80	645	409	549

TABLE XIV. THE PERFORMANCE OF CLASSIFICATION ON VALIDATION DATA

True Label	Accuracy	Precision	Recall	F1 Score
Positive	84.50%	87.16%	90.39%	0.89
Negative		81.18%	82.36%	0.82
Neutral		83.22%	77.67%	0.80

TABLE XV. THE PERDORMANCE OF CLASSIFICATION ON TESTING DATA (UNSEEN DATA)

True Label	Accuracy	Precision	Recall	F1 Score
Positive	84.55%	87.83%	89.37%	0.89
Negative		81.38%	87.42%	0.84
Neutral		82.13%	74.50%	0.78

According to TABLE XIV and XV, the classification model has the best performance on the positive statements in both the validation data and the testing data. All evaluation values (accuracy, precision and recall) of positive class are about 85% and higher. In the same way, there are good classification performance for the negative statements because all evaluation values are higher than 80%. All F1 score values are greater than 0.75 for all data groups, that means the overall performance of classification model is good in precision and recall rates. Therefore, the result can be concluded that the proposed model is able to classify the sentiments of statements effectively.

V. CONCLUSION

The proposed deep learning technique for identifying statement sentiments is the combination of triple layered CNNs which has filters of 32, 64 and 128 respectively, with the same kernel size of 2 in all three layers. The implementation was developed by Python language on the Jupiter Notebook and Keras library. The performance of classification model was evaluated by two confusion matrices on validation and testing data set. The result found that positive and negative statements can be classified exactly with all evaluation values more than 80% and some values up to 90%.

REFERENCES

[1] P. Vateekul, and T. Koomsubha, "A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data," in Proceeding of the 13th International Joint Conference on Computer Science and Software Engineering (JCSSE2016), Khon Kaen, Thailand, 2016

[2] J. Lertsiwaporn, and T. Senivongse, "Time-Based Visualization Tool for Topic Modeling and Sentiment Analysis of Twitter Messages," in Proceeding of the International MultiConference of Engineers and Computer Scientists (IMECS 2017), Hong Kong, 2017

[3] Twitter. https://twitter.com/.

[4] B. Deewattananon, and U. Sammapun, "Analyzing User Reviews in Thai Language toward Aspects in Mobile Applications," in Proceeding of the 14th International Joint Conference on Computer Science and Software Engineering (JCSSE2017), Nakhon Si Thammarat, Thailand, 2017

[5] Text Learning Group. SentiWordNet. http://sentiwordnet.isti.cnr.it/

[6] National Electronics and Computer Technology Center, "LEXiTRON," https://lexitron.nectec.or.th.

[7] P. Pugsee, V. Nussiri, and W. Kittirungruang, "Opinion mining for skin care products on twitter," in Communications in Computer and Information Science, vol.937, pp. 261-271, 2019.

[8] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, and A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review," International Journal of Advanced Computer Science and Applications, vol. 8, no. 6, pp. 424-433, 2017.

[9] V. Minaphinant. Whai is Machine Learning? (in Thai), 2018. https://blog.finnomena.com/machine-learning-fa8bf6663c07.

[10] L. Zhang, S. Wang and B. Liu, "Deep Learning for Sentiment Analysis: A Survey," arXiv, 2018. https://arxiv.org/ftp/arxiv/papers/1801/1801.07883.pdf

[11] C. Zhou, C. Sun, Z. Liu and F. C.M. Lau, "A C-LSTM Neural Network for Text Classification,". arXiv, 2015. https://arxiv.org/pdf/1511.08630.pdf

[12] X. Zhang, J. Zhao and Y. LeCun, "Character-level Convolutional Networks for Text Classification," in Proceedings of the 28th International Conference on Neural Information Processing Systems, vol. 1, pp. 649-657, Montreal, Canada: MIT Press Cambridge, 2015. https://papers.nips.cc/paper/5782-character-level-convolutional-networks-for-text-classification.pdf.

[13] T. Borovicka, M. Jirina Jr., P. Kordik and M. Jirina, "Selecting Representative Data Sets," in Advances in Data Mining Knowledge Discovery and Applications, intechOpen, 2012. https://www.intechopen.com/books/advances-in-data-mining-knowledge-discovery-and-applications/selecting-representative-data-sets

[14] TripAdvisor. https://th.tripadvisor.com/.

[15] Deepcut. https://libraries.io/pypi/deepcut.

[16] The Jupyter Notebook. https://jupyter.org/

[17] Keras Library. https://keras.io/

The analysis for quantitative evaluation of palpation skills in maternity nursing

Shunya Inoue
Informatics Course
Kochi University of Technology
Kochi, Japan
225113m@gs.kochi-tech.ac.jp

Sumika Yoshimura
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
yoshimuras@kochi-u.ac.jp

Miwa Saito
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
saitomiw@kochi-u.ac.jp

Kyoko Yamawaki
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
y-kyou@kochi-u.ac.jp

Mikifumi Shikida
School of Information
Kochi University of Technology
Kochi, Japan
shikida.mikifumi@kochi-tech.ac.jp

***Abstract*—In recent nursing education, it is said that there is a big gap between the practical ability of novice nurses and the level required in clinical site, and it is required to enhance the practical experience of nursing students and improve the educational effect. This research focuses Leopold Maneuvers and clarify whether the palpation according to the textbook or not. The palpation according to textbook tends to examine with less pressure and slowly touching tather than applying excessive pressure. So quantitative evaluation of touch pressure can improve effective education of future nurses.**

***Index Terms*—Maternity nursing education, Palpation training, Qualitative evaluate support, Pressure data analysis**

I. Introduction

In recently years, the environment surrounding health care and medical is changing greatly, and various changes have occurred in the work contents of nurse staff [1]. For example, it is indispensable to improve the clinical practice ability of nursing staff in order for nursing staff to adopt appropriately to issues such as advances in medical technology and ensuring medical safety. In particular, novice nursing staff need organized and systematic efforts to improve their clinical practice skills [1]. The Japan Nursing Association also is aiming expansion of medical care and improvement of people's life qualitatively, and quantitatively [2], but it is pointed out that there are differences in the achievement level of nursing skills because each school has set specific goals for achieving practical nursing skills until they graduate from basic nursing education. In addition, the scope and opportunities of nursing skills training for real patients tend to be limited so far [1], and problems such as ensuring nursing training facilities, education according to the learning ability of nursing students, and training of supervisors have been pointed out [3]. In particular, in maternity nursing on-site training, the number of facilities that can be used as on-site training facilities will be declining due to the recent declining birthrate in Japan [4] [5]. In the undergraduate process in nursing education, it is important that nursing students master basic skills to deepen their expertise throughout their lives [6], and exercises using simulated patients in basic nursing education have important educational effects such as accurately grasping the image of the subject of nursing [7] [8]. However, in one certain university, there are problems such as advices by a supervisor in simultaneous exercises by multiple nursing students and the lack of nursing students who can experience the exercises multiple times. Furthermore, in the achievement assessment of nursing practice ability required before graduation, it is important that the evaluation itself functions to improve and enhance the curriculum [6], but since the evaluation method during the exercise is generally qualitative, it is not easy for the supervisor to make all the nursing students understand the supervisor's experience and sense of skill, and as a result, the educational effect varies among students. The training in basic nursing education is an important educational process for developing high-quality nursing staff. However, the various issues and current situation of nursing education described above show that effective teaching and evaluation for all nursing students is insufficient.

Therefore, in this research, we collected pressure data during palpation in an experiment simulating maternal nursing practice, and performed data analysis and statistical hypothesis test. We also discussed the usefulness of quantitative palpation technique evaluation and the possibility of improving educational effectiveness through quantitative evaluation.

II. Review of related literature

A. Changes in hands pressure during palpation

Kaetsu et al. have researched the relevance between hand pressure and the posture of the nurse when changing the face-up position to lateral position [9]. On the experiment in this research, a pressure measuring films were used to measure hand pressure. If pressure measurement film is used to measure pressure, we can see the detailed maximum pressure of the

whole hand, while we cannot know the change in pressure over time.

The purpose of this research is to clarify the characteristics of how to use the finger palm during palpation, so we use a pressure sensor that can collect pressure data over time. In the experimental devices used in this research, multiple pressure sensors are fixed at specific positions on the finger palm, and the pressure of the finger palm during palpation can be collected from multiple positions over time.

B. Skills acquisition

Fukutani et al. have proposed the method that enables quantitative evaluation of qualitative evaluation in technical classes at junior high school [10]. Previous technical teachers had evaluated the work produced by students visually and qualitatively, and in them research, they proposed the quantitative evaluation method using smartphones. Although their method was able to confirm the effect of reducing the burden of teachers' evaluation work, the teaching regarding skills and educational impact of teachers were not taken into account.

This research aims not only quantitative evaluation of nursing skills, but also to improve educational effects. We collect palpation data of nursing students and supervisors and analyze how to use the palms in palpation skills.

C. Understanding of own palpation

Hosozawa et al. have developed a system to support clinical nurses acquire efficient physical assessment skills [11]. In the system they proposed, a pressure sensor was attached to the abdominal simulator for nursing. In addition, the pressure distribution and the place where the center of gravity is operating are displayed on the display simultaneously with the image being palpated. It supports nursing students understand the abdominal position and pressure changes that experienced nurses palpate. However, in clinical site, it is necessary to be able to practice nursing skills suitable for each patient. Therefore, it is important that nursing students deeply understand how to use the finger palm during their palpation at the stage of palpation training, which are basic nursing education.

In this research, we decided to wear a pressure sensor on the finger palm as a method of collecting pressure data of the finger palm during palpation. By analyzing which fingers and palms were used during palpation, we aim for effective learning tailored to nursing students so that nursing students can deeply understand their own palpation.

III. Data collection experiment

A. Experiment conditions

The experimental environment is shown in Fig.1.

We cooperated with Department of Nursing, School of Medicine in Kochi university for this experiment. The experiment is conducted in the environment shown Fig.1. There were a total of 20 subjects in this experiment, including 10 third-year undergraduate students, 6 second-year graduate students, and 4 faculty members at the university's nursing department. All subjects are right-handed female and are physically and mentally healthy. On this experiment, the subjects who wear experimental devices conduct the experiment which assumed maternity training Leopold Maneuvers first phase and second phase in the university. The object of palpation is a doll modeled a pregnant women. In the first phase, each subject palpates upper abdomen of the doll by her both hands, and understand the presentation. The presentation is the positional relationship between the womb and the fetus. In the second phase, each subject palpates both side abdomen of the doll by her each hand, and understand the position of presentation. The position of the presentation is the orientation of the back or face of the fetus. The condition that height of the bed and the air pressure of the doll's abdomen were also unified.

Fig. 1. The experiment environment

The goal of undergraduate students set by the university is that they can palpate according to the textbook. They have the knowledge regarding Leopold Maneuvers but do not experience the training of Leopold Maneuvers. So maternity nursing supervisor explained to them about the training in the same way actual the training in the university. Specifically, the explanation were brief reviews of the first and second phase of Leopold Maneuvers by verbal and white board, and observation of the supervisor's performance. And we told them that the results of the experiment had no bearing on their future assessments. Also, all postgraduate students have high-level of training and practical experience.

Before the experiment, when the subject finished palpating the each phase, we asked her to give any sign such as "The first phase was finished" and we asked come off her hands from the doll at the same time. The subject stand left side, and she starts the palpation as the same time our sign. We record her palpation by video, and the start time of recording video, finish time of first phase and second phase.

B. Experiment Devices

The experimental devices is shown in Fig.2.

The subject wears a black waist pouch containing several microcomputers and fixes the 16 sensitive-pressure sensors to specific positions on both fingertips and palms. The sensitive-

Fig. 2. The example of the palpation simulated the experience

pressure sensors used were 0.5 inch circular sensitive-pressure sensors from the company Spark Fan Electronics, and the Arduino UNO from the company Arduino were used as the microcomputers. The microcomputers send all the sensitive-pressure sensors' value to a collecting computer every 30ms. We used an Apple MacBook as the collecting computer. The collecting computer saves the values sent from the microcomputers in CSV format.

The sensitive-pressure sensors is shown in Fig.2.

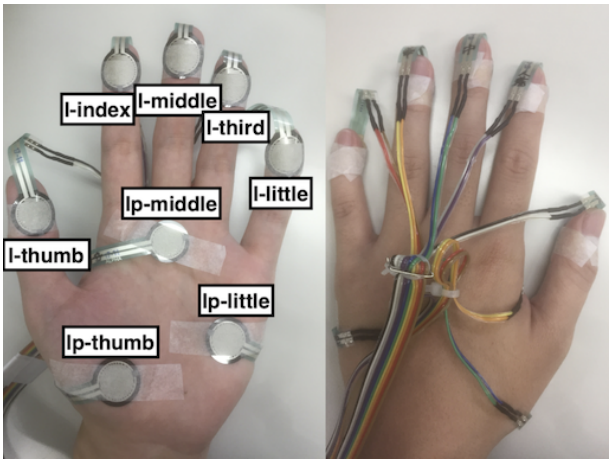


Fig. 3. The positions of each sensor

In this paper, each position of the sensitive-pressure sensors are called as shown in the Fig.3. Specifically, the sensors attached to the fingertips are named "little", "third", "middle", "index" , "thumb" in order from the little finger. And if the fingertip is of the left hand, "l-" is added to head to the fingertip's name, on the other hand, if the fingertip is of the right hand, "r-" is added to head of the right fingertip's name. Also, in the lower part of the little finger, the lower part of the middle finger, and the lower part of the thumb on the palm, if the position is on the left hand, "lp-" is added to the head of the name, on the other hand if the position is on the right hand, "rp-" is added to head of the name. There are 16 the sensors, 8 each in one hand, and paper white tape was used as the tape for fixing each the sensor. The position of each the sensor was united for each subject. The top of every fingertips' sensors were fixed upper 5 millimeter position from each distal interphalangeal joint. The side of both palms' sensor under little finger were fixed the position under 30 millimeter from the metacarpophalangeal joint of each little finger. The side of both palms' sensor under middle finger were fixed the position under 7 millimeter from the metacarpophalangeal joint of each middle finger. The side of both palms' sensor under thumb finger were fixed the position upper 20 millimeter from each wrist of both. The upper limit value that can be recorded by the sensor is 1023, but when the author applied a strong pressure to the sensor, the upper limit value was 900.

IV. Result of the analysis and Statistical hypothesis test

A. Analysis of the values collected

Before the data analysis, we pre-processed for data analysis accurately. First, the time of each value were adjusted to the multiple of 30 ms because the time of some value were not accurate record. Then, if there were the time not recorded, the time were complemented based on the before and after time, and the values were also complemented by the mean value of the before and after values. Next, we confirmed a few outlier caused by any problem of the experiment devices, so we confirmed the video and changed the value to 0. Finally, the value under 50 were changed to 0 for removal noises caused by the experiment devices.

For each subject, the total value of all fingertips and palms is used as the denominator and the value of each fingertip or palm is used as the numerator. The fingertip or palm which have largest rate of that ratio was defined as the "primary finger palm" of the fingertips and palms that was most used by the subjects during palpation. Also we considered it is the important thing that the combination between the position of each abdomen and number of hands if we defined the primary finger palm. In first phase, the subjects palpate upper abdomen by them both hands, on the another, in second phase, they palpate right abdomen by them left hand and left abdomen by right hand. So, we define the primary finger palm for each position of the abdomen. The primary finger palm of each subject is shown in Tab.I. Subjects starting with 'b' mean undergraduate students, subjects starting with 'm' mean postgraduate students, and subjects starting with 'n' mean supervisors.

The maximum value, average value, maximum rate of change, positive mean rate of change, negative mean rate of change and minimum rate of change were calculated for each subject's primary finger palm as the result of analysis. The change means the value that the difference between previous value and current value was divided by 30. And the goal of this university is "Every undergraduate students can palpate according to the textbook", so we separated two groups the analysis result for each abdomen. After this on this paper, N_0

TABLE I
THE PRIMARY FINGER PALM OF EACH SUBJECT

Abdomen	*Upper*	*Right*	*Left*
b-0	l-middle	lp-little	rp-thumb
b-1	l-middle	lp-thumb	r-middle
b-2	l-third	lp-little	r-third
b-3	l-third	lp-middle	r-index
b-4	l-third	lp-thumb	rp-third
b-5	l-index	lp-middle	r-middle
b-6	rp-little	lp-little	rp-thumb
b-7	rp-little	lp-middle	rp-thumb
b-8	rp-little	lp-middle	rp-middle
b-9	l-middle	lp-middle	rp-thumb
m-0	rp-little	lp-little	rp-middle
m-1	rp-little	lp-middle	rp-middle
m-2	l-third	lp-little	r-third
m-3	rp-little	lp-little	r-middle
m-4	lp-middle	lp-little	rp-thumb
m-5	rp-little	lp-thumb	r-index
n-0	l-middle	lp-middle	rp-middle
n-1	l-middle	lp-middle	rp-middle
n-2	rp-little	lp-middle	rp-middle
n-3	rp-little	lp-middle	rp-middle

means the group of palpation according to the textbook and N_1 means that not. Also the judgement of the separation was based on whether the primary finger palm of each subject is same with that of subject maternity nursing supervisor or not. So, the number of N_0 and N_1 is different from the palpation of each abdomen. The mean values for each analysis result is shown in Tab.I.

TABLE II
THE MEAN VALUES FOR EACH ANALYSIS RESULT

Abdomen	*Upper*	
Group	N_0	N_1
Max	347.889	642
Mean	183.483	363.654
Max rate of change	4.211	4.898
Positive mean rate of change	0.563	1.206
Negative mean rate of change	-0.556	-1.745
Min rate of change	-3.826	-7.321
Abdomen	*Right*	
Group	N_0	N_1
Max	341.7	650
Mean	172.962	319.624
Max rate of change	2.947	4.693
Positive mean rate of change	0.387	0.825
Negative mean rate of change	-0.482	-1.043
Min rate of change	-3.313	-5.623
Abdomen	*Left*	
Group	N_0	N_1
Max	260	545.462
Mean	133.459	261.766
Max rate of change	3.781	4.6
Positive mean rate of change	0.546	0.868
Negative mean rate of change	-0.554	-1.148
Min rate of change	-3.557	-4.769

B. Statistical difference between the two groups

We statistically tested the analysis results between two groups for each abdomens. Before testing whether there was a difference between the two groups, we performed a Shapiro-Wilk test to determine whether the population of the data set was normally distributed. As a result, the population of some data sets did not follow the normal distribution, so the Mann-Whitney U test was performed with a one-sided test in this test. Also, it is confirmed that the absolute value of the median value of N_0 is smaller than the absolute value of the median value of N_1 between all two groups . The null hypothesis H_0 is "There is no difference in the median value of the two groups", and the alternative hypothesis H_1 is "the difference in the median value between the two groups. If the same rank does not exist in the data set, the adoption of the null hypothesis is determined by the statistics and the rejection limit value, and if the same rank exists, the adoption of the null hypothesis is determined by the p-value. The test results are shown in Tab.III.

TABLE III
MANN-WHITNEY U TEST

Abdomen	*Upper*	*Right*	*Left*
Max	Reject*	Reject*	Reject*
Mean	Reject*	Reject*	Reject*
Max rate of change	Accept	Reject*	Accept
Positive mean rate of change	Reject*	Reject*	Accept
Negative mean rate of change	Accept	Accept	Accept
Min rate of change	Accept	Reject*	Accept

*: $p<0.01$

V. DISCUSSION

A. Importance of primary finger palm

On the palpation to each abdomen, it is the important information for nurse students to understand how to use own finger palm and the difference how to use finger palm between own and supervisor. On the maternity in this university, the supervisor teach how to palpate according to the textbook to undergradutate nursing students. In the first phase, the palpation way is mainly used under little finger with them both hands curved from distal interphalangeal joint of every fingers. In the second phase, the palpation way is mainly used center of them hand with every finger outstretched. If we focus on the supervisors in the Tab. I, the primary finger palm differs depending on each supervisor in the first phase, on the other hands, the primary finger palms of every supervisor is all same between them in the second phase. On the first phase, it seems that it is easy to differ primary of each subject because they palpate with them hands curved. Actually, the each fingertip available touching to upper abdomen is changed depending on the degree of them hands' curved and the degree of turning them wrist. In addition, we confirmed two supervisor of N_1 used "lp-little" or "rp-little" on the palpation to the upper abdomen, so it cannot be said that they have palpated using a different method of palpation than textbooks. On the other

hands, in the second phase, the subjects mainly palpate with them outstretched hands, so it seems that the primary finger palms of every supervisors were same.

There is a feature that the primary finger palm of every subjects is a part of them palm on palpation to right abdomen, but there is no feature on palpation to left. It seems that the feature caused by the position of the subjects and them postures. In the case of this palpation, right abdomen is located in the front for the subjects, and left abdomen is located in the back for them. In this conditions, it seems that it is hard for the subjects to palpate the left abdomen with them center of right hand without them devised postures.

From the above, we cannot overlook the relation between each abdomen and the primary finger palm on whether them palpation is according to textbook or not. Therefore, the information regarding the primary finger palm on each abdomen and the tend is extremely important for the realization of effectively education on the palpation training.

B. High-quality palpation

The important difference between actual clinical site and training is whether the object of the palpation is real patients or not. The high-quality nursing skills to real patients are not only nursing care based on speciality knowledge but also the skills make the patients feel relieved. Therefore, the palpation with less touching pressure is important factor in the view of high-quality nursing skills.

"Max" and "Mean" in the Tab.III show that there are significant differences between N_0 and N_1 on the palpation for each abdomen. Also, maximum and average values in the Tab.II show that the every pressure values of the N_0 are half of the N_1. So, regarding the strength of the pressure to each abdomen, N_0 is less pressure than N_1 statistically. Also, regarding the difference between the pressure to each abdomen, it clarified that the subjects of N_0 can palpate with half hands' pressure of N_1

In addition, right abdomen in the Tab.III shows there are significant differences between N_0 and N_1 on the "Max rate of change", "Positive mean rate of change" and "Min rate of change". On the palpation to the right abdomen, the subjects of N_0 tend to adjust the pressure slowly compered to the subjects of N_1. On the other hand, there is no significant differences statistically on the palpation to the left abdomen. It seems that these also caused by the position of the subjects and them postures. It seems that it was easy for the subjects to adjust the strength of pressure to right abdomen which located in the front compared with left abdomen. Actually, we confirmed the rate of both fingertip and palm most used, and it showed that 18 subjects used any finger or palm of left hand. However, on the change rate which accepted by the test, Tab.II shows every change rates of N_0 is near 0 than N_1, so it seems that subjects of N_0 adjusted them hands' pressure well than N_1.

From the above, there are many differences regarding the pressure of palpation for each abdomen and also the change rate between N_0 and N_1. In the view of effective education, it seems that there are important relations between primary finger palm and the pressure and the change rate.

C. Usefulness to effective education

We discuss effective education on the palpation training on this research in the two view.

First of the view is support to supervisor. In previous palpation training, in the case of this university, supervisor confirmed palpation of several nursing students by her eyes, and taught them if necessary. However, more effective education would be realized if the palpation in the same time, confirmation by eyes and qualitative teaching is improved. For example, even if several palpation training are in the same time, the overlooks of teaching by supervisor will decrease more than previous teaching way. Also, it will be possible that it supports for advices and evaluations to nursing students even if the contents of teaching are qualitative and hard to make them understand. The prevention of the overlooks and ease understanding of teaching seem important for effective education of nursing students.

Second of the view is the possibility according to level of each nursing student. The palpation data shows that which finger palm is used, how is touch of the palpation, how much is the pressure and so on. Until now, it is hard for nursing students to review own palpation skills after the training. However, it will be possible that the nursing students deepen to understand own palpation skills compared with before if the data regarding them touch on the palpation training. Also, it is possible compared the difference between the palpation of the nursing students and that of supervisor. So, supervisor will be able to teach and evaluate according to the level of each nursing student effectively, and it will be easy for the nursing students to learn about palpation skills than before. It seems that if the nursing students' understanding such as own features of finger palm, pressure value and etc on the palpation is important regarding the usefulness of effective education on the training.

VI. Conclusion

In recent nursing education, it is said that there is a big gap between the practical ability of new nurses and the level required in clinical site, and it is required to improve the training and practical training of the nursing students and increase the educational effect. Therefore, in this research, we focused on Leopold Maneuvers training among maternity nursing, and revealed differences in the finger palm used most and pressure differences between those who palpated according to the textbook and those who were not it was. From the results, it was clarified that the group that was palpated according to the textbook had less pressure applied and the action of applying pressure was gradual compared to the group that not. Finally, we discussed that quantitative analysis of finger palm usage during the palpation and its application to education would be useful for improving educational effectiveness and developing high-quality human resources.

REFERENCES

[1] Ministry of Health, Labor and Welfare,"Study Report on Improvement of Practical Practical Training Ability of Novice Nursing Staff," 2004.

[2] Japan Nursing Association, "Nursing Challenge for 2025 Future Vision of Nursing, Nursing for life, living and dignity," June, 2015.

[3] Keiko Itagaki, "Current status and issues of nursing education," Bulletin of Department of Nursing, Tohoku Bunka Gakuen University, vol. 4, No. 1, March 2015.

[4] Japan Cabinet Office, "FY2018, status of declining birthrate and overview of measures to cope with declining birthrate," Low birthrate society measures report, June 2019.

[5] Mika Shishido, Tomomi Omori, Kyoko Kubo, Hiroe Fujimura, "The actual state of maternity nursing practice in the nurse training course," Bulletin of Department of Nursing, Saitama Medical University, vol. 5, No. 1, pp. 47–53, March 2012.

[6] Ministry of Education, Culture, Sports, Science and Technology,"Achieving Goals at University Graduation for Enhancing Nursing Practice Capability Development (Report of Study Group on Nursing Education)," 26, March, 2004.

[7] Yumiko Matsui, "On the effect of Pediatric Nursing Practicum: In Light of the Results of Post-Practicum Questionnaires," Niigata journal of health and welfare, vol.9, No.2, pp.31–38.

[8] Miki Fukuma, Yuko Tsumoto, Hiromi Uchida, Yoshiko Sahara, Emiko Tarui, Kyoko Osada, "Effects and Problems of the Teaching Method on the Nursing Process Using Simulated Patient," Bulletin of Shimane University School of Medicine, Vol.29, pp.15–21, December, 2006.

[9] Mie Kaetsu, Naoko Hirahara, Shihoko Nomura, "Relationship of Nurse' s Motion and Hand Pressure to Patient' s Experience when Being Repositioned," Japanese journal of nursing research, vol. 36, No. 2, 2013.

[10] Ryota Fukutani, Akinobu Ando, Shota Itagaki, Hideyuki Takahashi, Tetsuo Kinoshita, IPSJ, CDS, vol. 7, pp. 51–63, May, 2017.

[11] Ayumi Hosozawa, Ryota Shibusawa, Hiroki Yamamoto, Hiroaki Yuze, Tomoyasu Aoyama, Naoyoshi Suzuki, "The Development of Support System for Physical Assessment Training for Nurses: Results and Future Research Directions," Vol.2009-CE-99, No.7, 23, May, 2009.

A Framework of Computer-Based Learning System Based on Self-Regulated Model in English Writing

***Abstract*—This paper presents a design phase of a computer-based learning system for English writing in Thai EFL learners. This system is designed to incorporate the self-regulated model and set the components of linguistics and machine translation as a learning environment. The system is designed based on three main phases of self-regulated model: forethought phase, performance phase, and self-reflection phase. The learning environment used to guide completely target sentence writing. Moreover, the display of user interface is designed for using as assisting tool for supporting a student self-regulated learning in English writing. There are three main modules of the system that consist of learning profile acquisition, learning behavior collection, and learning analytics. The system design is an important phase to encourages action between learners and computer-based learning system for English writing. Then, learners behavior are collected into data logs store for learning analysis. This system aims to collect Thai EFL learners behavior and find the behavioral pattern that could helpful reference for improve system and teaching materials in the future.**

***Index Terms*—computer-based learning system, self-regulated, machine translation, English writing**

I. Introduction

In language learning field, since English language is not a native language or first language (L1) of Thai learners and not national language of Thailand, Thai learners who learn English language are called English as a Foreign Language (EFL) learner. There are four main skills of English language: listening, speaking, reading, and writing. For Thai EFL learners, grammatical mistakes are often occurred in writings. Sometimes the English sentences are not convey what the writer would like to express. Therefore, the English writing is hard task for Thai EFL learners because they are not native English language and have limited chance to often write in English [1].

The effective strategies need to be emphasized for performance of Thai EFL learners in language learning. The learners are expected to be self-regulated in learning process [2]. The self-regulated model is one of the effective strategies for improving performance of learning [3]. The incorporating self-regulated strategies into foreign language learning supports the development of autonomous learners [4]. There are three main phases of self-regulated model: forethought phase, performance phase, and self-reflection phase [5], [6]. The self-regulated theory is an encouraging model for interactions of personal, behavioral, and environmental factors for effective learning [7].

The self-regulated model is needed to apply in digital learning for successful digital learning. This model is importance for digital learning to provide a learning environment conducive to learning through the development and support of reflective, self-regulated skills [8]. The digital technology are became ubiquitous in all levels of education [3], [9]. For this work, the learning environment in the computer-based learning system uses guideline components in Natural Language Processing (NLP) that consists of components of linguistics and Machine Translation (MT). These components are assisted in English writing. For MT, Thai learners mostly think in Thai (native language: L1) and translate. However, they do not understand the whole translation process. But machine translation involves understanding both languages in their translation processes. So, these components in MT could help learners to learn proper English writing. Moreover, MT is known as extensive digital tool use for support English writing in the field of foreign language learning. There are various reasons for using MT in foreign language learning such as helping increase vocabulary, helping increase grammatical accuracy, and saving the time [10]. Therefore, MT technology can be used for the learning environment of EFL learning to encourage interactive behavior between personal and learning environment.

In this paper, we aim to describe the design of computer-based learning system which designed based on self-regulated model and integrated components of linguistic and MT as a learning environment. The designed system affects to increase learners self-regulated and writing proficiency. The self-regulated strategies could support learner to set up goal, monitor learners behavior, and reflect own writing performance. Furthermore, the learning environment with components of linguistic and MT could encourage action between learners and system to create English sentences. This paper is structured as following: section 2 explains the system design based on the self-regulated model and system processing. Section3 describes an example of learner scenario. Finally, section 4 gives a discussion of the learning system and conclusion.

II. Computer-Based Learning System for English Writing

A. *System Overview*

The computer-based learning system of English writing is designed to record learning behaviors automatically while writing English sentences. The system based on self-regulated model and using components of linguistics and MT as a learning environment. This system consists of three main modules: learning profile acquisition module, learning behavior collection module, and learning analytics module. The system overview as illustrated in Fig. 1. All modules work

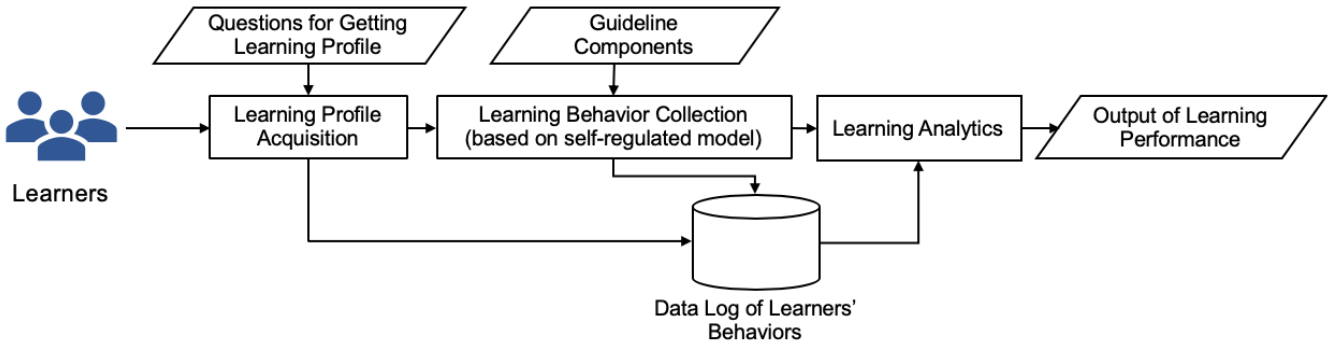


Fig. 1. An overview of computer-based learning system based on self-regulated model in English writing

coherently that start with the learning profile acquisition. Firstly, the learning profile acquisition is primary module that get learner information about existing English writing skill. Secondly, the learning behavior collection is learning behaviors recording module that works with data log storage and component of linguistics/MT. Moreover, This module is designed to encourage increasing of learners self-regulated by applying strategies of self-regulated model into process. Finally, learning analytics is a behavior analysis module that analyzes writing behavior from the log file. The result from behavior analysis will conduct to find the learning pattern of English writing in the Thai EFL learner.

*1) **Learning Profile Acquisition**:* The learning profile acquisition is an initial module that aims to collect learner information i.e. personal information, education information, and English grammatical skill information. Firstly, learners give personal and education information. Next, learners answer the provided questions for getting existing grammatical skill of writing. The existing grammatical skills are required for this process consist of three aspects: vocabulary usage, sentence type usage, and tense usage. Then, all answers are summarized in learning profile form before learners access to learning behavior collection as illustrated in Fig. 2.

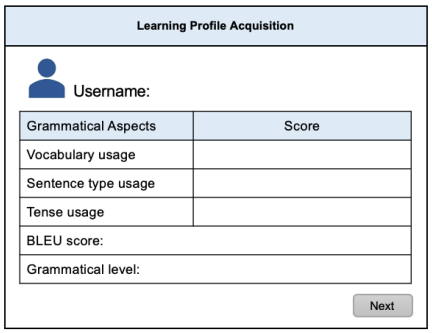


Fig. 2. An example form of learning profile in learning profile acquisition

*2) **Learning Behavior Collection**:* The learning behavior collection is central module that connects guideline components (linguistics and MT) and data log storage. The guideline components are set as a learning environment of this module. All learners behavior are collected into the data log store, then sending to analyze by learning analytics module. Moreover, this module is designed based on concept of three phases self-regulated model.

The self-regulated model is an encouraging model to interact between personal, behavioral, and environmental factors for effective learning. The main concept of self-regulated model consists of three phases: forethought phase, performance phase, and self-reflection phase [11]. The forethought phase is setting goal about learner need to learn. The performance phase is behavior recording that learners action with environment of the system and inform learners of their progress. The self-reflection phase is self-evaluation from behavior recorded and adapted behavior from the result of self-evaluated for increasing effectiveness method of learning.

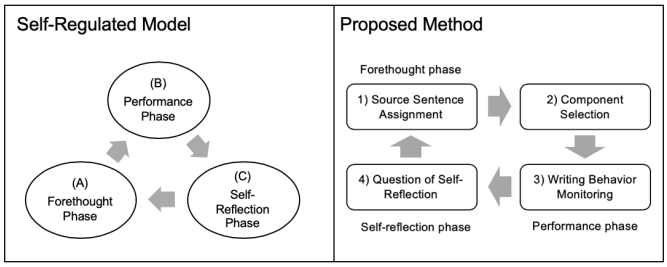


Fig. 3. A relation between parts of learning behavior collection and self-regulated model

The learning behavior collection includes four parts are source sentence assignment, components selection, writing behavior monitoring, and question of self-reflection. There are three parts of this module consistent with the concept of self-regulatory model i.e. source sentence assignment (related to forethought phase), writing behavior monitoring (related to performance phase), and question of self-reflection (related to self-reflection phase) as illustrated in Fig.3.

2.1) **Source sentence assignment:** The learners select the source sentence for trying to write the target sentence completely. Since learners assign source sentence by themselves that related to goal setting in the forethought phase. The selected sentence indicates the interest of

learner for how to write that sentence in the target language completely.

2.2) **Component selection:** When learner assigns source sentence, the learner could select provided components to guide for target sentence writing. The guideline components consist of components of linguistics and component of MT. The components are divided in to two levels: components for syntax level and lexical semantic level.

2.3) **Writing behavior monitoring:** The behavior monitoring part shows all activities of learner. When learner finished composing the target sentence in each round, the system will process all activities in the log file and show results for observation by learner. All activities are recorded from learner assigns source sentence, selects all components and submits the target sentence in each round. Since the learners observe or monitor their performance by themselves that related to self-observation in the performance phase in self-regulated model.

2.4) **Question of self-reflection:** Last step of learning behavior collection, the provided questions of self-reflection are assigned to ask learners for learners reflection. The answers of these provided questions are also recorded into log file. The questions will reflect the writing behavior in the previous round. The questions of self-reflection help to adapt learning behavior in next round. The relation of behavior between previous round and next round related to self-reaction in the self-reflection phase in self-regulated model.

*3) **Learning Analytics**:* The learning analytics module aims to analyze learning behavior in the log file from data logs stored. Learning analytics has become an important role that provides helpful information to optimize learning design, outcome, and environment based on analysis results [12]. Since data alone are insufficient to shape theory and guide practice, learning analytics is a useful process in the digital learning environment. In this work, learning analytics used to analyze writing behaviors for increasing self-regulation and writing proficiency. The module uses a statistical analysis method [13] to determine behavior transitions or find the learning behavior pattern. This analysis module aims to analyze learning behavior about the provided component in the learning environment that reflects the performance of English writing. Moreover, learning analytics uses to find learner's learning pattern who has improved for a suggestion to other learners.

B. Learning Environment

The learning environment composed of learning materials and linguistics/MT materials as illustrated in Fig. 4 The reason for using components of MT to guide English writing: widely language learning tool and characteristic of Thai EFL learners. For widely language learning tool, MT is an online tool that easy to access and help to write target sentence especially in EFL learning [14], [15], [16]. For characteristic of Thai EFL learners, Thai learners mostly think in Thai and translate. However, they do not understand the whole translation process. But machine translation involves understanding both languages in their translation processes. So, these components in MT could help learners to learn proper English writing.

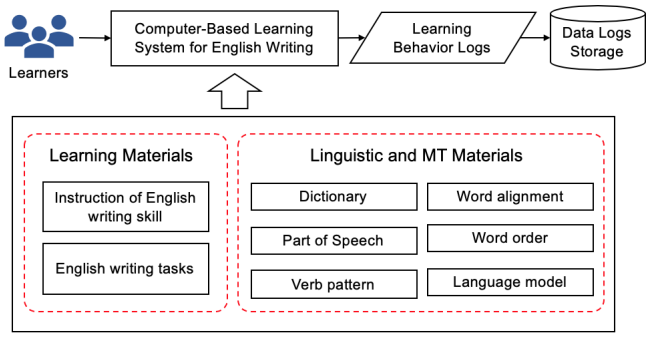


Fig. 4. The guideline components as a learning environment

The components are divided into two levels: syntax level and lexical semantic level. The components in syntax level help to write the target sentence in sentence structure and tense aspects. Furthermore, the components in lexical semantic level help to write the target sentence in vocabulary aspects. The components are separated into two levels as follows in Table I.

TABLE I
LIST OF GUIDELINE COMPONENTS AND THEIRS GRAMMATICAL ASPECTS

NLP Level	Grammatical Aspects	Components
Lexical Semantic	Vocabulary	Dictionary
		Word Alignment
Syntax	Sentence Structure	Word Order (Decoding)
		Language Model
	Tense	POS
		Verb Pattern

The computer-based learning system was developed and deployed for getting learning behavior. When learners use the system to practice English writing, the system collect behavior data like selection, composition, insertion, modification, substitution, and deletion. Those activities are stored to data log storage. Then, all learning behavior data were used to analyze learning behaviors.

- **Selection:** Learner can select interesting guideline components for helping sentence composition. They can change to select another component by clicking the component button and backtrack to the previous component by clicking the previously selected button.
- **Composition:** When learners want to type the whole target sentence, they can click composition button and textbox for sentence typing will be shown. After writing finishing, learner click OK button, the composed sentence will be save in log file.
- **Insertion:** When learners want to add some word or phrase in target sentence, they can click insertion button and cursor in textbox will be shown. Then, learners can save edited sentence by clicking the OK button.
- **Modification:** When learners want to delete some word or partial in the target sentence, they can click the

modification button. The words in the target sentence are represented in word buttons. They can select some word buttons for deleting.

- **Substitution:** When learners want to change the position of some word in the target sentence, they can click the substitution button. The words in the target sentence are shown in word buttons for moving position within the sentence.
- **Deletion:** When learners want to delete the whole target sentence, they can click deletion button and textbox will be cleared. Then, learners can compose new target sentence.

Moreover, The provided guideline components encourage these actions between learners and the learning environment. These behaviors are occurred during try to write target sentence that learners can select the provided components to guide for English writing. All activities are shown behavior transition of each learner.

C. *Participants*

The system used to collect learning behavior logs from Thai EFL learners. The undergraduate students are participated in this study. The participants are asked to write an English sentence via the computer-based learning system. All learning activities are collected into data logs store for sending to analyze in learning analytics phase. The participants information are protected by hiding their personal information by an authorized ethics committee in Thammasat University.

III. Learner Scenario

The learning activities of system usage consist of learning profile acquisition, source sentence assignment, component selection, target sentence submission, behavioral monitoring, and self-reflection respectively as illustrated in Fig. 5. All learning behaviors are recorded into log file. The log files are separated by a round of target sentence submission.

For explaining the system usage, the learner scenario is shown by a comparison between two learners. The learner starts with filling personal information, education information and answer of the provided question for learning profile acquisition. Next, the system provides a set of source sentence which depend on score of learning profile acquisition. Suppose that student A has a background English knowledge less than student B. The system provides sentence set in beginning level for source sentence assignment while provides sentence set in intermediate level for student B. After that, learners select guideline components to guide for target sentence composition.

Both learners have different learning profile and source sentence that affects to the component selection of each learner. Student A assigns a source sentence in beginning level and select three components i.e. dictionary, POS, and dictionary respectively. On the other hand, student B compose target sentence that related to source sentence in intermediate level by selecting two components i.e. verb pattern and word alignment respectively. When learners submit target sentence, the system generates the result of writing composition for learner monitoring. Finally, learners answer the provided self-reflection questions after observing own behavior from monitoring panel. The provided self-reflection questions use to evaluate the learner's self-satisfaction. The score of self-satisfaction and behavior transition are also summarized into the learner behavior form as illustrated in Fig. 6 and Fig. 7.

For sample case, student A has a few background English knowledge that affects to occur a problem of unknown vocabulary. From behavior transition, student A is not confident in vocabulary usage that observes from component about vocabulary is selected in many times as shown in behavior transition. Moreover, student A will take action to write after select guideline component. So, the guideline component can help the learner for writing. On the contrary, student B has more confidence in English writing that observe from selecting the Composition component in first time. The correction of the sentence is less than student A by comparing from Modification and Insertion components are selected as illustrated in Fig. 6 and Fig. 7.

IV. Discussion and Conclusion

This paper describes importance of system design. The system is designed to collect learning behavior of Thai EFL learners in English writing. The learning behavior data is necessary data for learning analytics. Since in Thailand does not have a system to collect learning behavior in English writing, the data from this system are useful for analysis and improving language learning of Thai EFL learners. The system is designed base on self-regulated model and integrated guideline components as a learning environment. This system aims to support self-regulated learning in English writing. The guideline components used to guide for English writing in syntax and lexical semantic level.

As self-regulated model mentioned, the strategies of this model are encouraging strategies for interactions of personal, behavioral, and environmental factors for effective learning. The self-regulated model is an effective process which learners establish goal, monitor self-behavior, and reflect own behavior. The behavior controlling guided by learners goal and the contextual features of the learning environment (M. Boekaerts, P. R. Pintrich, 2000). So, the system is designed based on strategies of the self-regulated model could support learner to regulate their behavior for efficiency of learning. Furthermore, This system is novel system that set the learning environment with component of linguistics and MT. since previous works [14], [15], [16] are used MT in language learning by using as a post-edit tool, but not extract components of MT for using into writing process.

In the future, the proposed system is planned to be developed and deployed for Thai EFL undergraduate learners. The system is set an assisting tool for self-regulated learning in English language learning. The provided guideline components are integrated as a learning environment to improve grammatical writing skill. Moreover, the deployed system is

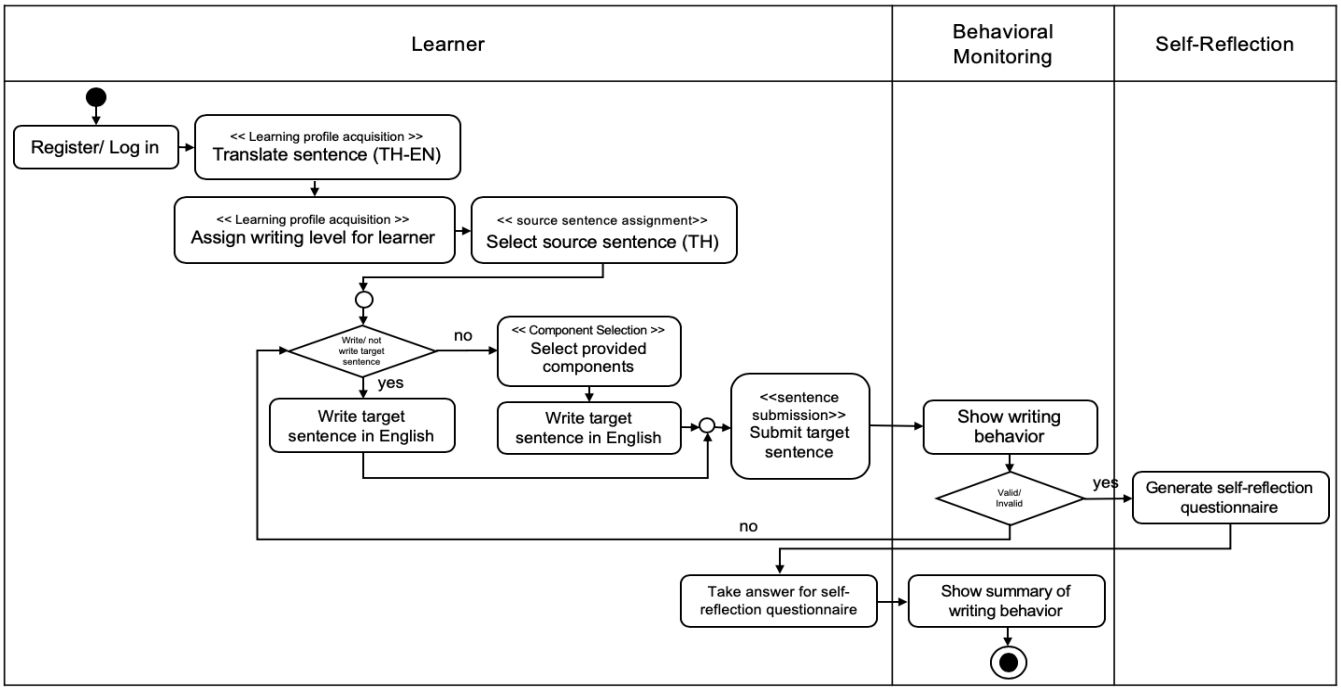


Fig. 5. An activity diagram of learner's activity

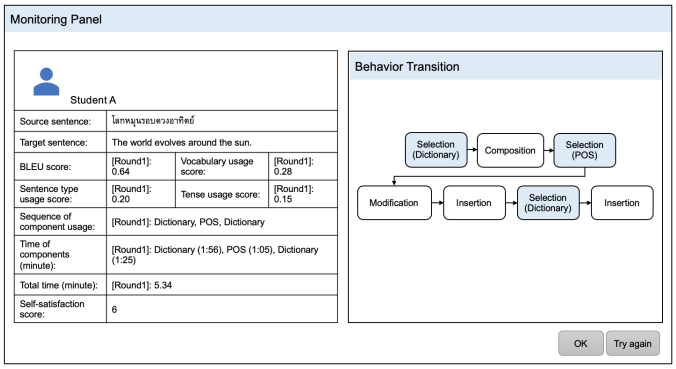


Fig. 6. An example summary of learning behavior and behavior transition (case: student A)

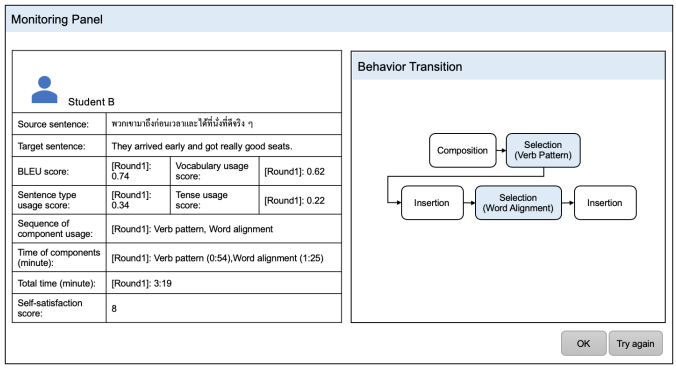


Fig. 7. An example summary of learning behavior and behavior transition (case: student B)

used to collect learning behavior and analyze for finding learning pattern in English writing of Thai EFL learners.

REFERENCES

[1] K. Sermsook, J. Liamnimitr, and R. Pochakorn, "An Analysis of Errors in Written English Sentences: A Case Study of Thai EFL Students," *English Language Teaching*, vol. 10, no. 3, p. 101, 2017. [Online]. Available: http://www.ccsenet.org/journal/index.php/elt/article/view/66264

[2] N. E. Perry and K. J. VandeKamp, "Creating classroom contexts that support young children's development of self-regulated learning," *International Journal of Educational Research*, vol. 33, no. 7, pp. 821 – 843, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0883035500000525

[3] M. S. Bagheri, S. Karami, M. J. Riasati, and F. Sadighi, "The Impact of Application of Electronic Portfolio on Undergraduate English Majors' Writing Proficiency and their Self-Regulated Learning," *International Journal of Instruction*, vol. 12, no. 1, pp. 1319–1334, 2019.

[4] M. Seker, "The use of self-regulation strategies by foreign language learners and its role in language achievement," *Language Teaching Research*, vol. 20, no. 5, pp. 600–618, 2016.

[5] B. J. Zimmerman, "Academic studing and the development of personal skill: A self-regulatory perspective," *Educational Psychologist*, vol. 33, no. 2-3, pp. 73–86, 1998. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00461520.1998.9653292

[6] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," *Metacognition in educational theory and practice. The educational psychology series.*, no. January 1998, pp. 277–304, 1998.

[7] L. S. Teng and L. J. Zhang, "Effects of motivational regulation strategies on writing performance: a mediation model of self-regulated learning of writing in English as a second/foreign language," *Metacognition and Learning*, vol. 13, no. 2, pp. 213–240, 2018.

[8] M. H. Dembo, L. G. Junge, and R. Lynch, "Becoming a Self-Regulated Learner: Implications for Web-Based Education." in *Web-based learning: Theory, research, and practice.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2006, pp. 185–202.

[9] J. M. Lodge, E. Panadero, J. Broadbent, and P. G. De Barba, "Supporting self-regulated learning with learning analytics," *Learning analytics in the classroom: translating learning analytics research for teachers*, no. October, pp. 45–55, 2019.

[10] C. Brockett, W. B. Dolan, and M. Gamon, "Correcting ESL errors using phrasal SMT techniques," *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06*, no. July, pp. 249–256, 2006. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1220175.1220207

[11] B. J. Zimmerman, "Self-Efficacy: An Essential Motive to Learn," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 82–91, 2000.

[12] W. Greller and H. Drachsler, "Translating learning into numbers: A generic framework for learning analytics," *Educational Technology and Society*, vol. 15, no. 3, pp. 42–57, 2012.

[13] H.-T. Hou, "Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (mmorpg)," *Comput. Educ.*, vol. 58, no. 4, pp. 1225–1233, May 2012. [Online]. Available: http://dx.doi.org/10.1016/j.compedu.2011.11.015

[14] I. Garcia and M. I. Pena, "Machine translation-assisted language learning: Writing for beginners," *Computer Assisted Language Learning*, vol. 24, no. 5, pp. 471–487, 2011.

[15] J. Clifford, L. Merschel, and J. Munné, "Surveying the Landscape: What is the Role of Machine Translation in Language Learning," *@Tic. Revista D'Innovació Educativa*, vol. 0, no. 10, pp. 108–121, 2013. [Online]. Available: http://ojs.uv.es/index.php/attic/article/view/2228

[16] N. Briggs, "Neural machine translation tools in the language learning classroom: Students ' use, perceptions, and analyses," *JALT CALL Journal*, vol. 14, no. 1, pp. 3–24, 2018. [Online]. Available: https://jcj.jaltcall.org/index.php?journal=JALTCALL&page=article&op=view&path%5B%5D=116

Predicting Drug Sale Quantity Using Machine Learning

Warayut Saena
Department of Computer Engineering
Faculty of Engineering, Mahidol University
25/25 Salaya, Phuttamonthon,
Nakhon Pathom, 73170 Thailand
warayut.sae@student.mahidol.ac.th

Vasin Suttichaya*
Department of Computer Engineering
Faculty of Engineering, Mahidol University
25/25 Salaya, Phuttamonthon,
Nakhon Pathom, 73170 Thailand
vasin.sut@mahidol.ac.th

***Abstract*—Medication is one of the essential parts of a patient's treatment. Therefore, it is important to have good medication storage administration in order to have effective medication storage. This study aimed to find a proper model used for the prediction of medication purchase amount by using machine learning to analyze medication purchasing amounts in the form of time series. In this research, the first 10 medicines in AV group were chosen. Then, Multilayer Perceptron (MLP), Long Shot-Term Memory (LSTM), and 1D Convolutional neural network with LSTM models were used together with Rolling Windows which were used to predict the purchase amount of each model. The periods of prediction were at 1 month, 3 months, and 6 months. The efficacy of each model was compared using their errors. CNN-LSTM model produces the better forecasting results. The result also shows that 1-month forecasting period is suitable for medicines that specific to disease. The 3-month forecasting period is suitable for commonly used medicines. The 6-month forecasting period is suitable for the medicines for chronic diseases.**

***Index Terms*—Time Series, Multilayer Perceptron, Convolutional Network, Long Short-Term Memory**

I. Introduction

Medicine costs are considered as one of the major expenses that contribute to the overall spending of hospitals. Hence, medical inventory must be properly managed. If the volume of medicine stored in a hospital exceeds the volume of medicine administered, such medicine may be expired before they are administered. Subsequently, this may lead to wasteful spending on purchases and storage of medicine. Moreover, some types of medicine require a long period of time to be purchased and transported [1]. As a result, medicine supply in hospital might not be able to satisfy patient's demand, which can cause in adverse and life-threatening effects on patients. Accordingly, medicine inventory management is one of the most important process of hospital due to its effects on hospital's finance and operations, specifically in the aspect of the fiscal budget for medical supplies. Poor management of medical inventory may cause financial crisis in hospitals [2].

This research proposed the incorporation of machine learning in time series analysis to forecast the amount of drug sale quantity that is appropriate for medical inventory management of the hospital. Machine learning will facilitate the hospital's decision-making process pertaining to drug purchases, which will result in efficient inventory management and reducing the storage cost of drug reserves, precluding a shortfall of inventory, and preventing drugs from being expired before they can be used. In addition, machine learning can be further used to formulate medical inventory management plans for other hospitals.

There are many studies that predict data trend by using machine learning. Janardhanan and Barrett [3] presented the time series forecasting of CPU usage of machines in data centers through the use of LSTM and ARIMA models. The time series dataset of CPU usage was extracted from Google's cluster data. According to the results, the LSTM model had a prediction error in the range of 17-23%, while the ARIMA model had the prediction error rate of 37-42%.

Selvin et al [4]. presented three different deep learning architectures for the prediction of the stock price. The researchers applied a sliding window approach for the short-term prediction of future values. According to the results, the deep learning models performed better than the ARIMA model. Moreover, the CNN model was found to give more accurate results than the RNN and LSTM models.

Hernandez et al [5]. presented a deep learning architecture for the prediction of accumulated daily precipitation for the next day. An Autoencoder was employed to reduce and capture non-linear relationships between attributes, and a Multilayer Perceptron was used for prediction. The results were compared with the following methods: the naive approach for the prediction of the accumulated rainfall of t-1 (Naive 1), the naive approach for the prediction of the average accumulated rainfall (Naive 2), the MLP approach, the Autoencoder (AE) and MLP approach, the Backpropagation Network (BP) approach, the Layer Recurrent Network (LR) approach, and the Cascaded Backpropagation (CBP) approach [6]. The best method was found to be the AE and MLP approach, which had the MSE and RMSE values of 40.11 and 6.33 respectively.

Kaneko and Yada [7] proposed a sales prediction model for retail stores by adopting the deep learning approach. The results was found that the deep learning approach had the highest prediction accuracy for all categories. The researchers then employed the L1 regularization for models that were constructed using deep learning. According to the results, deep

learning was able to improve the prediction accuracy by 1-3%.

Watanabe et al [8]. proposed a model for the prediction of goods demand in supermarkets and analyzed the effect of weather on the demand in order to forecast daily sales. The researchers examined two datasets, one of which was a collection of daily sales records obtained from Japanese retailers, and the other was the historical data of weather. The results obtained from the Linear Regression, Neural Network, and Deep Neural Network approaches were compared. In conclusion, the Deep Neural Network approach was the best method for improving the predictive performance of the model.

Kritchanchai and Meesamut [9] presented a method for analyzing of the most suitable inventory management policy. The drugs employed in the study were classified by using the ABC/VEN analysis, with ABC denoting the consumption value and VEN representing the clinical importance. Only drugs with a high consumption value (A class) were selected. The current inventory policy (Min/Max) and inventory policies from previous studies were compared to propose a new inventory policy that is appropriate for each drug category and characteristic. According to the results, the proposed inventory policy was able to reduce the number of shortages of drugs in the AV category by 92.98%, as well as reducing the total inventory cost of drugs in the AE and AN category by 14.63%.

II. Methodology

This research started by collecting the pharmaceutical sales data from Singburi hospital. The data was collected every month and stored in database. After that, the drugs were classified using the ABC/VEN analysis. Then, different machine learning models were applied to the selected medicines in AV group. Primarily, the effectiveness of such models for predicting drug demand was tested and evaluated. The methodological steps of this research are presented in Fig 1.

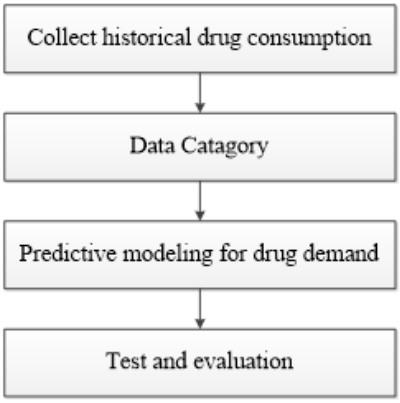


Fig. 1. Steps of work

A. Collect historical drug consumption

The pharmaceutical sales data was collected, which recorded 1,651 drug information from Singburi hospital. There are 72 records of each drug commencing from October 2009 to September 2015. This research selected sale quantity data from each month for prediction

TABLE I
ABC/VEN matrix

ABC/VEN	High Consumption Value (A)	Medium Consumption Value (B)	Low Consumption Value (C)
Vital (v)	AV	BV	CV
Essential (E)	AE	BE	CE
Non-essential (N)	AN	BN	CN

B. Drug Category

The ABC/VEN method was used to classify medicines. Table I represents the relation between ABC and VEN category. ABC is a drug classification based on the amount of drug usage and VEN classifies drugs according to their importance in treatment [10],[11],[12]. The medical sales quantity was collected from October 2009 to September 2015. The collected data is then sorted by drug sale quantity data and treatment importance. This research uses the data from AV group, which are about 70% of total drug usage in Singburi hospital. The list of considered medicines are as follows:

- Metformin Tab. 500 Mg.
- Enalapril Tab. 5 Mg.
- Enalapril Tab 20 Mg.
- Athenolol Tab. 50 Mg.
- Paracetamol Tab 500 Mg.
- Paracetamol Syr. 12 Mg/5 Ml.
- Omeprazole Cap. 20 Mg.
- Salbutamol Sulfate Inh. 0.1 Mg/Dose.
- Chlorpheniramine Syr. 2 Mg/5 Ml.
- Chlorpheniramine Tab. 4 Mg.

C. Rolling Windows

A Rolling Window is a technique used to extract features from time series data. It repeats regressions using subsamples of total data by shifting the start and end points with a fixed window [13],[14]. The stride of Rolling Windows shifting 1 month per step. For instance, predicting 1 month using Rolling Windows size 6. The data of each drug are 72 records, regressions are initially conducted using the time series of 1-6, followed by 2-7, 3-8, and finally 66-72. The example of fixed size window in the sequence data is presented in Fig 2. All models in this research compared between Rolling Windows size 1, 6, and 12.

D. Predictive modeling for drug sale quantity

The training data for 1-month prediction were selected from October 2009 to August 2015 and the remaining September 2015 were selected for testing. The training data for 3-months prediction were selected from October 2009 to June 2015 and the testing data were selected from July 2015 to September 2015. The training data for 6-months prediction were selected from October 2009 to March 2015 and the testing data were selected from April 2015 to September 2015.

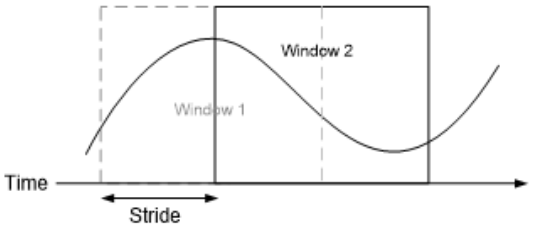


Fig. 2. Sliding the window

Dataset was converting to fit the proposed model, which performed by normalizing the data to obtain the range of 0 to 1. All models were trained with Adam optimizer. The number of epochs was set to 200 for MLP and LSTM. The number of epochs was set to 400 for CNN-LSTM model. These models are designed for drug sale quantity prediction.

1) Prediction by Multilayer Perceptron (MLP): MLP consists of 3 layers, which are input, hidden, and output layer. The size of input layer is varied, depended on the window size. ReLU is used as an activation function. The general overview of MLP model with window size 1, 6, and 12 are illustrated in Fig 3, 4, and 5 respectively.

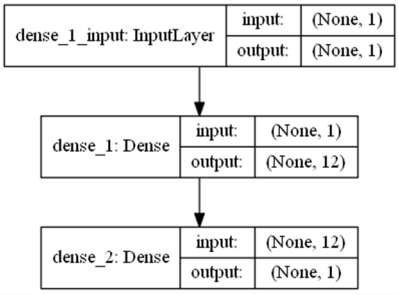


Fig. 3. MLP using Rolling Windows size 1

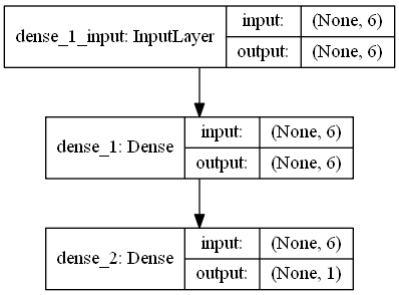


Fig. 4. MLP using Rolling Windows size 6

2) Prediction by Long Short-Term Memory (LSTM): LSTM using Rolling Windows size 1 consists of 3 layers, which are input layer, hidden layer, and output layer. LSTM using Rolling Windows size 6 and 12 consists of 4 layers, which are input layer, 2 hidden layers, and output layer. Tanh is used as an activation function. The general overview of LSTM model with window size 1, 6, and 12 are illustrated in Fig 6, 7, and 8 respectively.

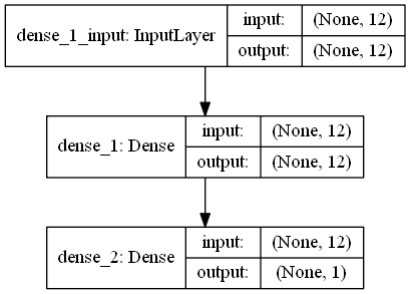


Fig. 5. MLP using Rolling Windows size 12

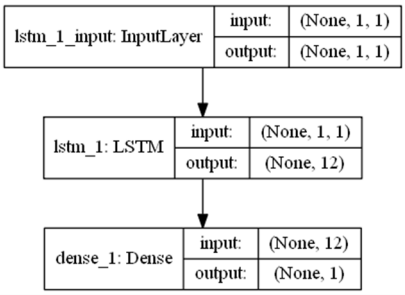


Fig. 6. LSTM using Rolling Windows size 1

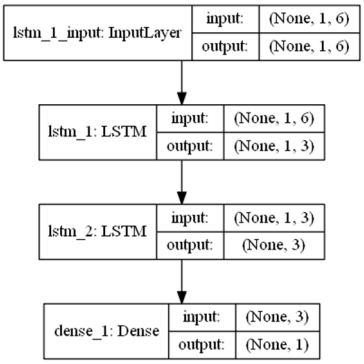


Fig. 7. LSTM using Rolling Windows size 6

3) Prediction by Convolution Neural Network and Long Short-Term Memory (CNN-LSTM): CNN-LSTM consists of 5 layers, which are input layer, 3 hidden layers, and output layer. The hidden layers consist of 1d Convolutional layer and 2 LSTM layers. The size of input layer is varied, depended on the window size. ReLU is used as an activation function for CNN layer and Tanh is used as an activation function for LSTM layer. The general overview of CNN-LSTM model with window size 1, 6, and 12 are illustrated in Fig 9, 10, and 11 respectively.

III. Test and Evaluation

Upon optimization of the proposed model, the results are compared with those obtained from other predictive models.

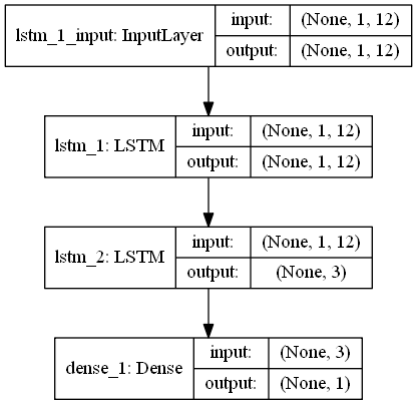


Fig. 8. LSTM using Rolling Windows size 12

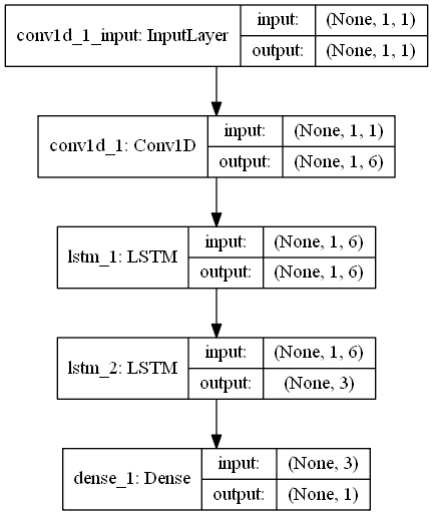


Fig. 9. CNN-LSTM using Rolling Windows size 1

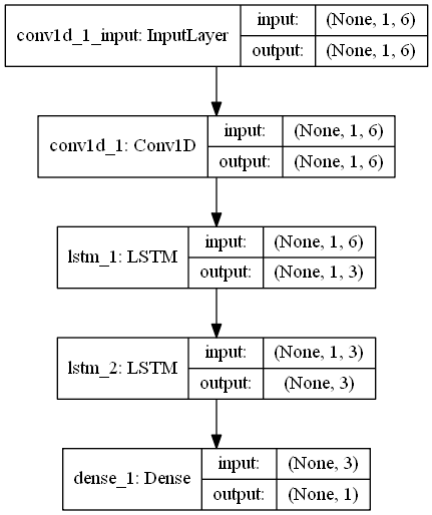


Fig. 10. CNN-LSTM using Rolling Windows size 6

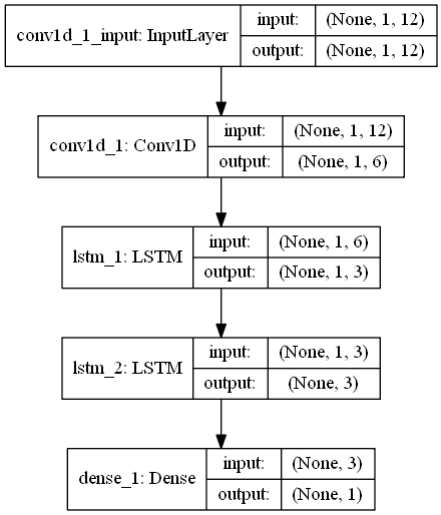


Fig. 11. CNN-LSTM using Rolling Windows size 12

Mean Square Error (MSE) is used as a loss function in order to optimize the models.

After obtaining the predicted output, the output was applied by denormalization and error was calculated using the relative error equation

$$error = \frac{|x^i_{real} - x^i_{predict}|}{x^i_{real}},$$

where x^i_{real} is the i^{th} real value and $x^i_{predict}$ is the i^{th} predicted value.

IV. Results and Discussions

We attempted to find the correlation between model with trend and variance of data, but none of correlation could be found. The experiment was done for three different machine learning models using Rolling Windows size 1, 6 and 12. Table II shows a comparison of the prediction method.

From Table II, CNN-LSTM can outperform other two models. The reason is CNN can automatically detect the features from spatial signal while LSTM receives high-level representation features from CNN for mapping to the output. However, the performance of CNN-LSTM depends on the initial parameter tuning. It requires a lot of experiments and computational power in order to fine tuning the parameters for CNN-LSTM.

The size of Rolling Window also affects to the performance of the proposed models since the window size relates to the structure of spatial data. A small size of window cannot capture the major pattern while a large size of window cannot extract minor detail that appears in the data. The result shows that the window of size 12, which represent the drug sale quantity in the period of 12 months, produces low relative error in many models. The reason is the sale quantity of medicines is directly influenced by disease seasonal factors.

The 1-month forecasting period is suitable for Omeprazole Cap 20 mg and Salbutamol Sulfate INH.0.1 mg/dose. Both

TABLE II
COMPARISON OF THE PREDICTION METHOD

Drug Name	Model	Windows Size	Error(%)		
			1 Month	3 Month	6 Month
METFORMIN TAB. 500 MG.	MLP	1	13.89	3.39	13.56
		6	**11.70**	14.63	0.78
		12	36.14	16.51	5.67
	LSTM	1	15.71	3.73	13.66
		6	38.54	7.03	2.57
		12	19.96	30.83	6.28
	CNN + LSTM	1	15.49	**3.24**	13.31
		6	37.45	6.66	**0.71**
		12	22.07	21.43	9.04
ENALAPRIL TAB. 5 MG.	MLP	1	24.78	3.55	12.23
		6	15.60	14.89	6.87
		12	**7.97**	15.57	**1.90**
	LSTM	1	27.20	5.20	14.44
		6	37.13	9.26	15.96
		12	21.13	2.07	3.38
	CNN + LSTM	1	24.33	3.87	13.73
		6	27.86	10.57	14.57
		12	19.93	**1.15**	7.70
ATENOLOL TAB. 50 MG.	MLP	1	114.95	113.05	88.18
		6	63.85	64.79	46.48
		12	31.80	79.02	25.40
	LSTM	1	114.03	114.04	91.13
		6	46.44	58.03	56.17
		12	18.30	41.16	14.17
	CNN + LSTM	1	111.95	112.69	82.63
		6	54.77	27.74	72.54
		12	**11.40**	**18.98**	**1.15**
ENALAPRIL TAB. 20 MG.	MLP	1	61.67	12.68	0.86
		6	68.22	24.20	1.43
		12	80.68	40.36	11.74
	LSTM	1	**56.16**	**11.73**	**0.31**
		6	60.32	24.20	7.95
		12	76.14	31.39	10.36
	CNN + LSTM	1	58.82	12.32	1.14
		6	69.35	26.88	7.30
		12	65.67	27.44	8.68
PARACETAMOL TAB. 500 MG.	MLP	1	8.85	17.41	21.49
		6	30.23	31.34	20.33
		12	16.48	3.49	40.83
	LSTM	1	6.89	18.43	25.79
		6	32.21	17.74	22.80
		12	10.93	21.31	24.46
	CNN + LSTM	1	7.58	17.92	32.48
		6	4.56	**2.46**	**14.32**
		12	**0.10**	9.29	17.53
PARACETAMOL SYR. 120MG/5ML.	MLP	1	345.03	31.01	25.24
		6	308.99	25.81	28.88
		12	**120.99**	45.96	**9.68**
	LSTM	1	343.49	31.55	22.65
		6	324.51	39.17	20.34
		12	305.90	37.57	18.27
	CNN + LSTM	1	342.51	30.55	22.13
		6	314.97	36.24	18.31
		12	200.28	**8.41**	14.50
OMEPRAZOLE CAP. 20 MG.	MLP	1	5.62	9.44	10.15
		6	2.55	**3.35**	**5.12**
		12	2.91	21.81	26.59
	LSTM	1	5.67	10.95	9.53
		6	5.20	6.79	5.18
		12	9.11	17.58	20.19
	CNN + LSTM	1	5.03	9.95	7.89
		6	4.77	4.97	7.93
		12	**1.47**	11.76	11.71
SALBUTAMOL SULFATE INH. 0.1 MG/DOSE	MLP	1	0.90	20.52	35.01
		6	9.41	22.47	24.66
		12	3.90	25.34	**3.17**
	LSTM	1	0.39	25.17	38.26
		6	11.83	18.56	29.50
		12	17.61	6.42	21.98
	CNN + LSTM	1	**0.09**	24.78	37.56
		6	17.20	**5.92**	22.67
		12	5.14	12.18	18.66
CHLORPHENIRAMINE SYR. 2 MG/5ML.	MLP	1	26.63	5.74	13.96
		6	63.34	5.44	21.37
		12	30.43	13.85	2.76
	LSTM	1	28.10	3.25	15.30
		6	57.53	8.65	18.91
		12	44.31	**0.55**	9.52
	CNN + LSTM	1	27.88	4.68	12.55
		6	73.07	8.18	21.36
		12	**5.61**	0.78	**2.16**
CHLORPHENIRAMINE TAB. 4 MG.	MLP	1	9.68	8.05	22.66
		6	14.46	3.70	23.56
		12	25.57	10.29	92.32
	LSTM	1	9.56	7.32	26.02
		6	21.92	14.95	30.11
		12	**9.25**	5.22	37.28
	CNN + LSTM	1	11.45	6.95	24.40
		6	33.15	**1.67**	29.13
		12	13.04	20.43	**20.21**

medicines can be categorized as the medicine for specific disease since Omeprazole is used to reduce the amount of acid in stomach and Salbutamol Sulfate is used to treat breathing problem. The 3-month forecasting period is suitable for Paracetamol Syr 120mg/5ml, Chlorpheniramine Syr 2 mg/5ml, and Chlorpheniramine Tab 4 mg. Both Paracetamol and Chlorpheniramine are nonprescription drugs that commonly used in Singburi province. The 6-month forecasting period is suitable for Metformin Tab 500 mg, Atenolol Tab 50 mg. and Enalapril Tab 20 mg, which can be categorized as the medicine for chronic diseases.

Paracetamol Tab 500 mg. is expected to have the best results in 3-month forecasting period. However, the result shows that the 1-month forecast period is suitable for this medicine. The result also shows that 3-month forecasting period is suitable for Enalapril Tab 5 mg, which is the medicine for chronic diseases. However, the relative error is not that much different. The reason is the dataset contains a highly dynamical value which can cause some deviation in the analysis process.

There is a paper [9] that closely relates to our research but the objective and the methods are different. In [9], the authors attempt to derive the inventory policy for each medication category in order to reduce the total inventory cost. In contrast, our proposed methods attempt to predicting the demand of medical usage by analyzing the sale quantity.

V. CONCLUSION

This research applies machine learning methods for predicting pharmaceutical sale quantity. The data was collected from Singburi hospital. We select the sale quantity of ten medicines in AV category, which cover 70% of total drug usage in hospital, for evaluating the proposed models. The empirical analysis shows that the changes occurring in drug sale quantity data not always be in a regular pattern or not always follow the same cycle. Moreover, the data also contains a highly dynamical value since the sale quantity can be affected by many factors, such as disease seasoning patterns and pharmacist's experiences.

MLP, LSTM, and CNN-LSTM are applied to analyze the data. The result shows that CNN-LSTM produces the better result. The reason is 1D CNN is developed for automatically

learn features. Thus, it can understand patterns occur in the current window. Moreover, LSTM is capable of learning long term dependencies in the time series. The prediction of Salbutamol Sulfate INH.0.1 mg/dose using CNN-LSTM with Rolling Windows size 1 obtained the best prediction result, which displayed the least error 0.09%.

The result also shows that 1-month forecasting period is suitable for medicines that specific to disease. The 3-month forecasting period is suitable for commonly used medicines. The 6-month forecasting period is suitable for the medicines for chronic diseases.

REFERENCES

[1] D. Kritchanchai, A Framework for Healthcare Supply Chain Improvement in Thailand, Operation and Supply Chain Management, Vol.5, No.2, pp.103-113, 2012.

[2] D. Kritchanchai, R. Suwandechochai, "Supply chain management in health sector in Thailand: a case study", International Journal of Services, Economics and Management, Vol.2, No.2, pp.211-224, 2010.

[3] D. Janardhanan and E. Barrett, "CPU workload forecasting of machines in data centers using LSTM recurrent neural networks and ARIMA models," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, 2017, pp. 55-60.

[4] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1643-1647.

[5] E. Hernández, V. Sanchez-Anguix, V. Julian, J. Palanca, N. Duque, "Rainfall Prediction: A Deep Learning Approach" Hybrid Artificial Intelligent Systems. HAIS 2016, pp 151-162

[6] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar. "A rainfall prediction model using artificial neural network". In 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC), July 2012.

[7] Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales," 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, 2016, pp. 531-537.

[8] T. Watanabe, H. Muroi, M. Naruke, K. Yono, G. Kobayashi and M. Yamasaki, "Prediction of regional goods demand incorporating the effect of weather," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 3785-3791.

[9] D. Kritchanchai and W. Meesamut, "Developing Inventory Management in Hospital ," International Journal of Supply Chain Management (IJSCM), vol. 4, 2015

[10] A. A. Koshkarov, A. A. Khalafyan, E. U. Dolzhkova and A. B. Semenov, "Automation of planning of medical-economic drug prescription control," 2016 IEEE Conference on Quality Management, Transport and Information Security, Information Technologies (IT&MQ&IS), Nalchik, 2016, pp. 103-107.

[11] Vaz, F.S., Ferreira, A.M., Kulkarni, M.S., Motghare, D.D, "A study of drug expenditure at a tertiary care hospital: an ABC-VED analysis", Journal of Health Management, Vol.10, No.1, pp.119–127, 2008.

[12] Vaz, Frederick. "Application of ABC-VED analysis in the medical stores of a tertiary care hospital," International Journal of Pharmacology and Toxicology, 2014

[13] S. Aparna, "Long Short Term Memory and Rolling Window Technique for Modeling Power Demand Prediction," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 1675-1678.

[14] [14] L. B. Amor, I. Lahyani and M. Jmaiel, "Recursive and Rolling Windows for Medical Time Series Forecasting: A Comparative Study," 2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), Paris, 2016, pp. 106-113.

Analyzing behavior in nursing training toward grasping trainee's situation remotely

Yuki Kodera
Informatics Course
Kochi University of Technology
Kochi, Japan
235061h@gs.kochi-tech.ac.jp

Miwa Saito
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
saitomiw@kochi-u.ac.jp

Sumika Yoshimura
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
yoshimuras@kochi-u.ac.jp

Kyoko Yamawaki
Department of Nursing, School of Medicine
Kochi University
Kochi, Japan
y-kyou@kochi-u.ac.jp

Kunimasa Yagi
University Hospital
University of Toyama
Toyama, Japan
yagikuni@med.u-toyama.ac.jp

Mikifumi Shikida
School of Information
Kochi University of Technology
Kochi, Japan
shikida.mikifumi@kochi-tech.ac.jp

***Abstract*—In nursing education, it is important to improve nursing ability, clinical training is conducted as part of this. It is conducted in parallel by many trainees. Therefore, it is very difficult for a teaching advisor to grasp many situations at the same time. This research goes toward realizing a remote situation grasping system in order to support teaching advisor to notice trainees who need teaching. In this paper, we analyzed differences of behavior between trainees based on nursing behavior data measured using sensors. Furthermore, we discussed the possibility of use for the supporting system based on the results.**

***Index Terms*—Analysing behavior, Clinical training, Nursing training, Multiple sensors, Grasping situation remotely**

I. Introduction

In recent years, Japanese healthcare has changed significantly. However, there are also many issues. For example, there are correspondence of new medical technologies and medical safety. Therefore, it is necessary to improve the clinical ability of nursing staff in order to solve that issues [1]. The basic nursing education carries out field training, which is aimed at applying knowledge and skills to actual situations. Nevertheless, there is actually a big gap between the practical ability immediately after graduation and the level of nursing practice required in clinical training [2].

In field training, a teaching advisor has many trainees. In addition, each trainee train in each patient room in parallel. Therefore, the teaching advisor needs to go to check each room at one's discretion. However, it is hardly that the checking and nursing behavior moment are match. Hence, the teaching advisors might overlook an opportunity in need teaching. In this research, we aim to realize an efficient teaching environment, and to reduce burden on the teaching advisor. This research supports in the future that the teaching advisor grasps the many trainee's behavior remotely. Therefore, this research requires to investigate what statistical information is effective for display to the teaching advisor in a remote place.

We implemented a method for collecting behavior data of trainees during nursing training. It collects the nursing behavior data by using sensors such as acceleration and pressure. We attached multiple sensors to each part of the trainee's body such as head, arm, wrist, waist, and foot. Then, during nursing training, we observed the movement of nurse's body. In addition, we analyzed differences of action between the nursing behavior of trainees based on the collected data.

Next, we describe related work in chapter 2. After that, we describe the implemented collection system in chapter 3, about experiment in chapter 4, and the results in chapter 5. Finally, we describe the differences of the nursing behavior revealed by the collected data in chapter 6. Furthermore, we refer to the appropriateness of collection for the devices used in the implemented system, the possibility of use for the grasping situation remotely.

II. Related work

A. Support for clinical training in the medical field

We have supported clinical training in the medical department. For instance, we have displayed statistical information to a teaching doctor and trainees who played a doctor as an attempt to analyze a medical behavior using sensors. The information is rate of the trainee who played the doctor turned their face to a patient and maximum time they did not talk during medical interview training [3]–[5].

In addition, we have proposed supporting method as well. This method enables trainees to study themselves between their group in medical interview training by using devices for inputting results of evaluation of query for the trainee who played the doctor. Furthermore, it assists a feedback learning for the trainee by playing video of only necessary scenes using the evaluation results. Moreover, it supports the teaching doctor to grasp the situation of parallel medical interview training remotely [5]–[7].

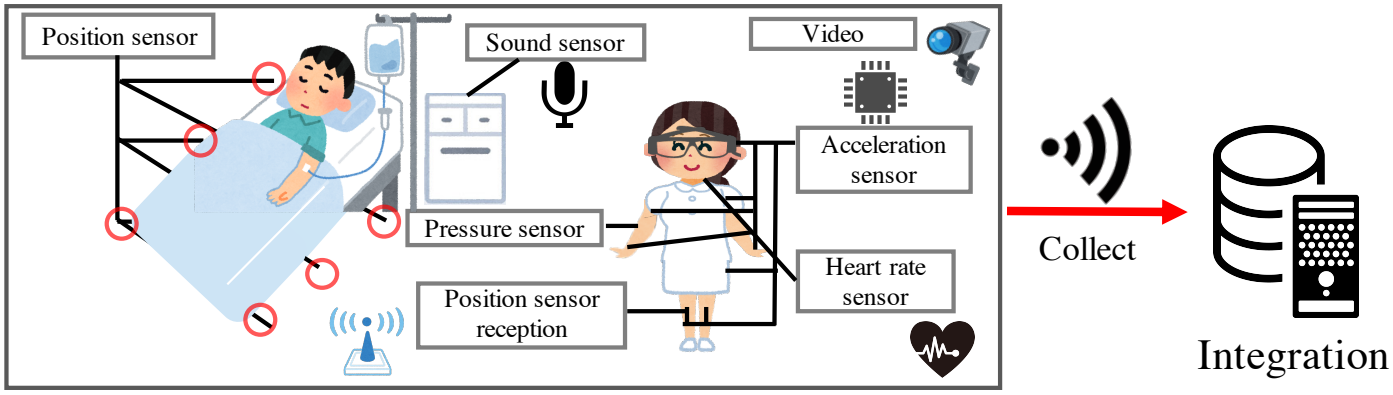


Fig. 1. Overview of a method for collecting data.

B. Analysing characteristics in nursing behavior

The article [8] analyzes characteristics of the nursing behavior of expert nurses and trainees who has not nursing skill during moving assistance to wheelchair. As above, it can discover an index of trainees who need teaching by analyzing the characteristics between skilled and unskilled person during the nursing behavior. In addition, it can grasp situations of nursing training of trainees by displaying statistical information used these indexes to the advisor in a remote place.

III. Implemented System

A. A method for Collecting data

This research needs to calculate statistical information based on multiple parts of the body. Hence, this collecting data method collects behavior data from each part of trainee's body during nursing training by using multiple sensors and a video images. Fig.1 shows an overview of the method. Multiple sensors measures accelerations of head, upper arm, wrist, waist, and foot, and trainee's voice, pressure on their right arm, heart rate, and their position from bed as behavior parameters. In addition, the measured data are collected in one place in order to be able to compute many kinds of statistical information. Moreover, the data are corrected so that there is no time difference between these the data. Besides, the situation of the nursing behavior is recorded by using the video.

B. Implemented devices

Fig.1 shows each point where attach to trainee's body each an implemented devices, and Installation position. Small smartphones (Jelly Pro) is attached to the upper arm, wrist, waist and foot, and the accelerations are sampled at 50Hz. Furthermore a wearable device(JINS MEME of JINS) is equipped the head, and the acceleration is sampled at 40Hz. In addition, a small device (M5stack) connected to a pressure sensor (ALPHA ELECTRONIC pressure sensor (round shape, large)) is attached to the right arm, as well as the small device connected to a heart rate sensor(Pulse Sensor) is attached to the neck. Moreover, BLE tags(Sanwa Supply MM-BTIB1) are Installed at 6 positions under the patient's bed, and the distances between the small smartphone attached to the left foot and each position is sampled at 1Hz. A sound sensor is also Installed near the bed. Fig.2 shows an image example of the devices wearing and installing.

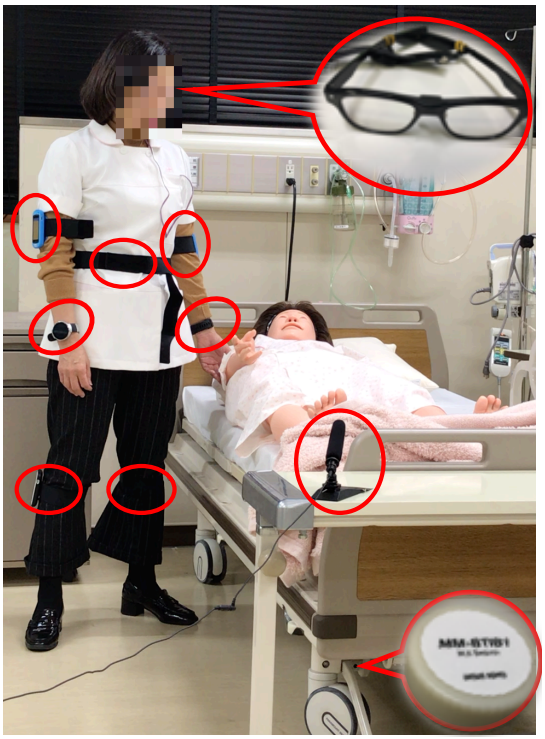


Fig. 2. Attachment position for implemented devices.

IV. Experiment

We collected the data of a postural change for a patient by nurse which is one of the nursing behavior in February 2019. Furthermore, examinees are 8 nursing trainees who are 3 years of Kochi University. In addition, they have experienced a lecture of the nursing behavior and never trained not. They took the postural change for doll similar to the patient which

is lying on one's back on bed in exercise room of Kochi University. The patient's posture change was performed by them in the following order the supine posture, the long sitting posture, sitting posture, standing posture, and moving assistance to wheelchair. Besides, we regard this flow as a session.

V. Results

We collected 18 sessions of the data. Fig.3 shows the graph of the accelerations of the head, upper arm, wrist, waist, and foot from the collected data.

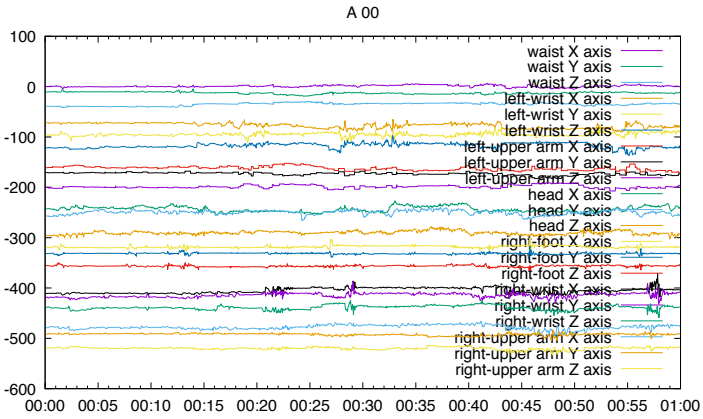


Fig. 3. Accelerations of the data collected from each part of the trainee.

tried to lift the patient several times by observing the video. Moreover, we also calculated the angles of right upper arm of others sessions in the same way. In addition, we calculated the number of times of peak of the right arm movement based on each obtained the angle in 18 sessions. TABLE I shows the results. As a result, we confirmed that 5 out of 8 trainees of the second time less than the first time about the counts of up and down movements of their right upper arm while they were performing the action of putting their wrist around the patient's back.

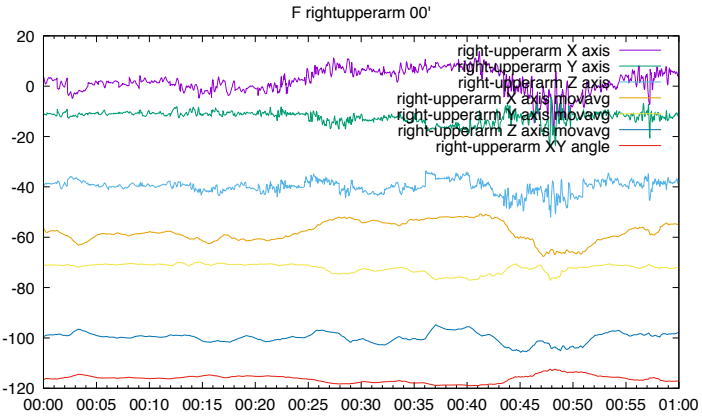


Fig. 4. Accelerations of right upper arm in examinee F.

VI. Discussion

A. Differences of behavior in collected data

We analyzed 18 sessions. We especially compared the differences of collected data between the examinees F and H. As the results, we perceived the difference whether or not they touch the patient's body during a speech at the beginning of the session. The examinee F did, however H touched patient's body after stood still and finished speaking.

Next, Fig.4 and 5 show graphs of the acceleration of the right upper arm during the sessions of examinees F and H. The acceleration of examinee F has dense between each step such as the session start and beginning of the posture change from the sitting posture to the standing posture. In contrast, the other has non-dense. Additionally, we confirmed that examinee H was performing trial and error between each step while they were performing by observing the video of the session H. We consider that this trial and error caused non-dense between each step. In particular, we perceived that examinee H was looking for a position to put their wrist on while they were performing the action of putting their wrist around the patient's back from the sitting posture to the standing posture. Furthermore, we calculated each moving average of interval 40 in the accelerations of X and Y of their right upper arm of the session F and H, and analyzed angles of their right upper arm based on it. As a result, we perceived a count which right upper arm of examinee F moved up and down is 1 time by observing the angle of Fig.4, on the other wrist, the other count is 4 times. In addition, we confirmed that actually examinee H

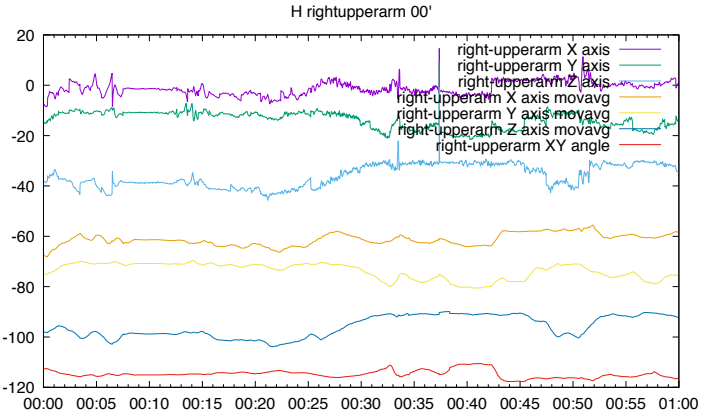


Fig. 5. Accelerations of right upper arm in examinee H.

TABLE I
Number of times of up and down of trainee's right upper arm.

	A	B	C	D	E	F	G	H
1st	4	10	14	11	6	18	2	44
2nd	2	9	17	17	5	3	14	8
3rd	16	8						

B. A collection aptitude of the devices

We perceived the difference of collection aptitude of the devices from collected the data. First, we consider difficult

that analyzing the behavior characteristic from the data similar to section VI.A due to several reasons. For example, if we analyze movements of wrist, we have to consider various factors such as the degree of freedom of range of joints. However, Analysing an upper arm is simpler than analysing a wrist because we do not need to consider a elbow joint.

In addition, the values of collected pressure sensor had hardly change. This result shows the possibility that the point where we expected and the point where the patient actually touched are different, and the attachment point may depend on the person. In other words, we should consider the position that a sensor can touch a point where the patient's body if we use the sensor which depends on whether it touches the point such as the pressure sensor.

Conversely, we have to analyze not only specific step but also a characteristic of whole the session in order to detect beginning each step. We think that it can detect the moments when the hand of the trainee put on the patient's arm by calculating the position of the them for using a device that can capture the whole such as a video. Additionally, we are set to analyze using OpenPose, a software that estimates a posture using deep learning. Fig.6 shows an example of the result of analyzing the coordinates of joints in the image by inputting the test video of the posture change to Open Pose. Each line in the figure shows joints of human body. However, it is not possible to measure a degree as opposed to a pressure sensor. Besides, There are also problems such as an analysis is difficult depending on an angle for taking video. As above, we conclude that it has different collection aptitude each device, and the characteristic of this system is collecting multidimensional data and using it appropriately.

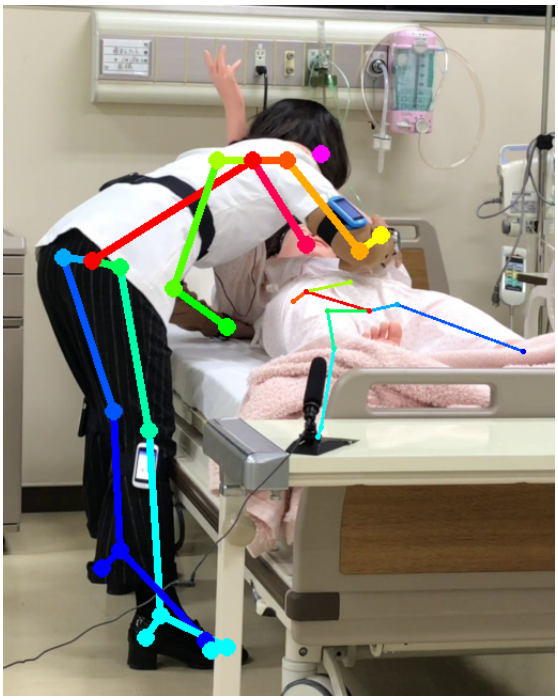


Fig. 6. An example of analyzed coordinates of joints during posture change.

C. A possibility of useful for supporting of remote situation grasp

As described above in section VI.A, we could discover the differences of behavior during the nursing behavior between trainees from the collected data. Furthermore, we also referred to the difference of the data that can obtain by using different devices in section VI.B. Therefore, we think that it is possible to reveal more characteristics which of the difference of behavior by integrating and analyzing each data. Moreover, we consider that it can be used for supporting remote situation grasp by detecting the characteristics of nursing behavior automatically and displaying these statistical information to the teaching advisor.

VII. Conclusions

In this paper, we described the data collection from multiple parts of the trainee's body during nursing training using the sensors and the video, and analyzing the differences of the data between trainees.

In recent years, Japanese healthcare has changed significantly, it has been required to improve the clinical ability. The basic nursing education carries out field training as part of that. In field training, there is the issue that teaching advisors might overlook an opportunity in need teaching as nursing training perform parallel.

This research supports the issue by displaying statistical information in nursing training to the teacher in a remote place. We attached multiple sensors to the body of the trainee of the nursing department, and collected the behavior data during nursing training. Furthermore, we analyzed the collected data, and described the differences in behavior between the trainees. Moreover, we referred to the necessary that integrate various devices and collecting multidimensional data because each device has the different collection aptitude.

In the future, we are set to evaluate whether the difference of the nursing behavior between the trainees becomes an index of statistical information display to the teaching advisor in a remote place.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 18K10204.

References

[1] Ministry of Health, Labour and Welfare. Consideration report related to improvement of clinical practice ability of new nursing staff, 2014.

[2] Sanae Oriyama and Aki Okamoto. The status of experiences of nursing skills in clinical practice among nursing students and factors influencing the subjective attainment level: An analysis of several nursing education institutions in the chugoku region. *Japan Academy of Nursing Science*, 35:127–135, 2015.

[3] Hajime Kira, Mikifumi Shikida, and Kunimasa Yagi. A study of analysis for the medical interview training based on speech attitude for development the automatic analysis. In *proc. of The 20th Spring Convention of Japan Association for Medical Informatics*, pages 122–123, 2016.

[4] Shunya Inoue, Mikifumi Shikida, and Kunimasa Yagi. A method to support behavior leaning using smart glasses in medical interview training. In *proc. of Multimedia, Distributed, Cooperative, and Mobile Symposium*, pages 1790–1794. Information Processing Society of Japan, 2018.

[5] Mikifumi Shikida, Yuki Kodera, Shunya Inoue, and Kunimasa Yagi. A method for supporting medical-interview training using smart devices. In *proc. of the 13th International Conference on Knowledge, Information and Creativity Support Systems*, pages 80–85, 2018.

[6] Yuki Kodera, Mikifumi Shikida, and Kunimasa Yagi. A trial of evaluation support system of inquiry in group of medical interview training The Institute of Electronics. *IEICE technical report, The Institute of Electronics, Information and Communication Engineers*, 117(389):31–35, 2018.

[7] Yuki Kodera, Mikifumi Shikida, and Kunimasa Yagi. Efficient review with communication records of medical interview training. In *proc. of 22nd Spring Symposium of Japan Association for Medical Informatics*, pages 130–131, 2018.

[8] Yuko Mito, Sooja Kim, Mikiko Take, Hiromi Jono, Yasuko Shijiki, Masahiro Fukushi, Kenji Iwasaki, and Hiroshi Saitou. An analysis by nursing students and nurses of techniques for transferring patients from wheelchair to bed, 3rd report. *Tokyo Academy of Health Science*, 1(1):21–27, 1998.

Using Conceptual Graph to Represent Semantic Relation of Thai Facebook Posts in Marketing

1st Kwanrutai Saelim
Department of Computer Science,
Faculty of Science and Technology, Thammasat University,
Pathumthanee, Thailand
kwanrutai.saclim@gmail.com

2nd Taweewat Luangwiriya
Department of Computer Science,
Faculty of Science and Technology, Thammasat University,
Pathumthanee, Thailand
taweewatlu@gmail.com

3nd Rachada Kongkachandra
Department of Computer Science,
Faculty of Science and Technology, Thammasat University,
Pathumthanee, Thailand
rdk@cs.tu.ac.th

***Abstract*—Conceptual Graph is a representation commonly used to express semantic relationship of natural language. This work presents a method to translate Thai natural language text to conceptual graphs regarding semantic relations based on semantic roles between predicate and its arguments. Shallowing parsing of Thai text and verb patterns as case frames are utilised in identifying core entities in a context and their semantic roles. Then, the argument with annotated roles are translated into conceptual graphs that are able to logically and visually represent relations of core terms. As a result, conceptual graphs of Thai natural texts from Facebook posts in a marketing group were generated. In the study, found issues regarding Thai specific natural style are encountered and discussed.**

***Index Terms*—Semantic Relation, Conceptual Graph, Information Extraction, Semantic Role**

I. Introduction

A semantic role (also known as thematic role or theta role) is the studying of underlying relationship of entities involved with the main predicate in a clause [1] . Semantic roles are an attempt to capture similarities and differences in verb meaning with generalizations that contribute to the mapping from semantics to syntax. Analysing semantic role is a basic step to understand semantic of context and gives comprehensive power to a computer system. Semantic role labelling (SRL, shallow semantic parsing) [2] is the computational process to analyse and assign labels to words or phrases in a clause for indicating their semantic role. It is a challenging task studied for decades. The clause (or sentence) level semantic analysis of text is concerned with the characterisation of events including ”who” does ”what” to ”whom” with ”whom” at ”where” and ”when” [3]. The main task of SRL is to identify semantic relations from a predicate and its associated participants (textual entities). The SRL are labels to let know about action each clause. The SRL have labels indicate each words in the sentence such as Agent, Patient, Location of other entities and Temporal of event each clause.

Shallowing Parsing is an analysis of a sentence which analyze using components of sentence structure. Therefore, Shallow parsing will be done after Part-of-speech tagging(POS). Part-of-speech tagging is the process of marking up a word in a text in the sentence(nouns, verbs, adjectives, adverbs, etc.) [4]. The POS procedure we get results is that smaller units. There are called chunks. Shallow parsing combines units into larger units and that have grammatical meanings. So we get phrases, phrases can reflect the relationship between the meanings of the basic components. And when the Shallowing Parsing process is completed, the results will be next step is to the Semantic Role Labeling process as follow Figure 1. When receiving the SRL result the final step is to create with Conceptual Graph as follow Figure 2.

A conceptual graph (CG) is a graph representation the meaning which reply knowledge of logic based on the semantic networks. Which a conceptual graph one of artificial intelligence. The research of conceptual graph have explored

Identify applicable funding agency here. If none, delete this.

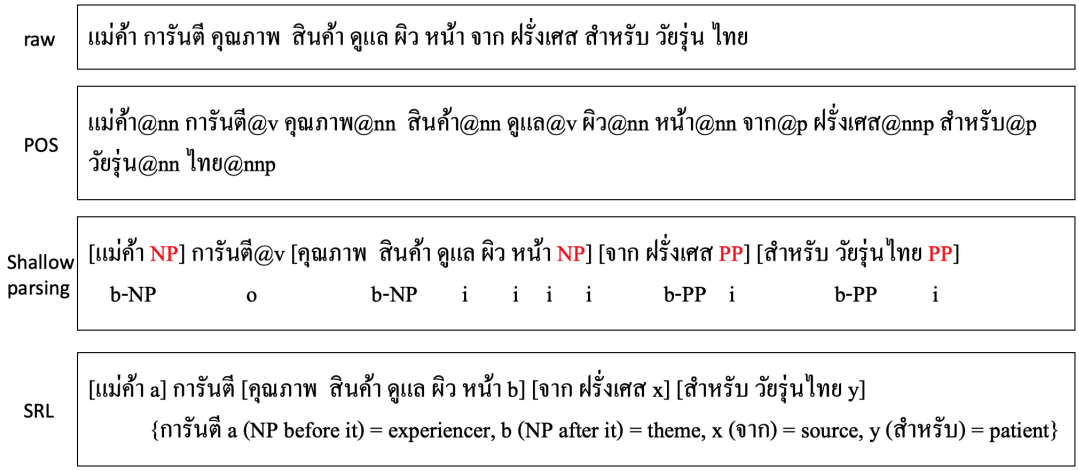


Figure 1. All step to get semantic labelling.

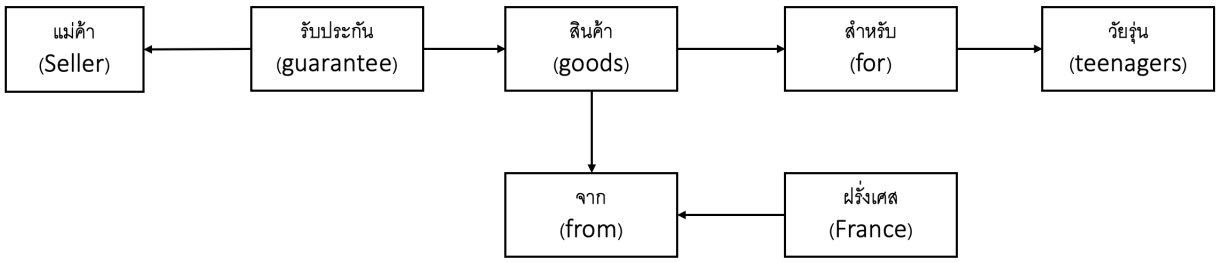


Figure 2. Examples of relations and a conceptual graph from marketing posts

the techniques for reasoning, knowledge representation from natural language into understanding natural language semantics. The semantics of the important base of the conceptual graph and conceptual graph is defined by a concept mapping to and from the ISO standard for Common Logic [5]. Mathematic is the important role of knowledge one in artificial intelligence (AI), Knowledge representation, their formal languages were not sufficiently expressive. Therefore the possibility of automated reasoning from the perspective of knowledge representation they were various in the use of syntactic constructs. knowledge representation to explain representing information in a form that a computer system can utilize to solve complex tasks. Knowledge representation and findings from logic to automate various kinds of reasoning the application of rules or the relations of sets and subsets.

In the past, there are several representations for labelling semantic roles, such as ”case frame” [1] [6] , FrameNet [7] and Conceptual Graph (CG) [8] . The CG is the most convenient on visualization and comprehensible while the FrameNet is designed to collect rich linguistic details. The most challenging point in SRL is that there is no specified set of roles that generalise for all contextual events. Moreover, ambiguity from natural language including subject-object omission and multi-meaning terms could render the sentence to non-analysable. For Thai language, very few studied and reported on semantic role labelled corpus [9] and Thai FrameNet as a resource for semantic analysis [10] . Their works were specified for a target domain such as agricultural context and tourist context in written Thai documents. Differently, the language used in social network, a major digital text resource nowadays, is more natural and vigorously stylish. Analysing of the semantic relation of words in social media is thus tough but necessary step towards understanding of Thai natural language.

In this work, we aim to create a framework for using conceptual graph to represent Thai social network posts. This work attempts to provide a standard framework to analyse and label Thai natural text with semantic role sets. The syntactic parser is applied to identify a predicate and its related entities.

II. Background

A. Semantic Role Labelling

There are many researches on manually created lexical and semantic resources as a lexical resource for natural language processing. Those works are guaranteed for their accuracy with the draw-back of labour-intensive task and restricted specific domains. For Thai, Suktarachan et al. presented a construction of Thai concept frames applied in language processing for agricultural domain based on the verb centric approach resulting in 5,784 Thai sentences annotated with POS and semantic roles. This is a good resource for semantic analysing for Thai text but apparently specified for agriculture documents. Their case frames and semantic roles however are adoptable as a standard for semantic role labelling task.

Semantic role labelling (SRL) is a task to identify the latent predicate argument structure of a clause/sentence, providing representations that answer basic questions about sentence meaning, such as ”who” does ”what” to ”whom”. General roles used in SRL are labels such as Agent, Patient, and Location for the entities participating in an event with temporal and manner details. These labels therefore provide a ground-level of semantic relation representation of the text. There are commonly used semantic roles [11] [12] as follows Figure 3.

- Agent: The ”doer” or instigator of the action denoted by the predicate.
- Patient: The ”undergoer” of the action or event denoted by the predicate.
- Theme: The entity that is moved by the action or event denoted by the predicate.
- Experiencer: The living entity that experiences the action or event denoted by the predicate.
- Goal: The location or entity in the direction of which something moves.
- Benefactive: The entity that benefits from the action or event denoted by the predicate.
- Source: The location or entity from which something moves
- Instrument: The medium by which the action or event denoted by the predicate is carried out.
- Locative: The specification of the place where the action or event denoted by the predicate in situated.

B. Conceptual Graph

A conceptual graph (CG) is a graph representation for logic based on the semantic networks [2]. The CG has been not only used to represent natural language semantics, but also knowledge representation and reasoning. In 1976, Sowa [2] proposed to use conceptual graphs (CGs) as an intermediate language for transforming natural language texts to machine-readable graph form. For example, CG of the sentence ”Kids went to Bangkok by train” is illustrated in Figure 3 In the CG, the rectangles refer to concepts representing entities of the text while the circles called conceptual relations are used to denote relations between concepts. Conceptual relations usually are based on the semantic role relation (see Section 2.1). An arc pointing toward a circle signifies the first argument of the relation, and an arc pointing away from a circle signifies the last argument.

CGs normally keeps the core information with concrete meaning of the text for simplification and ability of easy-to-comprehension; thus, not all the entities in the input text are kept. There are several researches on how to make use of CGs such as Question-answering, Diagrammatic reasoning, Entity-relationship model, Semantic web, etc.

III. Methods

We design a three-step approach for generating a conceptual graph representation of Thai texts. Same to other Thai natural processing tasks, pre-processing is required to handle Thai term-boundary and remove non-text entities such as emoticons and symbols. First, we identify the syntactic roles in a sentence using shallow syntactic parser. Second, we design a set of syntactic rules to semantic roles. Third, we construct a conceptual graph following the assigned semantic roles. An overview architecture of the system is drawn in Figure 4

In this work, the focused natural Thai texts are the Facebook posts about marketing. The posts related to product-selling advertisement and good details are collected from a marketing group. The target posts are text-based explanation excluding images and digital emotion expressions, i.e. emoticons and stickers. For pre-processing step, word segmentation is necessary to mark words ”boundary”. The preferable boundary approach of segmentation is a concept level since the entities should be understandable in themselves and do not need to combine for completing a word sense.

1) Firstly, for each clause in the sentence, we identify the main verb and build a sentence pattern using the parsed tree.

2) Secondly, for each verb in the sentence we extract a list of possible semantic frames from VerbNet, together with restrictions for each semantic role.

3) Thirdly, we match the sentence pattern to each of the available semantic frames, considering the semantic role’s constraints. As a result, we are presented with a list of all

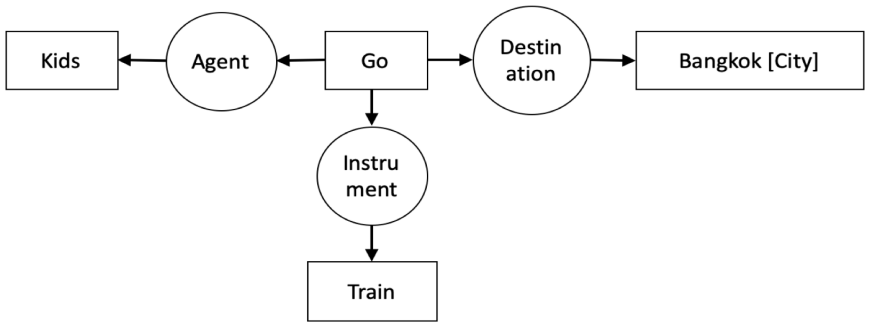


Figure 3. Conceptual graph representing a sentence "Kids went to Bangkok by train"

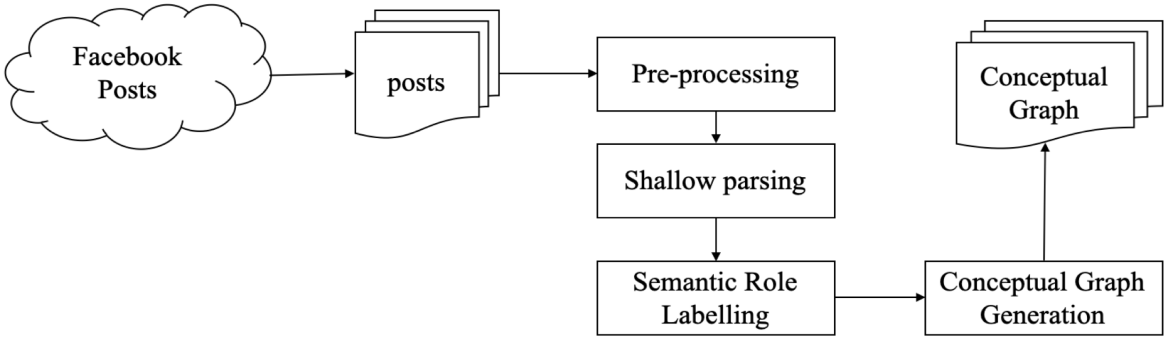


Figure 4. An overview architecture of Thai text based conceptual graph generation.

possible semantic role, s assignments, from which we have to identify the correct one.

A. Shallow Parsing for Detecting Clausal Core Entities

In SRL, the main entity to connect other entities is a predicate (typically a main verb of a clause) while the other entities such as a subject of the predicate and a direct object is handled as arguments of the predicate with roles. Thus, detection of these entities is undoubtedly essential for later processes. In fact, the other entities that represent a little to none meaning in the text, such as interjections and politeness ending markers, are ignored in this process to reduce complexity. A shallow parser is thus applied to separate an input text into phrases including Noun phrase (NP), Prepositional phrase (PP) and Verb phrase (VP). The process is conducted from left-to-right manner following a paradigm of Thai word composition. Unlike a normal syntactic parser, this shallow parsing only handles the VP for consecutive verbs (Serial Verbs) that has no NP and PP among them. In a case that NP belongs to a preposition, they will be grouped as PP while the NP with PP attached is counted separately as two distinct entities. The demonstration of the parsing results is given in Figure 5

A result of this process is phrasal chunks of Thai text. Please be noted that although shallow parsing in general has a fine accuracy result, shallow parsing for Thai is still a difficult and complex task from a nature of Thai in which is ambiguous and semantically implicit. Hence, the output chunks are required for manual post-editing to be reliably usable.

B. Semantic Role Identification

The algorithm for semantic role identification of a sentence that we propose consists of the following three steps:

1) Firstly, for each clause in the sentence, we identify the main VP and build a sentence pattern using the heuristic rules.

2) Secondly, for each verb in the sentence, we extract a list of possible semantic frames from VerbNet, together with restrictions for each semantic role.

3) Thirdly, we match the pattern to each of the available semantic frames, considering the semantic role constraints. As a result, we are presented with a list of all possible semantic

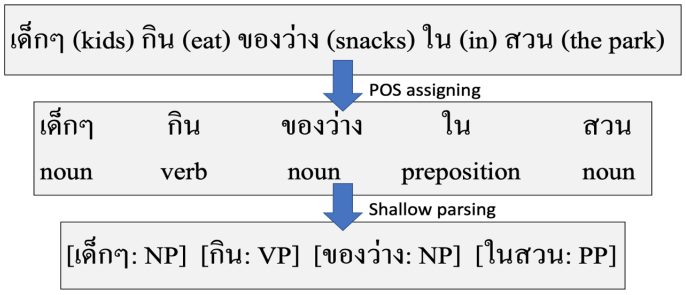


Figure 5. A demonstration of Thai shallow parsing

Verb	Preceding NP	Following NP	Related PP
"ซื้อ" (by)	Agent	Patient	"จาก"(from) – source
			"ที่" (at) –,locative
"ส่ง" (send)	Agent	Patient	"จาก" (from) – source
			"ถึง" (to) – destination
			"ที่" (at) –,locative
			"ด้วย" "โดย" (with) – instrument
"เห็น" (see)	Experiencer	Theme	"ด้วย, โดย" (with) – instrument
			"จาก" (from) – source

role assignments, from which we have to identify the correct one.

The rules for building pattern used in the step 1 mainly considers the verbs in a context. According to Thai serial verb construction, auxiliary verbs including temporal auxiliary (tense marker such as "กำลัง" (progressive) and "แล้ว" (past)), probability auxiliary (such as "น่าจะ" (should) and "อาจจะ" (may)), and directional auxiliary (such as "ขึ้น" (upward) and "ไป" (outward)) are ignored unless there are no other verbs. The remaining verbs with semantic meaning are treated equally. For the VerbNet, we collect Thai verbal information along with restriction to identify semantic relationship to other entities. We exemplify the information as shown in Table 1. For the current information, we have only collected verbs from the gathered corpus, but we plan to expand the database from other sources.

C. Conceptual Graph Generation

With annotated semantic roles, a conceptual graph can be generated accordingly. The CG as a whole is kept in a formula using logical expression. For an example from the context in Equation 1, the sentence can be formulated into the following formula.

$$(\exists x)(\exists y)(go(x) \wedge kids) \wedge City(Bangkok) \wedge train(y) \wedge$$
$$Agent(x, kids) \wedge Destination(x, Bangkok) \wedge Instrument(x, y))$$

As the exemplified formulation, the logical operators are conjunction and the existential quantifier. Those two operators are the most common in translations from natural languages. The formulae thus can be linked to form more detailed relationship as shown in Figure 6

With the conceptual graph and its logical reasoning, the obtained information can be utilised to classify sellers together based on their predicates such as type of products, locations, and marketing manners. Furthermore, this information could be used as knowledge base for further AI based application and data analytics.

IV. Discussions

In generating CGs from Thai Facebook posts on marketing, we encountered several issues. We found that the natural Thai textual expressions were difficult to handle perfectly. Firstly, 27 % of the inputs lack their subject part in which leads to incomplete graph. This circumstance in fact is a usual style of Thai natural language that typically causes ambiguity and incorrectness in automated processing. In this work, we handled the issue by analysing the verb and decide to add the subject part manually. There were two common solutions as 1) adding the post owner name as the subject of the predicate, and 2) considering NP previously mentioned in prior sentence as a subject of the predicate. We found that about 85 % of the cases were handled with the former solution. Secondly, many emerging terms or jargons were used with specific meaning such as "บ่องตง" (transformation of "'บอกตรง" (frankly speak)) and "ตะมุตะมิ" (stylish euphemism referring to "being cute"). These terms were used sparingly in many contexts in which requires specific design of patterns. Thirdly, the confusion

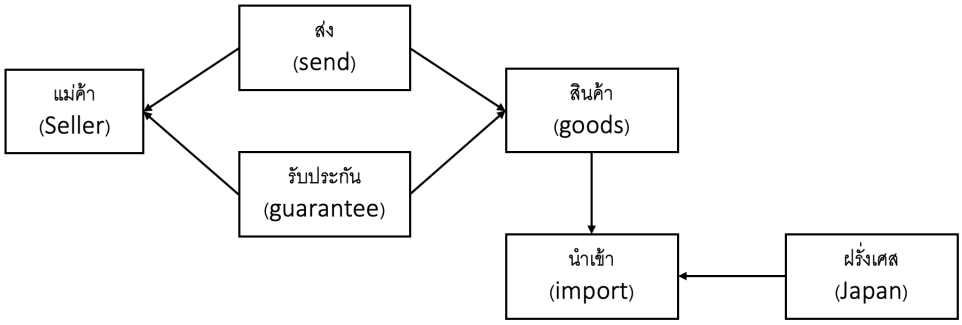


Figure 6. Examples of relations and concepts from marketing posts

between Thai serial verb construction and running causes. Thai language does not have an explicit marker for cause boundary. With the omission of subject part and continuous written text, verbs can be placed adjacently and lead to ambiguity in both shallow parsing and identifying semantic roles. For instance, Thai text as "ลูกค้า" (customer: NP) "สนใจ" (interesting: verb) "ชิ้น ไหน" (which product: NP) "อินบอค" (inbox: verb) "เลย" (mood marker: ignore in shallow parsing)" should be analysed for two main verbs sharing the subject "customer". However, the pattern can be confused that the second verb might take the object of the first argument as its own subject. This issue though can be solved if there is a dictionary to help declaring a concept of the noun as it is an object and cannot be considered as an Agent for the predicate.

V. Conclusions

This work presents a method to translate Thai modern natural texts to conceptual graph via the use of semantic role identification. The proposed method requires the shallow parsing to detect the predicate of the clause/sentence and verb information to identify semantic roles of predicate arguments. With the boundless style of Thai natural language, the design of processes is considerate of incomplete sentence and being flexible. In an attempt to generate conceptual graph, several issues leading to failure were found and discussed. The generated conceptual graphs can be used to classify posts according to predicates and their arguments such as type of products, locations, and marketing manners. Furthermore, this information could be used as knowledge base for further AI based application and data analytics.

References

[1] S. Padó and M. Lapata, "Cross-lingual annotation projection for semantic roles," *Journal of Artificial Intelligence Research*, vol. 36, pp. 307–340, 2009.

[2] J. F. Sowa, "Cognitive architectures for conceptual structures," in *International Conference on Conceptual Structures*. Springer, 2011, pp. 35–49.

[3] T. Cohn and P. Blunsom, "Semantic role labelling with tree conditional random fields," in *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2005, pp. 169–172.

[4] G. Neumann and J. Piskorski, "A shallow text processing core engine," *Computational Intelligence*, vol. 18, no. 3, pp. 451–476, 2002.

[5] J. Sowa, "Chapter 5 conceptual graphs," *Foundations of Artificial Intelligence*, vol. 3, 12 2008.

[6] C. J. Fillmore, "The case for case." 1967.

[7] D. Leijen, W. Schulte, and S. Burckhardt, "The design of a task parallel library," vol. 44, 10 2009, pp. 227–242.

[8] S. Hensman, "Construction of conceptual graph representation of texts," 01 2004.

[9] Z. Lin, Y. Duan, Y. Zhao, W. Sun, and X. Wan, "Semantic role labeling for learner chinese: the importance of syntactic parsing and l2-l1 parallel data," *arXiv preprint arXiv:1808.09409*, 2018.

[10] D. Leenoi, S. Jumpathong, and T. Supnithi, "Building thai framenet through a combination approach," in *2010 International Conference on Asian Language Processing*, Dec 2010, pp. 277–280.

[11] C. J. Fillmore and C. F. Baker, "Frame semantics for text understanding," in *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, 2001.

[12] C. J. Fillmore and C. Baker, "A frames approach to semantic analysis," in *The Oxford handbook of linguistic analysis*, 2010.

Object Distance Estimation with Machine Learning Algorithms for Stereo Vision

Pawarit Akepitaktam
Department of Computer Engineering
Image, Information and Intelligence Laboratory
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
ake.pawarit@gmail.com

Narit Hnoohom
Department of Computer Engineering
Image, Information and Intelligence Laboratory
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
narit.hno@mahidol.edu

***Abstract*—This paper presents a novel distance estimation to calculate distances from the stereo camera to the object accurately. This study collected stereo camera images as a dataset, each object determined at two different lighting environments and five different distances between the stereo camera and the object. To estimate the distance, researchers applied supervised learning methods to approach this task. There were performed with two machine learning algorithms: Linear Regression, and Artificial Neuron Network Regression. In the experimental results, the efficiency of the proposed method was examined by using the evaluation metrics to calculate the distance estimation errors. The results showed the model of convolutional neuron networks operated with densely connected neuron networks has the lowest errors rate in comparison with other models. The model eliminates the error rate of distance estimation at 0.000531, 0.014490, and 0.000048 meters, measured by mean square error, mean absolute error and mean logarithmic error respectively.**

Keywords—stereo camera, distance estimation, deep learning, regression

I. INTRODUCTION

Depth map data provides a useful reference point in the field of image processing and computer vision. Many research studies had already demonstrated the utility of depth map data, not only in object distance estimation but also in object detection. Object distance estimation in computer vision is most widely practiced by a stereo camera, where images from the stereo camera are used to estimate the object distances seen by cameras. Stereo vision is a typical view of a creature that possesses two eyes system. One of the advantages of the binocular creature is by taking beneficial of overlapped visions to percept depth from both eyes along with the brain reception. Moreover, in the intelligence animal, it is capable of applying experiences to sense the distance of an environment. The stereo vision is most similar human binocular vision, two eyes located parallel in front of the head which different from a most animal that located separately on each side of the head. The stereo vision was used in various applications such as an autonomous vehicle, three-dimensional image reconstruction, background blurring in portrait mode on dual-camera smartphone and estimate the distance from the camera to objects.

In addition, object detection is one of the methods in the image segmentation process, which intends to distinguish the objects from the surrounding environment in the image. The typical task of assigning a class label to each object within the image and the output is a bounding box of objects. Nonetheless, object detection might come along with an object classification, and this classification method performs to categorize detected objects into classes. As the rapid development of machine learning algorithms, the object detection algorithm had been developed for monocular camera images. Felzenszwalb et al. [1] proposed an object detection scheme based on a combination of a multiscale deformable multi-scale part model by using latent SVM. These models are trained using a discriminative process requiring only bounding boxes in a collection of images for the objects. The results showed that this work is efficient and accurate. Recently, deep learning has overcome the constraints of traditional hand-designed feature extractions and can also achieve objectives in complicated circumstances of angular rotation, occlusion, and lighting environments. Several techniques have been proposed for the depth map prediction using deep learning algorithms [2] [3].

The rest of this paper is arranged as follows. The related works are introduced in Section II. The process of the proposed method is briefly explained in Section III. Section IV describes the experimental method and experimental results, and the conclusion is presented in Section V. Finally, Section VI provides a discussion on the results of the experiment.

II. RELATED WORKS

Over the last few years, the depth estimation of deep learning approaches using monocular and binocular camera images has been intensively exploited. Deep learning networks have been validated to be more productive than the traditional hand-designed feature extractions. The most related works will introduce in this section. Zhang et al. [2] proposed a model of binocular stereo vision by using a region-based convolutional neural networks (R-CNN) algorithm. This work used the R-CNN to identify and locate obstacles in the image, and then replace the object in the model in order to compare the area coordinates of the obstacle object. Finally, the formula was employed for calculating the distance between the binocular camera and the obstacle. The experimental results showed that the model of this work is achieved position and target object identification of obstacles effectively. Although this work was successful, there were still several issues that need to be improved, such as, the limitation in terms of the only single obstacle was detected at a period of time.

Another study was conducted by Liao et al. [3] on the monocular depth estimation task, which proposed a method for distance estimation. The data from a laser range finder was used to generate a dense reference map for redefining the

depth estimation task as an estimation of the distance between the actual and the reference depth. This work used a combination of classification and regression for improving the accuracy of the distance estimation. Two datasets were used the NYUD2 for the indoor dataset, and the KITTI for the outdoor dataset. The accuracy rate and the error rate of the proposed method show better results compared with state-of-the-art techniques that used the same dataset.

Alternatively, Mustafah et al. [4] presented a stereo vision system to estimate an object distance and measure the size of an object. The purposed measurement algorithm of this work designed to work in real-time by using two cameras. The results showed that the average time of the distance estimation and size measurements were 65 milliseconds per cycle. However, this work has several limitations regarding the lighting environment. Due to the background subtraction method was used in the object detection process; it might use high effort to re-calculate every small lighting change to receive the result in the uncontrollable light environments.

In this study, the interesting object was detected within the stereo images, and the extracting attributes were investigated. The objective was to develop a predictive model for an estimate the object distance by using stereo camera images through the two machine learning algorithms. The results obtained from this study provided the predictive distance estimation model of stereo camera images.

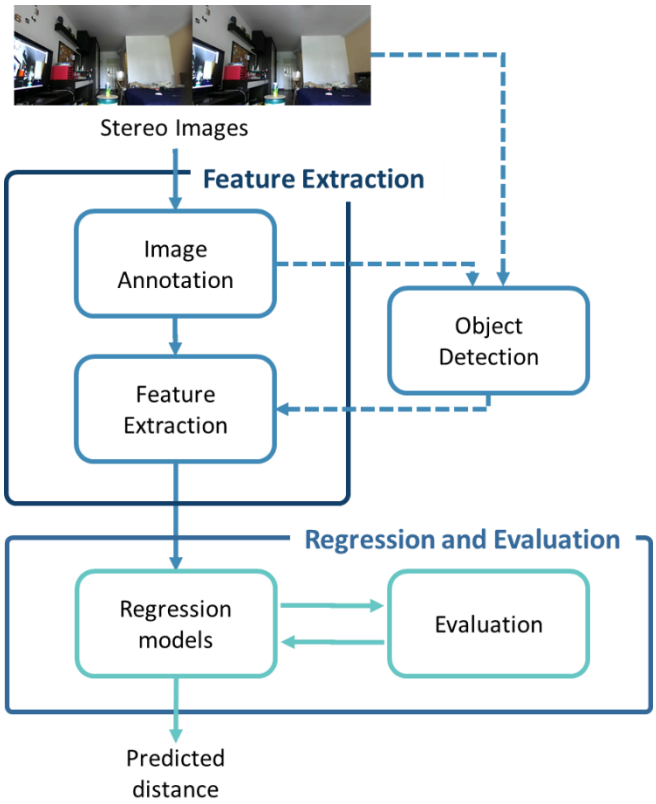


Fig. 1. Block diagram of proposed method. Feature extraction is provided dataset to training on our proposed regression models that estimate a distance.

III. Proposed Method

This research had the objective to predict the distance from the camera to the object in an image, by using a pair of images. Accordingly, as illustrated in Fig. 1, the structure of the proposed method was developed to achieve the research objective. The structure comprised of four processes: *Object Detection*, *Feature Extraction*, *Regression models*, and *Evaluation*.

A. Input images

The image dataset was constructed from 1,290 stereo images, that possessed 2,560 pixels in the width and 720 pixels in the height. Stereo images were taken by the stereo camera—the two lenses camera with the same focal length and image sensor size, that be able to take images from both lenses simultaneously; however, stereo images have an overlapped viewpoint, because two lenses share a rectilinear horizontal plane, that causes stereo images to perceive the depth even it is projected onto a two-dimensional surface.

For the diversity of the dataset, while the image dataset was being recorded, the environments were arranged in terms of numerous conditions as follows: light settings, object types, and object distances. Object types were comprised of two types: aluminum cans, and ceramic mugs. For the light settings, there was either the natural light source or the warm light bulbs.

The most considerable characteristic of the image dataset was the object distances, that were measured from the front of

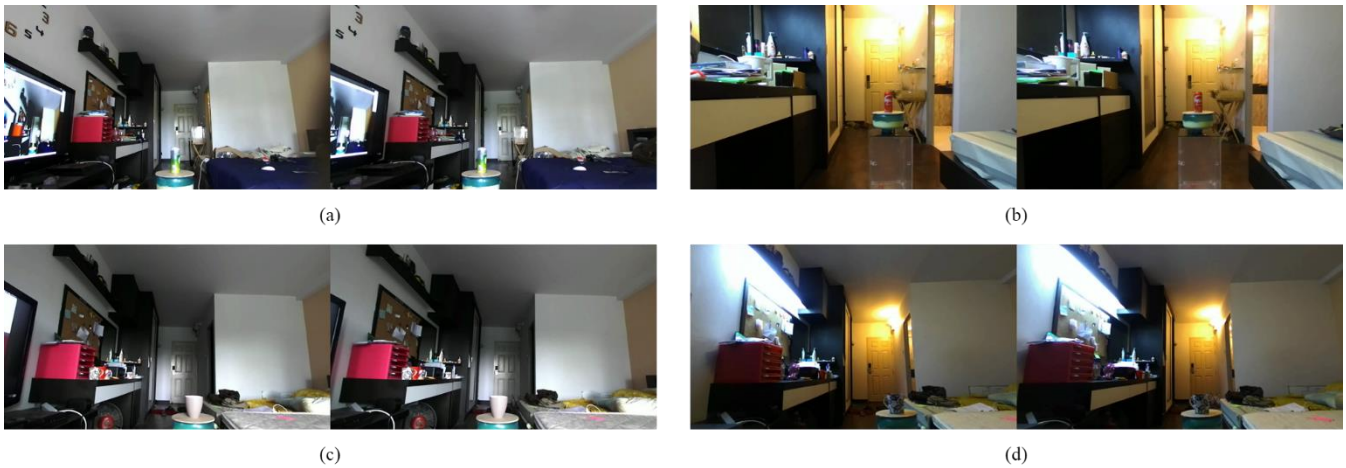


Fig. 2. Example of images from the image dataset. Stereo image is concatenated from left and right image, and interesting objects were in the center of each image. The interesting object in figure (a) and (b) is an aluminum can, and figure (c) and (d) is a ceramic mug. Figure (a) and (c) demonstrate the source of light that come from natural source, and figure (b) and (d) demonstrate light blub are the source of light.

TABLE I. EXAMPLE OF THE PRE-PROCESSED DATASET

Dis-tance	Left object position[a]				Right object position[a]			
	x TL	*y TL*	*x BR*	*y BR*	*x TL*	*y TL*	*x BR*	*y BR*
1.0	684	557	686	657	1851	555	1889	656
1.5	657	547	686	617	1881	545	1910	618
2.0	656	539	680	599	1891	539	1914	598
2.5	666	536	686	584	1909	536	1929	584
3.0	666	530	682	574	1914	531	1931	573

[a.] TL: The Top-Left position of the bounding box; BR: The Bottom-Right position of the bounding box

the stereo camera to the closest point of the interesting object. The image dataset was categorized into five groups: 1, 1.5, 2, 2.5, and 3 meters; consequently, the group of each distance had 258 images as shown in Fig. 2.

B. *Feature extraction*

The attention of this paper is concentrated on the distance estimation of an interesting object; consequently, the various other processes are described in this section. The basic concept of this process is the pre-processing of the image dataset, which contained raw images, in order to create a feature-linked dataset.

1) Image Annotations: the purpose of this process was to determine the position of the object in the image dataset by manpower with label image tool [5]. The area of the object was determined by constructing the most fitted rectangle around the object in the form of bounding box and annotate those object's areas according to their object types. Bounding boxes, the representative of the object position, contained four attributes assembled with the x-y coordinate of the top-left position and the x-y coordinate of the bottom-right position.The image annotation results were used in *Feature Extraction* process and another parallel process was *Object detection*.

2) Object Detection: from the box diagram in Fig. 1, this process is an alternative image annotation process that automatically determined the object position and annotated the object type which operated on Tensorflow Object Detection API [6]. To achieve this objective, the annotated data from *Image Annotation* process and the image dataset were the significant ground-truth dataset to tuning the object detection models. However, in *Experiments and Results*, the object detection process was not included, because the result of object position was not precisely determined.

3) Feature Extraction: this process aimed to collect information on the interesting objects from the image dataset along with results from *Image Annotations* or *Object Detection* process. After this process had been completed, two important details were collected; firstly, the object positions of the stereo image in terms of the bounding boxes; secondly, the ground truth distance, the actual object distance, which was measured between the front of the stereo camera and the interesting object in the real environment while the image dataset was being recorded. There were the pre-processed dataset that had 1,290 instances and contained nine attributes as shown in Table I.

C. Regression *models*

The machine learning method, that had been selected to contribute to this research, was regression models. The unique advantage of these methods provides continuous numeric values as a result, that corresponds with our expected results. Accordingly, regression models can be classified into two types based on the fundamental concept of the algorithms as explained below.

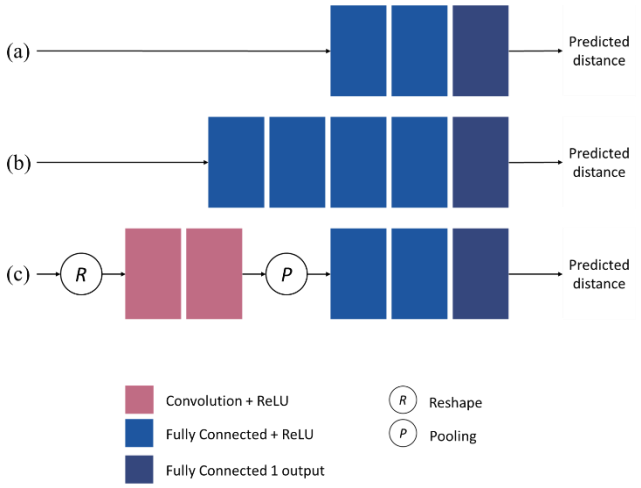


Fig. 3. The architecture of our neuron network regression models. Illustration (a) is the *three-layers Fullsy Connected* regression model, (b) is the *five-layers Fully Connected* regression model, and (c) is *Convolutional with the three-layers Fully Connected* regression model.

1) Linear Regression: Mathematical equations are used to explain the relationships between the attributes in the linear regression model. The prediction process of linear regression is performed by tuning coefficients and constants value of a mathematical equation, although the prediction model is capable to select the optimization function based on the algorithms. Two linear regressions, that were selected to perform in this research, comprised the Ridge regression and the Polynomial Ridge regression, and the linear least-squares function, which shown below in Equation (1), was selected to be the regression loss function.

$$\min_{w} \left|\left|Xw - y\right|\right|_2^2 + \alpha \left|\left|w\right|\right|_2^2 . \quad (1)$$

2) Artificial Neuron Network Regression: There is an unfamiliar *Linear Regression* that uses mathematical equations to explain the relationship of the data. The neuron network layers are performed as the predictive method. The two neuron network regressions, that were proposed to resolve the stereo distance estimation problem, are explained below:

a) Fully Connected model: The architecture of the network regression, illustrated in Fig. 3 (a) and (b), was assembled from three and five fully connected layers, respectively. The first part, the input layer, possessed eight inputs and returns as 64 output nodes; the second part, the hidden layer, the second layer in Fig. 3 (a) and the third and fourth layer in Fig. 3 (b), possessed 64 output nodes; and the final part, the predictive layer, provided a numerical number as a result. Every layer except the predictive layer used Rectified Linear Unit [7] (ReLU) as an activation function.

b) Convolutional with the three-layers Fully connected model: The network architecture was convolutional neuron networks concatenated with a Fully connected model from the model in step *a)*, which is illustrated in Fig. 3 (c). For a

perspicuous explanation, the model has been separated into three parts. The first part initiated by reshape layer; accordingly, two serial network layers were assembled from one-dimensional Convolutional Neuron Networks (CNN) that had ReLU as an activation function, three kernel sizes, one stride, and six filters. The second part was the global average pooling layer, which flattened CNN features for compatibility with the following part. The third part, the *Fully connected model* from strep *a)* was reused and returned the result.

D. Evaluation

The objective of this research is to estimate the distance between the stereo camera and the interesting object, the result which is given as a numeric value. Thus, standard regression metrics were applied to evaluate the performances.

Let $\bar{y}_i$ is the estimated distance and y_i is the ground truth distance, and N is the total number of images in the image dataset. The regression metrics were adopted in this research comprised:

- Mean Square Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=0}^{N}\left(\bar{y}_i - y_i\right)^2. \quad (2)$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=0}^{N}\left|\bar{y}_i - y_i\right|. \quad (3)$$

- Mean Square Logarithmic Error (MSLE):

$$MSLE = \frac{1}{N}\sum_{i=0}^{N}\left(\log \bar{y}_i - \log y_i\right)^2. \quad (4)$$

The most important metric in this research is the *MSE* as shown in (2). This was applied for many purposes, not only the evaluation of model performance but also the loss function of model optimization while the *Artificial Neuron Network Regressions* were deployed on the tuning process.

IV. Experiments and Results

The experiments consisted of two sections: evaluations and implementations. The evaluations discussed on the experimental methods and provided feedback on our proposed methods. The implementations reviewed on the software implementation, using API, and hardware details which described below.

A. Software implementation and Hardware details

The experimental environments, implementation method, and programming interface that were used in this research are reviewed. There were arranged based on the proposed method.

Input images: The only *camera*, that recorded the image dataset, is the ZED stereo camera, which is connected to a computer via USB 3.0 interface. The essentials of the image dataset were recorded in the SVO video file format, the specific file format of the ZED SDK. The image dataset was transformed into a pair of images via ZED python API [8].

Feature Extraction: TensorFlow Object Detection API [6] was deployed to extract the object position of the image dataset, that provided the results in term of bounding boxes.

Regression models and Evaluation: These were separated into two programming modules: first, the Linear Regression models adopted the ready-to-use Ridge regression model and implemented the Polynomial Ridge regression from Scikit-learn [9]; second, the Artificial Neuron Network Regression models were implemented by Keras, higher-level machine learning library interface, sub-module within TensorFlow [10].

Apparently, TensorFlow is a primary module that was used in our various processes, and there also supports to compute and deploy to the graphics processor. Accordingly, the experimental environments were set up on an 8th generation Intel Core i7 with Nvidia GTX 1070 CUDA support.

B. Models evaluation

A total of 80 percent of the pre-processed dataset was randomly selected to tuning the proposed models from Section III.C., *Regression models*. The other 20 percent was reserved to be the testing dataset. Furthermore, the initial random seed, random algorithm, and initialize value of the training models were deliberately set for reproducibility. After the proposed models had been trained, 20 percent of the pre-processed dataset that was reserved was used to evaluate the model performance by using the evaluation metrics that were described in section III.D., *Evaluation*.

In the experiment, to achieve the best results, there had many experimental methods were designed for the *Artificial Neuron Network Regression* models as follows; optimization algorithm comprised of Adam [11] and RMSprop [12] with either full batch method or mini-batches methods; and the learning rate possessed two types, that consisted of constant value and exponential decay, 0.001 initial learning rate with 0.0001 decay rate on 4,000 decay steps. The results in terms of the evaluation metrics are reported in Table II and Fig. 4,

TABLE II COMPARISON OF OPTIMIZATION METHOD AND FEEDING DATASET METHOD OF NEURAL NETWORK MODEL

Model[b]	Optimizer	Batch	LR[c]	Evaluation Error[d]		
				MSE	*MAE*	*MSLE*
FC	Adam	Full	Exp-d	1758.086	968.343	140.801
		8	Exp-d	7802.153	2379.154	818.442
		8	Const	17.217	99.291	1.719
	RMSprop	Full	Exp-d	344.882	415.859	48.087
		8	Exp-d	2074.986	958.569	190.490
		8	Const	38.369	149.896	3.299
FC5	Adam	Full	Exp-d	171.011	346.622	24.331
		8	Exp-d	73.446	210.937	15.217
		8	Const	10.649	75.863	0.902
	RMSprop	Full	Exp-d	5.969	51.695	0.513
		8	Exp-d	13.871	86.477	1.227
		8	Const	2.001	28.380	0.160
CNN+FC	Adam	Full	Exp-d	114.333	278.320	12.003
		8	Exp-d	222.941	376.441	24.773
		8	Const	9.566	76.881	0.882
	RMSprop	Full	Exp-d	103.374	235.482	10.734
		8	Exp-d	87.285	221.348	8.275
		8	Const	0.531	14.490	0.048

[b.] FC: three layers Fully Connected; FC5: five layers Fully Connected; CNN+FC: Convolutional with Fully Connected

[c.] LR: Learning rate

[d.] unit: 10^{-3}

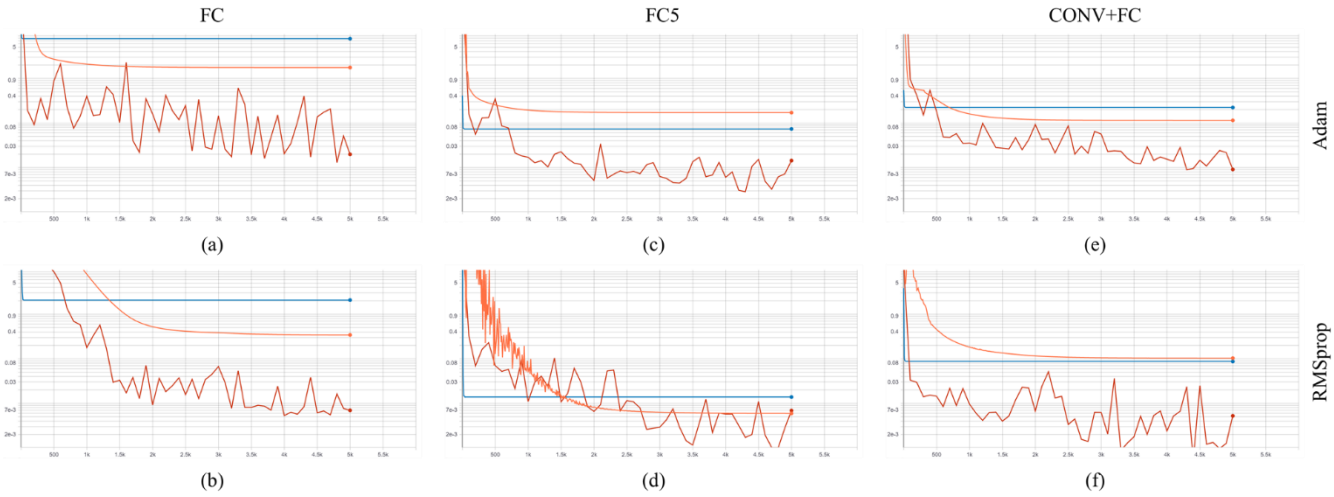


Fig. 4. The experimental result of Artificial Neuron Network Regression models on various optimization, batch feeding, and learning rate method using identical evaluation dataset. In each figure, the x-axis is the training step, and the y-axis is the logarithmic scale of the mean square error evaluation metric; the yellow line denotes full batch with exponential decay learning rate, the blue line denotes mini-batches with exponential decay learning rate, and the red line denotes mini-batches with constant learning rate. Figure (a), (c), and (e) denote Adam [11] optimization, and figure (b), (d), and (f) denote RMSprop [12] optimization. Figure (a) and (b) denote the *three-layers Fully Connected model*, figure (c) and (d) denote the *five-layers Fully Connected model*, and figure (e) and (f) denote *Convolutional with the three-layers Fully Connected model*.

TABLE III COMPARISON OF EVALUATION ERRORS OF PROPOSED MODELS ON IMAGE DATASET

Model	**Evaluation Error[e] (*lower is better*)**		
	MSE	***MAE***	***MSLE***
Camera API	76.509	202.943	4.017
Ridge	84.355	244.763	6.840
Polynomial Ridge	7.121	52.151	0.723
FC	38.369	149.896	3.299
FC5	2.001	28.380	0.160
CNN+FC	**0.531**	**14.490**	**0.048**

[e] unit: 10^{-3}

TABLE IV COMPARISON OF DISTANCE IN TERM OF MEAN SQUARE ERROR OF PROPOSED MODELS ON IMAGE DATASET

Model	**Mean Square Error[f] of Actual Distance(meter)**				
	1	***1.5***	***2***	***2.5***	***3***
Camera API	8.919	5.524	3.017	**1.619**	**1.285**
Ridge	67.132	98.017	67.770	17.051	145.505
Polynomial Ridge	11.391	8.045	2.220	7.973	5.721
FC	0.395	2.541	7.096	8.010	14.502
FC5	**0.128**	0.694	1.049	4.084	22.617
CNN+FC	0.632	**0.219**	**0.393**	8.143	13.572

[f] unit: 10^{-3}

that showed the RMSprop optimizer, mini-batches method, and the constant learning rate significantly reduced the evaluation errors.

Accordingly, the first row in Table III is an essentially mathematical pixel-by-pixel distance calculation method provided by the camera API [8]; these were converted to object distance by averaging pixels of the interesting object in the area of the bounding box, and the estimation metrics were applied. The comparison of the state-of-the-art models, which reported in Table III, used our best performing method, suggested by the results in Table II, the RMSprop optimizer with mini-batches and the constant learning rate. The report showed the *Convolutional with fully connected* model was the best performing model in all evaluation metrics.

To observe the variance of state of the art, predictions of the testing dataset were evaluated by MSE to demonstrate the model performance based on the actual distance between the stereo camera and the interesting object. The variances of the state-of-the-art models, which reported in Table IV, showed *Convolutional with fully connected* and *five-layers Fully connected* model efficiently operated on a short distance, but higher MSE error on long distance might cause by the insufficient dataset or model complexity.

Accordingly, the results from the camera API calculation possessed naturally likelihood to convergence when the distance gradually far away, because the depth equation, that calculated the disparity and used in camera API, was directly developed based on mathematical geometry theorem {ref}. Therefore, the depth equation generated three-dimensional measurement based on the distance between pixel of the object. If the object in the image far away, errors of the distance between pixels was dramatically reduced that influenced MSE as well.

V. CONCLUSION

This paper presented the novel distance estimation for calculating distances from the stereo camera to the interesting object using machine learning algorithms. The image dataset for distance estimation was organized, that contained the ground truth distance in the meter unit. There might be a not solid precision, but based on other research, e.g., Liao et al. [3] and Mustafah et al. [4], had justified the possibility for manipulating the results appropriately.

As described in the above section, the report indicated the potential of machine learning algorithms to resolve the stereo distance estimation problem even by using the straightforward model and a raw dataset from the feature extraction technique.

VI. DISCUSSION

In the experiment, we had a curiosity about the performance of *Convolutional with Fully Connected* compared with *Fully Connected*. We started to investigate, and the first hypothesis was the two additional layers of CNN over the FC might provide the chance of attributes to spread throughout the layers, so we experimented by adding two FC layers over

FC model, called FC5. The results suggested the number of layers in models has significant.

ACKNOWLEDGMENTS

This research was supported by the Department of Computer Engineering, Faculty of Engineering, Mahidol University.

REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1627-1645, 9 2010.

[2] J. Zhang, S. Hu and H. Shi, "Deep Learning based Object Distance Measurement Method for Binocular Stereo Vision Blind Area," *International Journal of Advanced Computer Science and Applications*, vol. 9, 1 2018.

[3] Y. Liao, L. Huang, Y. Wang, S. Kodagoda, Y. Yu and Y. Liu, "Parse geometry from a line: Monocular depth estimation with partial laser observation," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[4] Y. M. Mustafah, R. Noor, H. Hasbi and A. W. Azma, "Stereo vision images processing for real-time object distance and size measurements," in *2012 International Conference on Computer and Communication Engineering (ICCCE)*, 2012.

[5] Tzutalin, "LabelImg," Git code, 2015. [Online]. Available: https://github.com/tzutalin/labelImg. [Accessed 23 9 2019].

[6] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," *CoRR*, vol. abs/1611.10012, 2016.

[7] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, USA, 2010.

[8] "GitHub: Python API for the ZED SDK," Stereolabs Inc., [Online]. Available: https://github.com/stereolabs/zed-python-api. [Accessed 28 August 2019].

[9] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015.

[11] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[12] G. Hinton, N. Srivastava and K. Swersky, "Neural Networks for Machine Learning," [Online]. Available: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. [Accessed 28 August 2019].

Automatic Football Match Event Detection from the Scoreboard using a Single-Shot MultiBox Detector

Rungroj Somwong
Department of Computer Engineering
Image, Information and Intelligence Laboratory
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
rungroj.sow@student.mahidol.ac.th

Narit Hnoohom
Department of Computer Engineering
Image, Information and Intelligence Laboratory
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
narit.hno@mahidol.edu

***Abstract* — During a football match, the information is manually collected by humans. However, the correctness of the football match data is difficult to check because of the game's speed, and thus, human errors can occur. This paper presents an automatic football match event detection from the scoreboard using a deep learning algorithm. The proposed method can reduce human error and performs the detection faster. The study included 68 match of English English Premier League 2017-2018 broadcast videos and tested with 80 testing videos. These videos were prepared from 20 matches broadcast videos, which consisted of twelve matches from the year 2017-2018 and eight matches from the year 2018-2019. The proposed method contains three main steps: data gathering and augmentation, object detection for scoreboard visualization forms, and the event classification. The scoreboard detection is performed with an SSD. The event classification employs the majority vote and time frame technique. The experimental results show an accuracy rate of 1.00 with the expected event scoreboards, comprised of Goal, Substitution, and Card events.**

Keywords — football match, data augmentation, image processing, deep learning, object detection

I. INTRODUCTION

Football is the most popular sport in the world. Many methods have been developed to improve the game in different fields. Significant methods include Goal-line Technology (GLT) and Video Assistance Referee (VAR), which were officially applied in the 2018 FIFA World Cup in Russia.

GLT is a high accuracy goal event detection. The concept is to detect the location of the ball and goal line with high-speed cameras that are set up around the stadium. If the ball passes the goal line, then it is a goal event. It supports the referees when judging an ambiguous goal event.

VAR is an assistance replay video to determine an unclear event. Replay videos are recorded by various types of cameras around the stadium. They are watched by the VAR team in the VAR room in order to assist the referees. The VAR team will notify the referees when an unclear event occurs. A referee can decide to check the replay videos and correct the judgment.

These methods mentioned are current football match quality improvements. Another essential improvement involves the football match information, which is important for most everyone in football. Teams and players require it to develop themselves for the competition, and football fans require it to follow the match results and events. There are websites that provide match event information such as goals, substitutions, cards, etc. At present, this information is collected by humans manually, and human error can occur. Thus, automatic event information detection will prevent human error and reduce human labor.

In this research, the researchers provide a method to detect the match events automatically. The primary concept is that different scoreboard visualization forms can represent the different match events. In a football match event, this paper focuses on three match events: Goals, Substitutions, and Cards. The proposed method can be separated into three main steps as follows.

1) Data gathering and augmentation: The data are gathered from football broadcast video frames and augmented by using image processing techniques.

2) Object detection for scoreboard visualization forms: A Single-Shot MultiBox Detector (SSD) is used as an object detection model. The model was trained with three classes: Goal, Substitution, and Card.

3) Event classification: Classification is needed after the detection to filter incorrectly detected events by using a majority vote and time frame technique.

The rest of this paper is organized as follows. The related works are introduced in Section II. The proposed method is briefly described in Section III. Section IV explains the dataset preparation. Section V outlines the process of the scoreboard detection. The event classification is investigated in Section VI, and the experimental results are summarized in Section VII. Finally, Section VIII provides the conclusion and discussion on the results of the experiment.

II. RELATED WORKS

The creation of a new dataset was needed in this research because there is no prepared data set for football broadcast scoreboards. Unfortunately, the amount of created data was not efficient to train the model. Thus, the data augmentation technique was required. We adapted the data augmentation technique from the Learning Data Augmentation Strategies for Object Detection research [1]. Zoph et al. proposed these data augmentation strategies for object detection. The strategy is to randomly select a group of image processing operations, which are called the "sub-policy". Each image processing operation has a probability that is applied to the source image.

For the deep learning model, the focus was on an object detection model, and the Single-Shot MultiBox Detector (SSD) [2] was selected as the object detection model. The

model contains two parts: Classification layers and Detection layers. It has fast detection and good accuracy. The original SSD300 has a speed of 46 frames per second and an accuracy of 77.2 mAP.

There are similar works related to football event detection from broadcast video or scoreboard detection in other sports. Firstly, Yang et al. [3] presented a football goal event detection from the scoreboard in videos. They used image processing techniques, such as the Sobel operator and morphology, to locate the scoreboard, and then detect the goal by checking the differences of the scoreboard in every video frame. The results of the experiment had nearly 95% recall and a low average precision of 75%.

Hung et al. [4] proposed a baseball event detection model in broadcast videos. The method started with scoreboard information extraction. The scoreboard was located by using image averaging in videos. Then, the text was recognized with a Neural Network Classifier. After the scoreboard extraction, the shot transition pattern of the video was used to improve the event detection. The Bayesian Belief Network (BBN) classified the shot transition pattern. This research had 92% recall and 95% precision.

Agyeman et al. [5] proposed a football video summarization by using deep learning. They used an improved 3D action recognition Convolutional Neural Network (CNN) for football videos feature extraction. Then, the 3D-CNN and Long Short-Term Memory (LSTM) were used as the recognition framework and to summarize the video. The accuracy of this research was 96.81%.

Eldib et al. [6] proposed a football video summarization using enhanced logo detection. This research used image processing techniques to detect a replay logo to get a replay shot and then, classified the shot and detected an event with a rule-based classifier. The proposed method had 100% recall and 91.7% precision.

Bojukrapan et al. [7] proposed an automatic football summarization by using time constraints. First, they used image processing techniques to obtain a shot boundary. Then, the event from the shot was detected with various image feature extractions, and the event was assigned with a weight. Finally, the events were summarized with the knapsack problem and generated into a 1-minute clip. This method was evaluated with a satisfaction survey, and the satisfaction level was an average of 4.07.

III. PROPOSED METHOD

The match event detection method begins with the dataset preparation. This step is very important for the proposed method as it can determine the method of performance. Then, the training of the SSD model, which is the object detection model used to detect the visualization of the scoreboard to indicate the match event, was conducted, which are displayed in Fig. 1. The detection will provide the raw event detections, which need to be clarified with the scoreboard visualization criteria in order to reduce the detection errors and summarize the outputs of the events.

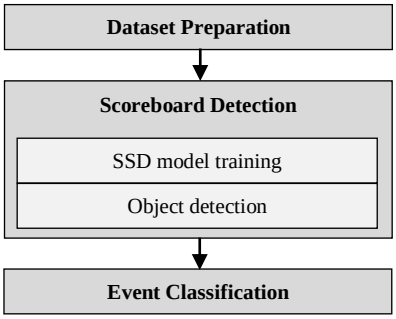


Fig. 1. Block diagram for the proposed method

IV. DATASET PREPARATION

A. Data Gathering

The dataset was gathered from 68 full match broadcast videos of English Premier League in 2017-2018. The video was extracted to frames with a one frame per second extraction rate. Each frame contained the scoreboard, which indicates the match event. Three match events were considered: Goals, Substitutions, and Cards. Every frame of these classes were gathered into the dataset.

B. Data Annotation

The scoreboards in the videos can be separated into two types: the main scoreboard and the popup scoreboard. The main scoreboard is always displayed on the frame to show the basic match information such as the teams, score and time. The popup scoreboard displays the match event information such as cards and goals only when these events are occurring.

However, the main scoreboard does not display only the basic match information. It also displays some match event information by transforming the visualization when the event occurs. The remaining event information is still displayed in the popup scoreboard with a different visualization for each event. Therefore, we can use the different visualizations from the main scoreboard and the popup scoreboard to indicate the events. The scoreboard information display criteria can be different in each football league according to the league broadcast design.

For English Premier League 2017-2018, Substitution and Card events are displayed on the main scoreboard, while Goal events are displayed on the popup scoreboard. We annotated the area in the frame for these three event classes based on these criteria, which are displayed in Fig. 2.

C. Data Augmentation

We adapted the data augmentation strategy from the research of Zoph et al. [1]. Only the policy application part is used in this research. A policy for the match event detection

(a) (b)

(c)

Fig. 2. Annotation area for English Premier League 2017-2018: (a) Annotation area for a Substitution event on the main scoreboard; (b) Annotation area for a Card event on the main scoreboard; (c) Annotation area for Goal event on popup scoreboard.

Fig. 3. Data augmentation with sub-policy (Annotation area flip with probability of 0.6 and brightness with probability of 0.6 and magnitude of 3)

dataset contains 20 sub-policies. The sub-policy is randomly selected and then applies a group of image processing operators to the source data for new data generation. Each sub-policy has two different image processing operators with the same applied probability (0.6) and a different magnitude (1-10). There are five operators that are suitable for the match event detection dataset.

1) Annotation area flip. It flips the image only in an annotation area. Scoreboard visualization can be symmetrical or mirrored due to the information for both teams. This operation does not use magnitude in the process.

2) Horizontal transition. It transitions the overall image and annotation area to align with the horizontal axis.

3) Vertical transition. It is similar to the Horizontal transition, but the transition is aligned with the vertical axis.

4) Gaussian blur. This simulates the degraded video quality by using the Gaussian blur technique.

5) Brightness. It can be adjusted to dark (magnitudes 1 - 5) or bright (magnitudes 6 - 10) in order to generate different match brightness environments.

We created 30,000 augmented training data items, which is sufficient for object detection model training. Fig. 3 is an example of the augmented data. There are 11,100 data for Goal class, 14,643 data for Substitution class and 4,257 data for Card class.

V. Scoreboard Detection

A. Detection Model

The Single-Shot MultiBox Detector (SSD) [2] is an object detection model that has fast detection and good accuracy. An original SSD300 has a speed of 46 frames per second and an accuracy of 77.2 mAP.

The SSD is used in this research to detect the match events via scoreboard visualizations from the frames. Each event can be indicated by the specific visualization of the scoreboard. We used the SSD with the inception from a TensorFlow Zoo model with slight parameter modification.

1) Input image size. It was modified from 300×300 pixels to 640×360 pixels for the frame size compatibility and aspect ratio. The frame size is 1,280×720 pixels, which is too large to train the model. Thus it was decided to modify the model input image resizer to 0.5 times of the frame size (640×360 pixels) in order to maintain the correct aspect ratio and size compatibility without a significant model performance reduction.

2) Optimizer. After numerous model training experiments, a suitable optimizer for this detection was found to be the Adam optimizer with an exponential decay learning rate. The initial learning rate is 0.002 and the decay factor is 0.75 in every 1,000 steps.

B. Training and validation

The model was trained with the three classes (Goal, Substitution, and Card) in 40,000 steps. The training dataset had 30,000 items from the data augmentation, and the validation dataset included 1,621 items (5.13% of total data). The final total losses of this model training were 0.15.

C. Detection

Scoreboard visualizations are detected from the frames, which are extracted from the broadcast videos at two frames per second extraction rate. Detection outputs are able to represent the match event occurring within the frame. They are filtered with a 95% detection score threshold.

VI. Event Classification

Outputs from the detection are the event detection for each frame. Each match event in the videos consists of many frames containing the same detected event. Thus, the detected event frames need to be grouped and clarified to summarize the match events in the videos.

Detected event frames are grouped by event type. Each group contains attributes for event classification, and every group is stored in a pool. The group is created in the pool if the detected event frame contains the event type that is a non-existing group in the pool. If this event type group exists, it will be updated with a frame counting number to count the valid detected frames in the group and the group interval time to count the interval time of the group. In every interaction, the total frame counting number is continually increased for every group and the timeout value is increased for non-match event type groups. The timeout is reset when the group is updated. It is used as a flag to complete the group by breaking the sequence of the detected frames in the group. The timeout is set to 1 second in this research. The group will be complete and ready for event classification when the timeout is greater than 1 second.

Event classification is based on the scoreboard event visualization interval time. The group interval time should be equal to or greater than the shortest visualization interval time. We observed and found that the shortest interval time for the scoreboard event visualizations of English Premier

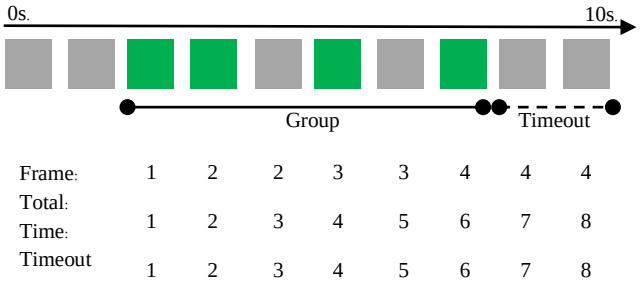


Fig. 4. Example of detected event frame grouping (Gray boxes are frames with undetected events or different detected event types. Green boxes are frames with the same detected event type.)

League 2017-2018 is 2 seconds. Then, the majority vote is used to verify and summarize the match event by checking the frame counting number, which should be more than half of the total frame counting number.

VII. Experimental Results

The performance speed is 200 milliseconds per frame or 5 frames per second. The detection was trained with 30,000 data of Goals, Substitutions and Cards scoreboard from 68 matches The experiment was separated into two parts: scoreboard detection and event classification. They were examined with 80 testing event videos from 20 matches of English Premier League broadcast videos, which consist of twelve matches from the year 2017-2018 and eight matches from the year 2018-2019. The videos are Goals, Substitutions, Cards and No events, which is an extra event to test the accuracy of the scoreboard detection between expected and unexpected events. Examples of No event are displayed in Fig. 5. Substitution events is the group with the most videos (32 videos) due to it being the most common event occurring in a football match. For Goal and Card events, there were fourteen videos for each event because their average event occurring number is almost half that of the substitutions. The remaining 20 videos are No events, as shown in Table I.

Experimental testing of the proposed method was conducted on a personal computer with an Intel Core i5-4570 CPU, Nvidia GTX 1060 6G graphics card and 10 GB RAM.

The duration time of the testing videos for both parts of the experiment were different. The videos for the scoreboard detection experiment contained only a set of frames that display unaltered event types of the scoreboard. This set was defined as the main event frames. An example is a video that contains five frames that are displaying the Goal event scoreboard.

The video duration time for the event classification is longer, lasting for 10 seconds. It consists of the main event frames and 5-second video sections that appeared before and after the main event frames. The extra video section is needed due to the process of event classification.

A. *Scoreboard Detection Experiment*

In the first experiment, the focus was on the accuracy of the scoreboard detection between the expected and the unexpected events from the frames. The expected events include Goal, Substitution and Card events, while the others are considered an unexpected event. We tested the detection with all frames from the 80 testing videos. Detected results were grouped into Expected and Unexpected classes in order to determine the accuracy.

As shown in Table II, the results indicate that the higher true positive (TP) rate of 96.94% belonged to the *Unexpected Events*. The lower TP rate was 90.92%, which belonged to the Expected Events.

The results were reported with respect to the true positive (TP), true negative (TN), false positive (FP), false negative (FN), recall, precision, and accuracy, as shown in Table III. The detection accuracy was 0.92. There were several incorrect detections in which unexpected event scoreboards were classified as an expected event based on the precision of the expected event class. The root cause is the similarity of the visualization of the information scoreboard

Fig. 5. No event scoreboard as unexpected event detection: (a) Normal scoreboard with time (most displayed scoreboard in the football broadcast video), (b) Information scoreboard style 1, (c) Information scoreboard style 2, (d) Added time information scoreboard.

TABLE I. Testing Event Videos
(G = Goal, S = Substitution, C = Card and N = No Event)

Match broadcast videos	Testing video			
	G	S	C	N
Bournemouth vs. West Bromwich Albion (2017-18)	1	2	0	1
Huddersfield Town vs. Crystal Palace (2017-18)	1	2	1	1
Stoke City vs. Everton (2017-18)	1	1	1	1
Newcastle United vs. Huddersfield Town (2017-18)	0	2	1	1
Watford vs. Bournemouth (2017-18)	1	1	1	1
West Ham United vs. Southampton (2017-18)	1	1	1	1
West Bromwich Albion vs. Swansea (2017-18)	1	2	1	1
Burnley vs. Leicester City (2017-18)	1	2	1	1
Huddersfield Town vs. Watford (2017-18)	0	2	1	1
Liverpool vs. Bournemouth (2017-18)	1	2	0	1
Swansea vs. Everton (2017-18)	1	2	0	1
Leicester City vs. Southampton (2017-18)	0	2	0	1
Brighton vs. Watford (2018-19)	0	1	1	1
Arsenal vs. Bournemouth (2018-19)	1	1	0	1
Crystal Palace vs. Huddersfield Town (2018-19)	1	1	0	1
Watford vs. Southampton (2018-19)	0	1	1	1
Wolverhampton vs. Arsenal (2018-19)	1	1	1	1
Burnley vs. Arsenal (2018-19)	1	2	1	1
Leicester City vs. Chelsea (2018-19)	0	2	1	1
Watford vs. West Ham United (2018-19)	1	2	1	1
Total	14	32	14	20

TABLE II. Results for Expected and Unexpected Events Detection

	Expected	Unexpected
Expected	761	76
Unexpected	9	285

TABLE III. Measurements for Expected Events Detection

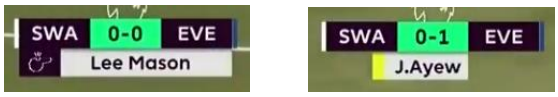
TP	TN	FP	FN	Recall	Precision	Accuracy
761	285	76	9	0.99	0.91	0.92

(unexpected event) to the expected event scoreboard. For example, a referee information scoreboard (a) is similar to a Card event scoreboard (b), as seen in Fig. 6.

For false negative (FN) value, some frames of expected events were not detected. These are Card event from Stoke City versus Everton match. The frame displays the card event scoreboard with a player name who got the card. The name length is too short and similar to the No event scoreboard visualization. A comparison between the problem and No event scoreboard visualization is shown in Fig. 7. This problem occurs only on Card event scoreboard detection because its visualization is the most similar to the No event scoreboard. Player name length is only one feature that can separate class of these events. If the player name is too short, these events will not be separated (Card event cannot be detected). Therefore, the problem depends on the player's name length.

Following this, the scoreboard detections were tested with all frames from the 79 videos, which exclude the No event videos and one video that missing to detect the event as described in the previous paragraph, in order to determine the accuracy of the events identified in the scoreboard detections. This experiment considered three event classes: Goal, Substitution, and Card.

The results of the *Goal*, *Substitution*, and *Card* classes are shown in Table IV. All TP rates were 100%. All events can be detected to the correct class due to the class visualization which is completely different from each other classes as seen in Fig. 8.



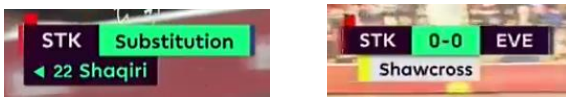
(a) (b)

Fig. 6. The similarity of the information scoreboard and event scoreboard: (a) Referee information scoreboard; (b) Card event scoreboard.

(a) (b)

Fig. 7. The similarity of Card event scoreboard and No event scoreboard: (a) Card event scoreboard with short player name; (b) No event scoreboard.

(a)



(b) (c)

Fig. 8. Correct event scoreboard detection: (a) Goal class, (b) Substitution class, (c) Card class.

The accuracy rate was 1.00 as seen in Table V. It demonstrates the excellent accuracy in detecting the correct event in the scoreboard detections.

In all of the scoreboard detection experiments, there were detection problems with the unexpected event scoreboard when an information scoreboard was detected as an expected event and the short player name length of Card event scoreboard was not detected as an unexpected event. The similarity of scoreboard visualizations is a root cause of these problems.

On the other hand, when the expected event scoreboard is detected, the detection method can identify the event type with a 100% success rate.

B. Event Classification Experiment

The experiment was conducted with 79 testing event videos for Goal, Substitution and Card events. All frames of each video were processed by scoreboard detection before the classification. Only the correct detection frames were used as an input in the event classification. Thus, one event video with Card event detection missing problem is excluded from a total of 80 event videos.

The results of the *Goal*, *Substitution*, and *Card* classes are shown in Table VI, and the TP rate was 100%. The detection result is the input of the event classification. The input of this experiment was included only with correct detection results to determine the classification process.

As shown in Table VII, the event classification had excellent accuracy with a rate of 1.00. All frames of each video were grouped and classified as the correct event.

TABLE IV. RESULTS FOR GOAL, SUBSTITUTION AND CARD EVENTS DETECTION

	Goal	Substitution	Card
Goal	190	0	0
Substitution	0	481	0
Card	0	0	90

TABLE V. MEASUREMENTS FOR GOAL, SUBSTITUTION AND CARD EVENTS DETECTION

	TP	TN	FP	FN	Recall	Precision	Accuracy
Goal	190	571	0	0	1.00	1.00	1.00
Substitution	481	280	0	0	1.00	1.00	
Card	90	671	0	0	1.00	1.00	

TABLE VI. RESULTS FOR EVENT CLASSIFICATION

	Goal	Substitution	Card
Goal	32	0	0
Substitution	0	14	0
Card	0	0	13

TABLE VII. MEASUREMENTS FOR EVENT CLASSIFICATION

	TP	TN	FP	FN	Recall	Precision	Accuracy
Goal	32	27	0	0	1.00	1.00	1.00
Substitution	14	45	0	0	1.00	1.00	
Card	13	46	0	0	1.00	1.00	

VIII. Conclusion

This research was focused on automatic football match event detection from scoreboards through the use of a deep learning algorithm. The proposed method consisting of two main processes, scoreboard detection and event classification, was used to extract the football match events. Videos from ten matches of English Premier League broadcasts were used for the prediction testing. The experimental results showed that the accuracy rate of the event classification is 1.00.

The proposed method works well with the expected event scoreboards with regard to the scoreboard detection for expected events and event classification accuracy. However, there are problems with the scoreboard detection for unexpected events and Card events with short player name length. The detection results are an input of the event classification, and the incorrect data has an effect on the classification results and the accuracy of the method.

Therefore, studies for the future research works should include scoreboard detection improvement for the event scoreboards with similar visualization as well as enhancement of the accuracy of the event information extraction from the classified events.

Acknowledgments

This research was supported by the Department of Computer Engineering, Faculty of Engineering, Mahidol University.

References

[1] Barret Zoph, Ekin D. Cubuk, Golnaz Ghiasi, Tsung-Yi Lin, Jonathon Shlens and Quoc V. Le, "Learning Data Augmentation Strategies for Object Detection", arXiv preprint, arXiv:1906.11172, 2019.

[2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, "SSD: Single Shot MultiBox Detector", arXiv preprint, arXiv:1512.02325, 2016.

[3] Song Yang, Wen Xiangming, Sun Yong, Zhang Liang, Yan Lelin and Lin Haitao, "A Scoreboard Based Method for Goal Events Detecting in Football Videos", 2011 Workshop on Digital Media and Digital Content Management, 2011.

[4] Mao-Hsiung Hung and Chaur-Heh Hsieh, "Event Detection of Broadcast Baseball Videos", IEEE Transactions on Circuits and Systems for Video Technology, 2008.

[5] Rockson Agyeman, Rafiq Muhammad and Gyu Sang Choi, "Soccer Video Summarization Using Deep Learning", 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019.

[6] Mohamed Y. Eldib, Bassam S. Abou Zaid, Hossam M. Zawbaa, Mohamed El-Zahar and Motaz El-Saban, "Soccer Video Summarization Using Enhanced Logo Detection", 2009 16th IEEE International Conference on Image Processing (ICIP), 2009.

[7] Sutthikarn Bojukrapan and Punpiti Piamsa-nga, "Automatic Soccer Archive Summarization Using Time Constraint", 2015 International Computer Science and Engineering Conference (ICSEC), 2015.

A Light-Weight Deep Convolutional Neural Network for Speech Emotion Recognition using Mel-Spectrograms

Kamin Atsavasirilert*, Thanaruk Theeramunkong*, Sasiporn Usanavasin *,
Anocha Rugchatjaroen †, Surasak Boonkla †, Jessada Karnjana †, Suthum Keerativittayanun *†,
and Manabu Okumura ‡

* Sirindhorn International Institute of Technology, Thammasat University,
Pathum Thani, Thailand
† NECTEC, National Science and Technology Development Agency,
Pathum Thani, Thailand
‡ Tokyo Institute of Technology, Tokyo, Japan
kamin.atsavasirilert@gmail.com

***Abstract*—At present, speech emotion recognition is a challenging field of studies about emotions of speakers. Speech emotion recognition also enhances interaction between humans and machines. In many situations (e.g. embedded systems), we have to detect emotion in speech within limitation of both computing and memory resources. Even several previous works reported that a reasonable recognition rate can be achieved using transfer learning techniques with popular models, such as AlexNet, they suffered with a large model size and can not be executed on an embedded system. To address this problem, we propose a light-weight deep convolutional neural network architecture, which utilizes only partial component of the AlexNet with Log-Mel-Spectrograms as input. Our result shows that the proposed light-weight model can achieve a comparable recognition rate with the state of the art, but the number of parameters used in our model decreases around 272 times from the AlexNet.**

***Index Terms*—speech emotion recognition, Mel-spectrogram, light-weight model, convolutional neural network, deep learning**

I. Introduction

Speech Emotion Recognition (SER) is one of vital tasks since it can improve human-machine interactions [1]–[4]. SER is usually based on supervised learning. Many classifiers, such as k-nearest-neighbor [5], artificial neural network [6], Gaussian mixture model [7], hidden Markov model [8] and support vector machine [9], have been proposed to achieve high recognition rates. However, Iliou and Anagnostopoulos recently showed that a method based on deep learning outperforms other classifiers in terms of recognition rate [10]. There are three main types of classical neural networks. The first one is the fully-connected neural network (FNN). the second is the recurrent neural network (RNN), and the last is convolutional neural network (CNN) [11]–[13].

There are features used in SER [14]–[17]. For example, Liu *et al.* investigated feature fusion between cepstum, mel-frequency cepstral coefficients (MFCC) and synthetically enlarged MFCC [15]. Lalitha *et al.* studied combining of spectral-based features and pitch-based hyper-prosodic features [14]. Wang *et al.* proposed a new Fourier parameter model using voice quality and the first-order and second-order differences [16]. Bandela and Kumar proposed a new feature called Teager-MFCC, which is the combination of MFCC and Teager energy operator (TEO) [17]. However, Niu *et al.* and Zhang *et al.* recently proved that using only spectral features can achieve high recognition rates. They also used the transfer learning technique from AlexNet model [13], [18], [19].

For real world applications, any model should be implemented on an embedded system of which it resources are limited. Although AlexNet is proved to be an effective model to achieve a high recognition rate, it takes too many hardware resources in terms of size of the model. Another drawback of using AlexNet for SER is that it was originally designed for the image classification task, which is called ImageNet [20], [21]. Thus, the model's architecture is not effectively suitable for SER. Moreover, Qian *et al.* showed that using a non-square kernel can increase recognition rate in the automatic speech recognition [20]. Therefore, our study aimed to reduce this pain point by proposing a new light-weight deep convolutional neural network (DCNN). Because it was proved that AlexNet's model architecture can achieve good performance, we adapt and adopt some parts from it. Also, we apply a non-square kernel to the first two convolutional layers [20].

The rest of this paper is organized as follows. Section II describes our proposed method. Section III gives details of our experiments and results. Discussion is made in Section IV, and Section V concludes our paper.

II. Proposed Method

To design DCNNs for SER, there are two aspects to be considered. The first one is input features, which are computed from a speech signal, and the second one is the architecture

of the model. In this section, we address these two aspects of our model.

A. Input Feature

Speech features used in speech emotion recognition can be intuitively divided into four types, which are acoustic features [16], [22], language features [23], [24], context information [25], and hybrid feature [26].

The most popular features are of the acoustic feature, such as prosody features, voice quality features, and spectral features [13], [27]–[31]. Pitches, durations, and loudness, which are prosody features, are commonly used in speech emotion recognition because they express the stress and intonation patterns of a speaker [27]. Tato *et al.* showed that voice quality features can be used to distinguish negative and positive emotions [28]. the first three formants and spectral energy distribution are examples of widely used voice quality features. Later, Tahon *et al.* combined some voice quality features and acoustic features together and showed that the emotion recognition rate improved, compared with using any voice quality features or acoustic features alone [29]. However, recently, Zhang *et al.* oppositely showed that using only a spectral feature and its derivatives could achieve a high recognition rate [13]. In their work, static, delta and delta-delta log Mel-spectrograms were used as the input of their classifier [13].

In addition, Huzaifah *et al.* compared the performance of their model of which the input features were Mel-spectrogram, linear-scale spectrogram, constant-Q transform (CQT) spectrogram, continuous Wavelet transform (CWT) scalogram and MFCC cepstrogram and found that using Mel-spectrogram could achieve the highest recognition rate [30]. In this work, we therefore use the log Mel-spectrogram as the model input.

Additionally, Puterka and Kacur showed that the longer the duration of input, the better the recognition rate [31]. Therefore, we used a sequence of log Mel-spectrograms, as shown in Fig. 1, in our work and compared the recognition rate with the work that used the static, delta and delta-delta log Mel-spectrograms [13].

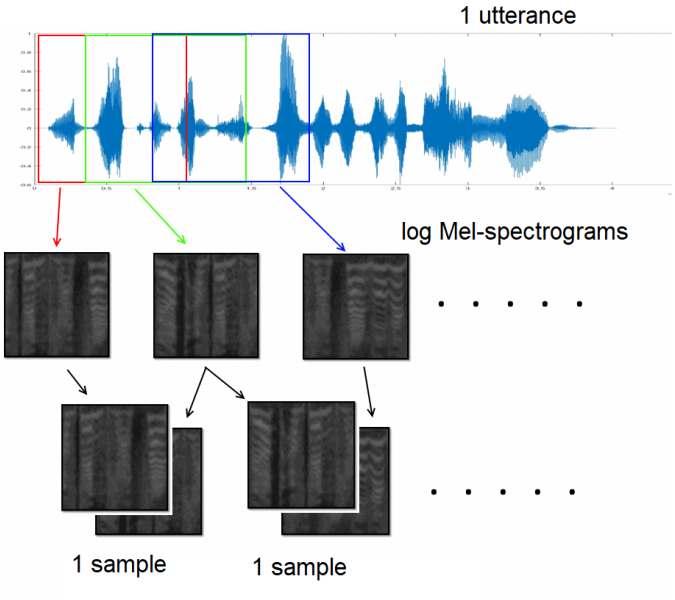


Fig. 1. Example of input construction when one input consists of a sequence of 2 log Mel-spectrograms. Note that we apply 50 percent overlapping.

B. Model Architecture

Recently, the DCNN for SER proposed by Zhang *et al.* is the state of the art in terms of the recognition rate [13]. They used the transfer learning technique on AlexNet [13], [18] in order to achieve a high recognition rate. Since AlexNet consists of about 60 million parameters, which is not suitable for embedded systems, as mentioned before, our proposed model is therefore adopted and adapted from it. Thus, we purpose a new model architecture as shown in Fig. 2.

The proposed architecture consists of two 3×12 convolutional layers, with 32 filters, that follows by a 2×2 max pooling layer. The outputs of the last max pooling layer are fed to three 3×3 convolutional layers with 64 filters and one 2×2 max pooling layer. Then, the last three layers are a flatten layer, a 16-node fully-connected layer, and a 7-node fully-connected layer, which is the output layer.

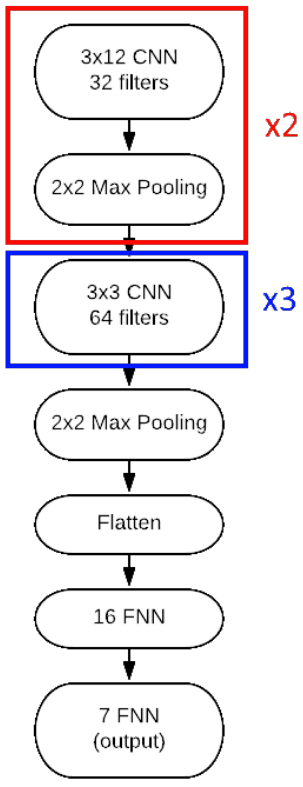


Fig. 2. Proposed model architecture.

III. Experiment and Result

In our experiment, the EMO-DB speech corpus [32] was used to evaluate the model performances. It contains 535 emotional utterances with 7 different emotions, which are anger, joy, sadness, neutral, boredom, disgust, and fear. Those utterances were from 10 German native speakers, which are 5 male and 5 female speakers. The EMO-DB corpus consists of 5 short and 5 long daily life sentences. All speech signals were sampled at a 16 kHz with a 16-bit resolution. The average duration of each utterance is about 3 seconds.

Four objective measures, which are weighted average recall (WAR), unweighted average recall (UAR), f1-score, and accuracy, are used to evaluate our proposed model. In multi-class classification, using accuracy alone is not sufficient because a good model should be able to equally handle each class well. Therefore, WAR, UAR and f1-score are typically used to asses because they take each class recognition rate into account. For class-imbalance dataset, using UAR is more suitable than WAR because it equally reflexes each class recognition rate. An f1-score can clearly indicate balancing between precision and recall.

In our simulation, we use 64-long context window size, with 25 ms long window size and 10 ms shifting, and the number of the Mel-filter bank is set to 64, as suggested by Zhang *et al.* [13]. Hence, our context window size is $63\times10+25=655$ ms long. We used the overlap length of $26\times10+25=285$, which is greater than 250 ms, which is the number suggested by Woellmer *et al.* [33]. Moreover, to avoid the overfitting problem, two dropout layers, a zero mean Gaussian-noise layer, and L2-weight regularization were used. To speed up our training, batch normalization layers were placed after every CNN layer, after every fully-connected layer, and before every ReLU activation function layer [34], [35]. In addition, Bergstra *el at.* have proved that using a random search for the hyper-parameter optimization is effective [36]. Therefore, a random search for the hyper-parameter optimization was deployed in our work to tune the hyper-parameters, dropout-percents, standard deviation of Gaussian noise, L2-weight regularization. We performed the random search for 150 trials. We then selected the best f1-score from those trials.

Before evaluating our proposed model, we firstly investigate the effect of number of log Mel-spectrograms, which is the input of our model. In our experiment, the number of log Mel-spectrograms, which is denoted by N, was varied from 1 to 3. The comparison results are shown in Table I. It can be seen that 3 log Mel-spectrograms can achieve the highest recognition rate. Therefore, we set $N = 3$ for comparison between our proposed model and the state of the art.

TABLE I
Effect of varying number of log Mel-spectrograms.

N	WAR	UAR	f1-score	Accuracy
1	76.84	80.28	62.16	73.49
2	77.75	70.04	64.15	76.73
3	85.54	87.16	78.07	77.99

TABLE II
WAR and UAR comparison between our proposed model and the state-of-the-art models.

References	Features	Classifier	WAR	UAR
Schuller *et al.* [37]	Prosody, MFCC	SVM	85.60	84.60
Stuhlsatz *et al.* [38]	Prosody, MFCC	GerDA	81.90	79.10
Eyben *et al.* [39]	ComParE set	SVM	N/A	86.00
Zhang *et al.* [13]	Static, Delta and Delta-Delta Log Mel-spectrogram	DCNN	87.31	86.30
Proposed model	Sequence of Log Mel-spectrograms	DCNN	85.54	87.16

Table II shows the comparison between our proposed model and the state of the art. It can be seen that our proposed model is slightly better than the others in terms of UAR. However, the WAR values are comparable. We compare our recognition performance with these results because they used the same evaluation method called Leave One Speaker Out [37].

IV. Discussion

In this section, we discuss 2 issues. The first topic is the reason we varied N from 1 to 3, and the second one is about the merit of our proposed model.

First, from Table I, we found that more number of frames can lead to higher f1-score and accuracy. The reason that the max number of frames used in our experiment is 3 is because the shortest utterance duration in the EMO-DB dataset is 1.23 seconds, and 3 frames take $285\times2+655=1.225$ seconds. That means if the number of frames (N) is greater than 3 per input, some data must be omitted which is not suitable for the comparison.

Second, even though we cannot conclude that our proposed model's recognition rate outperforms the others [13], [37]–[39], our model is more light-weight due to the fact that Zhang *et al.* used transfer learning technique from AlexNet model, which contains about 60 million parameters [13], [18]. In contrast, our proposed model takes only about 220 thousand parameters. Therefore, embedded systems can take benefits from this merit.

V. Conclusion

In our work, we studied the effect of varying number of log Mel-spectrograms of the model input. We used random search with 150 trials to compare the results. The results shows that increasing number of log Mel-spectrograms can improve the model's recognition rate. Moreover, our model achieved a comparable recognition rate from the state of the art, which used transfer learning techniques from the AlexNet, by using only 220 thousand parameters.

Acknowledgment

This research is financially supported by Thailand Advanced Institute of Science and Technology, National Science and Technology Development Agency, and Tokyo Institute of Technology under the TAIST Tokyo Tech Program. In addition, it is also partially supported under the Thammasat University's research fund, Center of Excellence in Intelligent

Informatics, Speech and Language Technology and Service Innovation (CILS), and Intelligent Informatics and Service Innovation (IISI) Research Center, the Thailand Research Fund under grant number RTA6080013, as well as the STEM workforce Fund by National Science and Technology Development Agency (NSTDA). Lastly, we would like to express our gratitude to Mr.Parinya Siritanawan from Japan Advanced Institute of Science and Technology (JAIST) for providing us a computing engine.

REFERENCES

[1] Ramakrishnan, S. and El Emary, I.M., 2013. Speech emotion recognition approaches in human computer interaction. Telecommunication Systems, 52(3), pp.1467-1478.

[2] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. and Taylor, J.G., 2001. Emotion recognition in human-computer interaction. IEEE Signal processing magazine, 18(1), pp.32-80.

[3] Walther, J.B. and D'Addario, K.P., 2001. The impacts of emoticons on message interpretation in computer-mediated communication. Social science computer review, 19(3), pp.324-347.

[4] Calvo, R.A. and D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing, 1(1), pp.18-37.

[5] Dellaert, F., Polzin, T. and Waibel, A., 1996, October. Recognizing emotion in speech. In Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96 (Vol. 3, pp. 1970-1973). IEEE.

[6] Nicholson, J., Takahashi, K. and Nakatsu, R., 2000. Emotion recognition in speech using neural networks. Neural computing and applications, 9(4), pp.290-296.

[7] Ververidis, D. and Kotropoulos, C., 2005, July. Emotional speech classification using Gaussian mixture models and the sequential floating forward selection algorithm. In 2005 IEEE International Conference on Multimedia and Expo (pp. 1500-1503). IEEE.

[8] Nwe, T.L., Foo, S.W. and De Silva, L.C., 2003. Speech emotion recognition using hidden Markov models. Speech communication, 41(4), pp.603-623.

[9] Schuller, B., Rigoll, G. and Lang, M., 2004, May. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 1, pp. I-577). IEEE.

[10] Iliou, T. and Anagnostopoulos, C.N., 2010. Classification on speech emotion recognition-a comparative study. animation, 4, p.5.

[11] Han, Z., Lun, S. and Wang, J., 2012, March. A study on speech emotion recognition based on CCBC and neural network. In 2012 International Conference on Computer Science and Electronics Engineering (Vol. 2, pp. 144-147). IEEE.

[12] Zhang, T. and Wu, J., 2015, July. Speech emotion recognition with i-vector feature and RNN model. In 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP) (pp. 524-528). IEEE.

[13] Zhang, S., Zhang, S., Huang, T. and Gao, W., 2017. Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching. IEEE Transactions on Multimedia, 20(6), pp.1576-1590.

[14] Lalitha, S., Geyasruti, D., Narayanan, R. and Shravani, M., 2015. Emotion detection using MFCC and cepstrum features. Procedia Computer Science, 70, pp.29-35.

[15] Liu, G., He, W. and Jin, B., 2018, August. Feature Fusion of Speech Emotion Recognition Based on Deep Learning. In 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC) (pp. 193-197). IEEE.

[16] Wang, K., An, N., Li, B.N., Zhang, Y. and Li, L., 2015. Speech emotion recognition using Fourier parameters. IEEE Transactions on Affective Computing, 6(1), pp.69-75.

[17] Bandela, S.R. and Kumar, T.K., 2017, July. Stressed speech emotion recognition using feature fusion of teager energy operator and MFCC. In 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[18] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[19] Niu, Y., Zou, D., Niu, Y., He, Z. and Tan, H., 2017. A breakthrough in speech emotion recognition using deep retinal convolution neural networks. arXiv preprint arXiv:1707.09917.

[20] Qian, Y., Bi, M., Tan, T. and Yu, K., 2016. Very deep convolutional neural networks for noise robust speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(12), pp.2263-2276.

[21] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[22] Luengo, I., Navas, E. and Hernáez, I., 2010. Feature analysis and evaluation for automatic emotion identification in speech. IEEE Transactions on Multimedia, 12(6), pp.490-501.

[23] Jin, Q., Li, C., Chen, S. and Wu, H., 2015, April. Speech emotion recognition with acoustic and lexical features. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4749-4753). IEEE.

[24] Schuller, B., 2011. Recognizing affect from linguistic information in 3d continuous space. IEEE Transactions on Affective computing, 2(4), pp.192-205.

[25] Tawari, A. and Trivedi, M.M., 2010. Speech emotion analysis: Exploring the role of context. IEEE Transactions on multimedia, 12(6), pp.502-509.

[26] Cao, H., Savran, A., Verma, R. and Nenkova, A., 2015. Acoustic and lexical representations for affect prediction in spontaneous conversations. Computer speech and language, 29(1), pp.203-217.

[27] Petrushin, V.A., 2000. Emotion recognition in speech signal: experimental study, development, and application. In Sixth International Conference on Spoken Language Processing.

[28] Tato, R., Santos, R., Kompe, R. and Pardo, J.M., 2002. Emotional space improves emotion recognition. In Seventh International Conference on Spoken Language Processing.

[29] Tahon, M., Degottex, G. and Devillers, L., 2012. Usual voice quality features and glottal features for emotional valence detection. In Speech Prosody 2012.

[30] Huzaifah, M., 2017. Comparison of time-frequency representations for environmental sound classification using convolutional neural networks. arXiv preprint arXiv:1706.07156.

[31] Puterka, B. and Kacur, J., 2018, September. Time Window Analysis for Automatic Speech Emotion Recognition. In 2018 International Symposium ELMAR (pp. 143-146). IEEE.

[32] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F. and Weiss, B., 2005. A database of German emotional speech. In Ninth European Conference on Speech Communication and Technology.

[33] Woellmer, M., Kaiser, M., Eyben, F., Schuller, B. and Rigoll, G., 2013. LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework. Image and Vision Computing, 31(2), pp.153-163.

[34] Ioffe, S. and Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[35] Santurkar, S., Tsipras, D., Ilyas, A. and Madry, A., 2018. How does batch normalization help optimization?. In Advances in Neural Information Processing Systems (pp. 2483-2493).

[36] Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), pp.281-305.

[37] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G. and Wendemuth, A., 2009, November. Acoustic emotion recognition: A benchmark comparison of performances. In 2009 IEEE Workshop on Automatic Speech Recognition and Understanding (pp. 552-557). IEEE.

[38] Stuhlsatz, A., Meyer, C., Eyben, F., Zielke, T., Meier, G. and Schuller, B., 2011, May. Deep neural networks for acoustic emotion recognition: raising the benchmarks. In 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 5688-5691). IEEE.

[39] Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., André, E., Busso, C., Devillers, L.Y., Epps, J., Laukka, P., Narayanan, S.S. and Truong, K.P., 2015. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), pp.190-202.

The Development of Eyes Tracking System in Smartphone for Disabled Arm Person

Thidarat Pinthong
Department of Computer Business
Phetchaburi Rajabhat University, Thailand
Phetchaburi , Thailand
thidarat.pin@mail.pbru.ac.th

Mahasak Ketcham
Department of Information Technology Management
King Mongkut's University of Technology North Bangkok, Thailand
Bangkok, Thailand
mahasak.k@it.kmutnb.ac.th

***Abstract*—This research objectives were to design and develop algorithms for detecting and tracking human eye movements. The Researchers use image processing techniques combined with Haar Classifier and Region of interest for control eye. The system designed to control the operation of the eyes to perform the functions of controlling the smart phone. The researcher simulating cursor with microcontroller. The result is showed that system can detect face and eye precisely. The accuracy of this test is 78 percent.**

***Keywords*— smartphone, image processing technique, Haar-Like Model**

I. INTRODUCTION

Nowadays, the access to information and communication is essential in learning and development of various potentials. The technology is essential to use in people's daily lives. Everyone can access technology, such as using a mobile phone for chatting or using tablet applications. There are some people who are unable to use the phone and tablet, such as people with disabilities, patients, or people with physical disabilities. From the statistics of computer usage between 2009 and 2013, it was found that proportion of computer users increased by 35.0 percent (22.2 million people), Internet users increased by 28.9 percent (18.3 million people) and Mobile phone users increased by 73.3 percent (46.4 million people) show figure 1 .

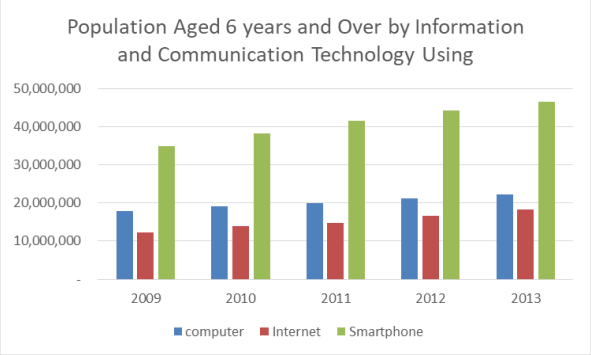


Figure.1. Population Aged 6 years and Over by Information and Communication Technology Using.

The number of people using information technology is increasing because the smartphones and tablets is work the same way as computers such as use to search data, communications, various transactions and the education. In the design of smartphones and tablets, it is small and light. The control smartphones and tablets use your fingers to control and open the application. Therefore it is difficult for people with disabilities or patients.

The researcher developed a device to assist patients or the disabled. To be able to use smartphones and tablets using eye detection.

II. RELATED LITERATURE

A. Haar-Like Model

Haar-like feature proposed by Viola and Jones . Haar-like feature extraction is a popular Technique implemented for finding the difference of the intensity between white and black areas.

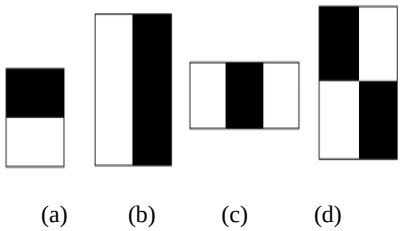


Figure 2. (a)Haar-like feature characteristics, (b) Two-rectangle features. (c) Three-rectangle features. (d) Four-rectangle features.

The area for searching Haar-like feature can be calculated by an integral image. The vertical and horizontal intensity summed can make an integral image. One, two, three and four can be used as the reference points to calculate the intensity in D area. The sum of the intensity of the A area equals to the one while the intensity of the A and B areas equals to the two. The sum of the intensity of the A and C areas equals to the three while the sum of the intensity of the A, B, C, and D areas equals to the four. The following equation (1) can be used to calculate the sum within D area.

$$D = 4 - 3 - 2 + 1 \quad (1)$$

Figure. 3 shows that five is the value of the sum of the intensity within D area which is used to find Haar-like feature. Two adjacent squares are derived from six reference points. Three and four adjacent squares are derived from eight and nine reference points. The computation time has the same

range for every square size which means the computation time is affected by a strong point of an integral image.

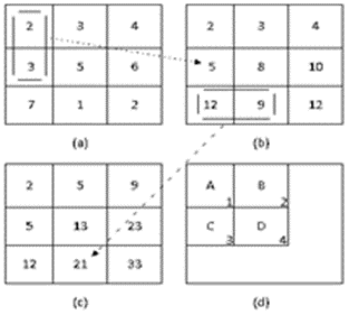


Figure 3. Shows an Integral Image technique, (a) shows Original image, (b) shows Sum of the vertical and horizontal intensities, (c) shows an Integral image and (d) shows Computation of the intensity of the D area.

B. Arduino Pro Micro

Microcontroller chip ATmega32U4 is connected via USB port. A micro-USB connector with Male Pin-Headers separated in two rows and had distance between each pin of 2.54 mm. was used to plug on breadboard. It is used to record commands to control the movement of the cursor via Arduino program.

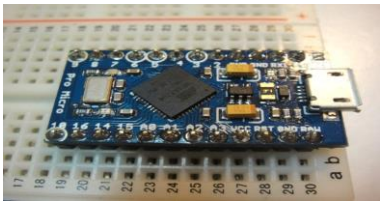


Figure. 4. The Arduino Pro Micro

C. Wireless Serial 4 Pin Bluetooth

It is a transmitter used to control devices between smartphone and Arduino. A simple principle for using Bluetooth is to pair devices. After connected, devices can be commanded via application on smartphones.

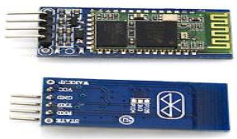


Figure. 5 . Wireless Serial 4 Pin Bluetooth

D. USB On-The-Go (OTG)

When connecting devices to USB for the first time requires driver installation. However, Android does not allow users to install driver, OTG is required for connection to display a cursor on devices. OTG is a wire that enables USB storage to work compatible with Android devices so that the stored data in USB storage can be retrieved. USB storage consists of 2 ports; 1) USB Port works as a socket to connect other devices such as a cursor or a keyboard and; 2) MicroUSB Port is a part to be connected to a smartphone or a tablet.

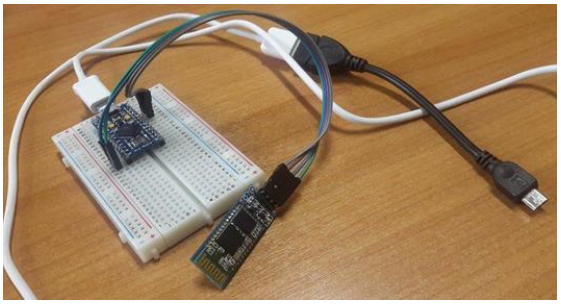


Figure. 6. Elements of a device.

III. PROPOSE METHODOLOGY

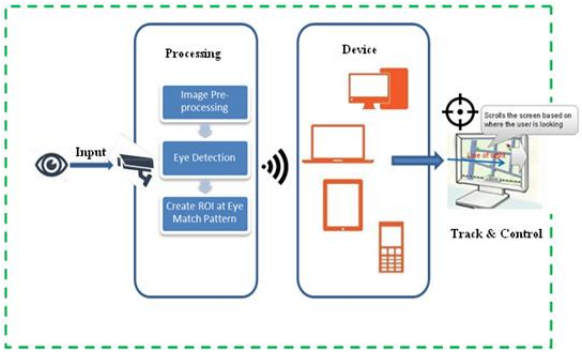


Figure 7 Show the process of system.

The researcher design system show that figure1. The system dived process follow as:

A. Image Pre-Processing.

When the system receives an eye image from the camera on the device, the image will be processed into digital data.

B. Eye Detection

The system will detect eye positions to identify eye position data.

C. Eye Math Pattern

The data is transferred to the microcontroller to process a command in accordance with eye movements to monitor such as left move, right move, bottom move and top move. The movement of the eyes is the movement of the mouse.

D. The Control of the mouse

The last stage is to display results on a smartphone and a tablet to control system such as open website .

E. The process of the eye movement detection system

The process of the eye movement detection system can show as follows:

Input Image from Webcam

Firstly, The camera captures the user's eyes using the commands from HighGUI when open Smartphone.

- Image Pre-processing

Image enhancement is essential for pre-processing process. The RGB color image requires conversion into Gray-scale image. The calculation is as follows.

$$\text{Gray} = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (2)$$

Where

Gray = gray intensity value is between 0 to 255
R = red intensity value is between 0 to 255
G = green intensity value is between 0 to 255
B = blue intensity value is between 0 to 255

- Detecting eye by Haar Classifier

AdaBoost Algorithm. Haar-like feature includes plenty of features yet a small number of the features are preferred. To improve features, AdaBoost algorithm is thus integrated. It discovers weak classifiers and combines them to become strong classifier.

- Let images are $(x_1, y_1), \ldots, (x_{n,}, y_n)$, where $y_i = 0, 1$ for negative and positive image, respectively.
- Define the first weight as the equation (3)

$$W_{1,i} = \frac{1}{2m}, \frac{1}{2l}, \quad (3)$$

Where $y_i = 0,1,$

m is the number of negative image.
l is the number of positive image.

- Let $t = 1, , T$

a) Normalize weights

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}} \quad (4)$$

Where w_t is a probability distribution.

b) Compute a weak classification function (h_j) and an error (ϵ_j) in each feature (j)

$$h_j(x_i) = \begin{cases} 1, \text{ for } p_j f_j(x) < p_j \theta_j, \\ 0, \text{otherwise}, \end{cases} \quad (5)$$

$$\epsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right|. \quad (6)$$

c) Select the weak classification with the lowest error.

d) Adjust weights.

$$w_{t+1,i} = w_{t,i} \beta_t^1 \quad (7)$$

Where

$e_i = 0$, When x_i is classified correctly

$e_i = 1$, Otherwise

And $$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}. \quad (8)$$

- The final classification is calculated as follows.

$$h(x) = \begin{cases} 1, \text{ if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=}^{T} \\ 0, \text{otherwise}, \end{cases} \quad (9)$$

Where

$$\alpha_t = \log \frac{1}{\beta_t}.$$

Cascade classification. When the strong classifier equation is obtained, it is divided by cascade classification into n stages. This means that it is not necessarily required the inputs to have computing in all stages. Nevertheless, recognition of the input passing through all stages will take place or it will be rejected**.**

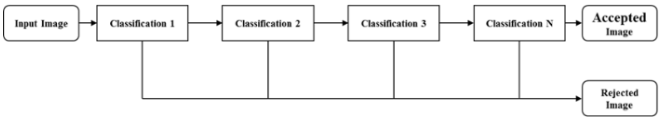


Figure 8. Shows Cascade classification

Eye Detection is shown in Figure 9.

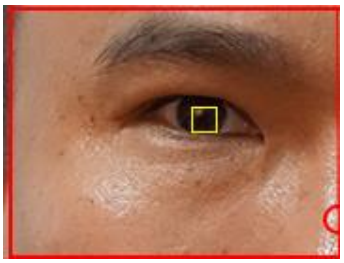


Figure.9. Eye position detection ClassifierCascade

- Detecting Eye Eage by ROI

Region of Interest (ROI) is employed for pupil detection which creates a circle around pupil position. The center point of the pupil (x,y) is identified by the circle function of Open CV and the radius (r) is determined referring to the circle equation as shown in equation (10).

$$\text{radius } (r) = \sqrt{(x - Center\_x)^2 + (y - Center\_y)^2} \quad (10)$$

Figure 10. ROI technique is used to detect the pupil position.

- Eye Tracking by template matching

Eye tracking based template matching is the method to track the directions of the eye movement for control the operation of the cursor. The example is if the eyes move to the left which the cursor will move to the left.

The equation below shows the calculation of an initial coordinates used for the movement.

$$(x_c, y_c) = \left(\frac{x_r + x_l}{2}\right), \left(\frac{y_r + y_l}{2}\right) \quad (11)$$

Where

$(x_c, y_c) =$ center coordinate between the eyes

$(x_r, y_r) =$ right eye coordinate

$(x_r, y_r) =$ right eye coordinate

When the initial coordinates are obtained, the distance and movement of the cursor is affected by the change of the position of any point to other positions. The equation below shows the calculation of the distance (12).

$$distance = \sqrt{(x_c - x'_c)^2 + (y_c - y'_c)^2} \quad (12)$$

Where

$x_c, y_c =$ center coordinate between the pupils

The equation (13), (14), (15), and (16) show the calculation of the new coordinate from the mouse movement.

$$\theta = \arctan\left(\frac{y_c - y'_c}{x_c - x'_c}\right) \quad (13)$$

$$x_{c\_update} = x_c + \frac{distance}{3} cos\theta \quad (14)$$

$$y_{c\_update} = y_c + \frac{distance}{3} sin\theta \quad (15)$$

$$(x_{c\_update}, y_{c\_update}) = \text{coordinate of mouse} \quad (16)$$

MotionEvent on android mobile works with Match template function of OpenCV.

- Face Detection for sent data in click mouse

CascadeClassifier library of OpenCV for detection is utilized to detect a region of face in the receive image or known as face detection. To detect a face, the eye detection is considered. Then, the Histogram values of the left eye and the right eye within the image are compared. The system will create a square box around the face region. But if unable to detect the face in the specified time the system will send commands to click once.

The process of the eye movement detection system can flowchart in figure 11.

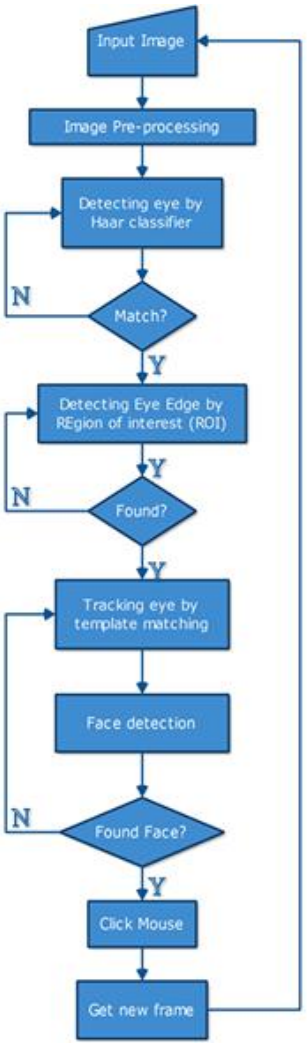


Figure 11 the receive picture and Eye detection.

IV. THE EXPERIMENTAL AND RESULT

We tested the system 100 times.

A. *Eye Detection Testing*

The image receiving and location tracking procedures while starting the program is processed by the eye detection testing.

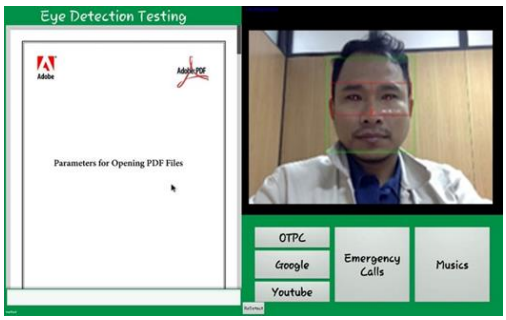


Figure 12 the receive picture and Eye detection.

B. *Test cursor control*

The test cursor control is test cursor movement to simulate the direction of the eye movement as follows.

- The left side is tracked when the eyes turn left.
- The right side is tracked when the eyes turn right.
- The topside is tracked when the eyes move to above.
- The bottom side is tracked when the eyes move to below.

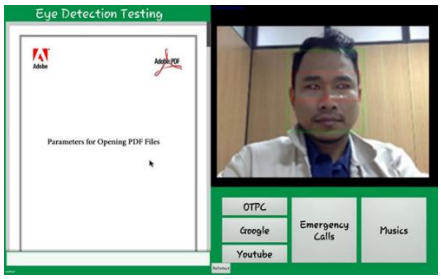


Figure 13. The picture shows the test, instructing the cursor to move to the left

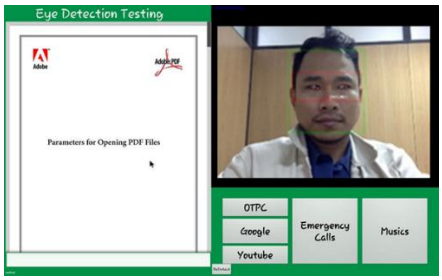


Figure 14. The picture shows the test, instructing the cursor to move to the right.

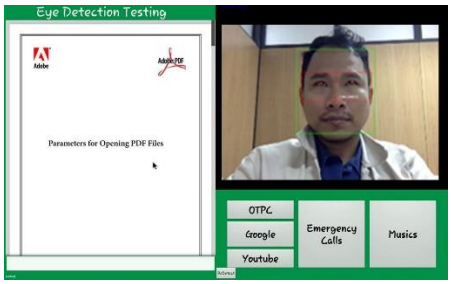


Figure 15. The picture shows the test, instructing the cursor to move to the above.

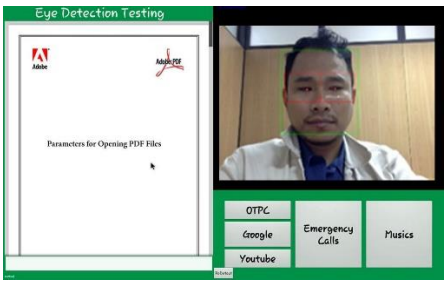


Figure 16. The picture shows the test, instructing the cursor to move to the below.

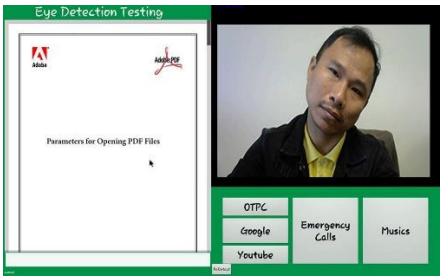


Figure 17. The picture show click test.

Two types of the experiment are identified including performance testing for the eye detection and eye tracking. Below is the experimental results.

TABLE I. EXPERIMENTAL RESULTS OF THE PERFORMANCE OF THE EYE DETECTION

	The position of the eye detection					Mean
	Straight	Top	Bottom	Left	Right	
Accuracy	0.96	0.87	0.89	0.96	0.95	0.93

TABLE II. EXPERIMENTAL RESULTS OF THE PERFORMANCE OF THE EYE TRACKING

	The position of the eye tracking					Mean
	Top	Bottom	Left	Right	Straigh t	
Accuracy	0.73	0.76	0.69	0.75	0.98	0.78

Based on the 100-time experiments, the accuracy of the eye detection and eye tracking performances can be described as follows.

The detection capability accuracy when looking straight on, looking up, looking down, turning to the left and turning to the right was 96, 87, 89, 96, 95 and 93 percent respectively. The eye tracking capability accuracy when looking up, looking down, turning to the left, turning to the right and looking straight on was 73, 76, 69, 98 and 78 percent respectively.

V. CONCLUSION

This paper is proposed to design eyes tracking system in smartphone for disabled arm person. The system can help people with disabilities or patients with disabilities to move their arms, hands, and fingers to have access to information technology. The results showed that the operation of the system is able to detect the face and find the eye position accurately, with an average accuracy of 93 percent. The cursor control has an average value of 78 percent.

REFERENCES

[1] C.Pornpanomchai, S.Rimdusit, P.Tanasap and C.Chaiyod. “Thai Herb Leaf Image Recognition System (THLIRS).” Nat. Sci., pp. 551-562, 2005.

[2] Chumuang N., Ketcham M., Sawatnatee A. (2019) Criminal Background Check Program with Fingerprint. In: Theeramunkong T.

et al. (eds) Advances in Intelligent Informatics, Smart Technology and Natural Language Processing. iSAI-NLP 2017. Advances in Intelligent Systems and Computing, vol 807. Springer, Cham.

[3] S. Thaiparnit, N. Khuadthong, N. Chumuang and M. Ketcham, "Tracking Vehicles System Based on License Plate Recognition," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 220-225. doi: 10.1109/ISCIT.2018.8588008

[4] S. Thaiparnit, N. Chumuang and M. Ketcham, "Weapon Detector System by Using X-ray Image Processing Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 214-219. doi: 10.1109/ISCIT.2018.8587853

[5] B. Narin, S. Buntan, N. Chumuang and M. Ketcham, "Crack on Eggshell Detection System Based on Image Processing Technique," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 1-6. doi: 10.1109/ISCIT.2018.8587980

[6] N. Chumuang, P. Chansuek, M. Ketcham, A. Silsanpisut, T. Ganokratanaa and P. Selarat, "Analysis of X-ray for locating the weapon in the vehicle by using scale-invariant features transform," 2017 Fourth Asian Conference on Defence Technology - Japan (ACDT), Tokyo, 2017, pp. 1-6. doi: 10.1109/ACDTJ.2017.8259599

[7] S. Suwannakhun, N. Chumuang and M. Ketcham, "Identification and Retrieval System by Using Face Detection," 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, 2018, pp. 294-298. doi: 10.1109/ISCIT.2018.8587856

[8] Lowe D.G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60(2): 91–110.

[9] Cecilia Di Ruberto and Lorenzo Putzu. “A Fast Leaf Recognition Algorithm based on SVM Classifier and High Dimensional Feature Vector.” 2014 International Conference on Computer Vision Theory and Applications (VISAPP), 2557.

[10] Xinhong Zhang and Fan Zhang. “Images Features Extraction of Tobacco Leaves.” 2008 Congress on Image and Signal Processing, 2551: 773-776

[11] T. Yingthawornsuk, N. Chumuang and M. Ketcham, "Automatic Thai Coin Calculation System by Using SIFT," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, 2017, pp. 418-423. doi: 10.1109/SITIS.2017.75

[12] W. Yimyam and M. Ketcham, “The automated parking fee calculation using license plate recognition system.” In 2017 International Conference on Digital Arts, Media and Technology (ICDAMT) (pp. 325-329). IEEE.

[13] M. Ketcham, W. Yimyam, and N. Chumuang, “ Segmentation of overlapping Isan Dhamma character on palm leaf manuscript’s with neural network”. In Recent Advances in Information and Communication Technology 2016 (pp. 55-65). Springer, Cham.

[14] S. Phatchuay, and W. Yimyam, “The System Vehicle of Application Detector for Categorize Type”. In 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 683-687). IEEE.

[15] W. Yimyam and M. Ketcham, “The Electroencephalography Signals Using Artificial Neural Network for Monitoring Fatigue System.” In Pacific Rim International Conference on Artificial Intelligence (pp. 160-169). Springer, Cham.

[16] T. Pinthong, W. Yimyam, N. Chumuang, and M. Ketcham, “License Plate Tracking Based on Template Matching Technique.” In 2018 18th International Symposium on Communications and Information Technologies (ISCIT) (pp. 299-303). IEEE.

[17] W. Yimyam and M. Ketcham, “Video Surveillance System Using IP Camera for Target Person Detection.” In 2018 18th International Symposium on Communications and Information Technologies (ISCIT) (pp. 176-179). IEEE.

[18] W. Yimyam and M. Ketcham, “The Grading Multiple Choice Tests System via Mobile Phone using Image Processing Technique.” International Journal of Emerging Technologies in Learning (iJET), 13(10), 260-269.

[19] W. Yimyam and M. Ketcham, “The System for Driver Fatigue Monitoring Using Decision Tree via Wireless Sensor Network for Intelligent Transport System.” International Journal of Online Engineering (iJOE), 14(10), 21-39.

[20] T. Pinthong and W. Yimyam, “The Model of Teenager’s Internet Usage Behavior Analysis Using Data Mining ”. In The Joint International Symposium on Artificial Intelligence and Natural Language Processing (pp. 196-203). Springer, Cham.

[21] W. Yimyam and M. Ketcham, “Eye Region Detection in Fatigue Monitoring for the Military Using AdaBoost Algorithm”. In International Symposium on Natural Language Processing (pp. 151-161). Springer, Cham.

Design and Implementation of A Smart Shopping Basket Based on IoT Technology

Sakorn Mekruksavanich
Department of Computer Engineering
School of Information and Communication Technology
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th

***Abstract*—In metro cities we can see people a huge rush at shopping malls on holidays and weekends. This becomes even more when there are huge offers and discounts. Today purchasing various products in supermarkets require a trolley or a basket. However, the product procurement represents a complex process. Each time customers have to carry the basket for getting the items and placing them in the basket and also they has to take care of expense computation. After shopping, most customers have to wait in a long queue for product scanning and bill payment. Therefore for dealing with this problems, developing a smart basket for shopping is presented in this research. Each and every product commonly contains barcode tag. The smart basket will consists of a barcode reader of mobile phone. When the customer scans and places any product in the basket, cost and the name of the product will be displayed on display mobile phone. The sum total cost of all the products will be added to the final bill, which will be stored in the micro controller memory. It will transfer the product information of the items placed in the basket using a transmitter to the main computer. Weight sensor system on the basket is used to validate in the shopping process accuracy. So, the proposed basket will support to avoid waiting in billing queue while constantly thinking about the budget.**

***Keywords*—smart shopping cart, Internet of Things (IoT), weight sensor, load cell**

I. Introduction

In the era of the Internet of Things (IoT), interactions among physical objects have become a reality. Everyday objects can now be equipped with computing power and communication functionalities, allowing objects everywhere to be connected. This has brought a new revolution in industrial, financial, and environmental systems, and triggered great challenges in data management, wireless communications, and real time decision making. Additionally, many security and privacy issues have emerged and lightweight cryptographic methods are in high demand to fit in with IoT applications. There has been a great deal of the IoT research on different applications, for example smart home, IoT-health system [1], wearable devices [2], [3], etc.

Today's world have a fast growing population with a wide range of demand from a variety of domains. With the development of people society, supermarket has been part of their daily life. Supermarkets are self-serving in nature, where customers use shopping carts or baskets in the store, search for the items they want to buy, place them into the containers and then proceed to the checkout counters. With little to none assistance, locating the shopping items in a big store can be very time-consuming, physically exhaustive and mentally frustrating. Due to the wide variety of commodities in the market, customers can buy anything they need.

However a coin has two sides, the more goods there are, the more time customers will spend on shopping. Customers may waste a lot of time on searching what they need. At billing stand the cashier takes one by one item from the cart for barcode scanning to make the bill due to which more time is going to be wasted and this results in long queue at billing places. So we thought of reducing the time consumption of the customer. Shopping mall is a place where people get their necessities ranging from food products, biscuits, clothing and electrical appliances. The numbers of large as well as small shopping malls has increased throughout the world due to increase in public demand and customer spending. Customers face difficulties about getting information about the products they wants to purchase and waste of time at the billing counter after shopping. Continuous improvement to the outdated billing system is needed to improve the quality of shopping practice to the customers. The program is intended to allow customers to feel the convenience that the IoT smart supermarket brought about to people's lives and understand what is IoT and how does it affect people's lives really and truly.

So in this paper, a smart shopping basket based on IoT technology is focused, which has not been well-studied in the past. Such a system, the customer scans the product using the barcode scanner from mobile phone application, information of product is shown and the bill is automatically updated. The weight sensor is placed under the cart which provides the total weight of the product in the cart, and the total weight of the products is also obtained from bar code scanner. If both of these values are matched, then the billing process is proceeded. Customers are also provided with option of removing the product from the basket where once again they need to scan the product. The updated list of the products is displayed on the mobile display. So, this system ensures that only scanned products are packed and billed.

The remainder of this work is arranged as follows: the recommended related works is offered in Section II, while Section III details the proposed shopping system to tackle the shopping problem. Section IV shows the shopping prototype.

Finally, section V provides the conclusion.

II. Related Work

The smart shopping system with IoT technology and artificial intelligence is recently very popular. Chandrasekar and Sangeetha [4] explain the centralized and automated billing system Using RFID and ZigBee communication. However the research in limited only to the RFID tags and it is Difficult to attach and detach RFID tags to each item. In paper [5], authors explain the shopping cart with three key components. First component is the shopping cart and the second key component is the communication system and the last component is the centralized system. But this paper is also limited to the RFID tags. Aryan [6] explains the localization algorithm based on hybrid sensor system with application to an active shopping cart. But the technology used in the paper is not recommended to the Indian marketing.

The implementation of the smart shopping cart using RFID tags were given in the paper [7], a system is urbanized that can be used in malls to solve the mentioned challenge. The system will be placed in all the carts. It consists of a RFID reader. All the products in the mall will be set with RFID tags. When a person puts any products in the trolley, its code will be detected and the price of those products will be stored in memory. As we put the products, the costs will get added to total bill. Thus the billing will be done in the trolley itself. Item name and its cost will be displayed on LCD. Also the products name and its cost can be announced using headset. Smart shopping cart using wireless communication is implemented in the paper [8].

III. The Proposed Shopping System

A. The System Design

The goals of the work is to provide pleasant shopping experience for the customers. Product information of all store merchandises and reducing delay billing time are mainly considered in shopping basket implementation. Customers face difficulties about waste of time at the billing counter after shopping and getting information about the products they wants to purchase. Continuous improvement to the automatic billing system is needed to improve the quality of shopping practice to the customers. So, the design of the proposed shopping basket can be shown in Fig. 1. With the smart shopping basket, the weighted sensor system and mobile application in mobile phone is proposed. The weighted sensor system is consists of load cell module which is connected to the load cell amplifier *HX711* module. The total weight of the product from the shopping basket is transfered to *ESP8266 NodeMCU* for sending weight data to record in the cloud server through Wi-Fi network. The shop customers scans the product using the application of mobile for showing the product information and its promotion and the bill is automatically updated. The weight sensor is placed under the cart which provides the total weight of the product in the cart, and the total weight of the products is also obtained from the mobile application. If both of these values are matched, then the billing process proceeds. The customer is also provided with option of removing the product from the basket where once again customer needs to scan the product. The updated list of the products is displayed on the mobile screen. So, this system ensures that only scanned products are packed and billed.

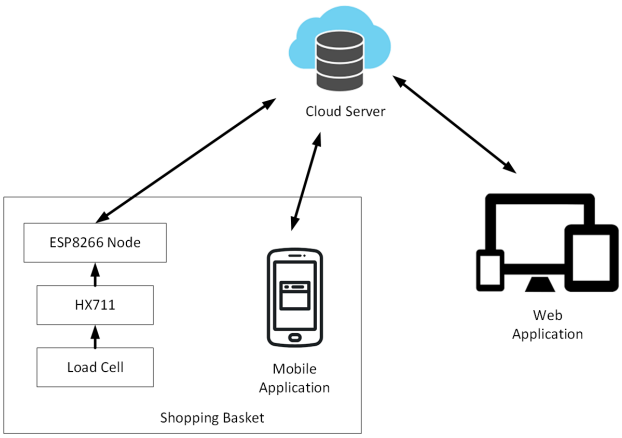


Fig. 1: The proposed structure of the shopping basket

The smart shopping basket integrates a shopping basket with barcode reader placed at the top of the shopping basket. It facilitates the customer to self-scan the barcode of the purchased products which each customer intends to purchase. If the customer wants to remove any products that can be done by scanning the product again while removing from the basket. A smart phone with an android application is used here. As soon as customer are logged in, a customer are assigned with a basket id which customers will be using throughout their shopping.

An android application facilitates customers to set the budget limit before they start their shopping. An android application makes note of all the scanned commodities of the particular basket and is linked with the Supermarket's backend database which contains details of the products such as price, stock amount, its promotion and etc. If the shopping amount reaches close enough to the budget limit or goes beyond the budget limit then the customer is notified through the same application. A customer can also increase the budget limit and set new budget limit once he is notified, or else he can generate the bill.

The scanned products are automatically billed in the android application, thereby significantly reducing turnaround time. The scanned products are also transmitted to the Shop's central billing program through a Wi-Fi network. By using this mechanism, the tedious work of scanning and billing every single product at the cash counter can be avoided. A weight sensor is also integrated with the shopping basket at the bottom of it. It is just to ensure if any product is added without getting scanned, so that the extra weight in the basket can be sensed. Finally, after the shopping and bill payment the bill is sent to the customer's registered E-mail through the same application mentioned.

The proposed structure of the shopping basket can be shown in Fig. 2. Three main components of the shopping basket

consist of:

1) *The control box*: this box contains *HX711* module and *ESP8266 NodeMCU* modules. Fig. 3 shows the connected networks between HX711 and ESP8266 NodeMCU module. When the shopping basket is online, weighted data of shopping merchandises will be sent from *HX711* to *ESP8266 NodeMCU*.
2) *The weight sensor module*: the module is on the button of the shopping basket. All weight data on its is sent to the control box part for showing the useful information of products and processing the billing information.
3) *The mobile phone holder*: the mobile phone is holden with this part. It is used to scan the barcode of each products and shows their information. Moreover, it also provides shopping amount and budget of their customers.

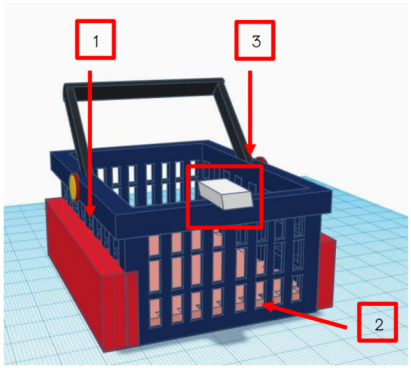


Fig. 2: The proposed structure of the shopping basket

IV. The System Prototype

According to the design of shopping basket in the previous section, the shopping basket is implemented as the prototype basket in this section. Fig. 4 shows the connecting among the *Load Cell*, *HX711* and *ESP8266 NodeMCU*. After that, the experiment is conducted by scaling some commodities and the weighted results is shown in Fig. 5.

The top-view completed prototype of shopping basket is shown in Fig. 6. The weight sensor module is on the bottom of the basket. The control box is beside the basket. Fig. 7a and Fig. 7b show the login and budget on the mobile application, respectively. The mobile application performing a commodity scanning can be showed in Fig. 8.

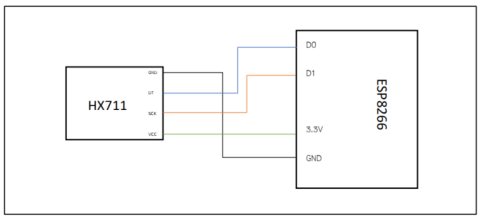


Fig. 3: The connected networks between HX711 and ESP8266 module

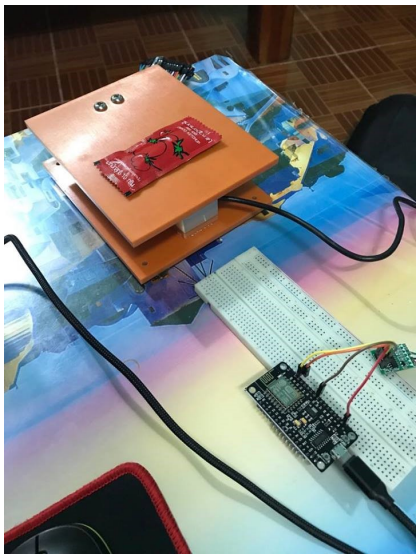


Fig. 4: The connecting among parts of the shopping basket implementation

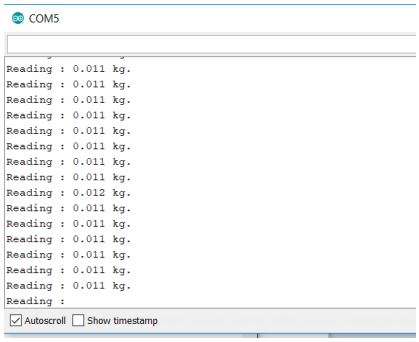
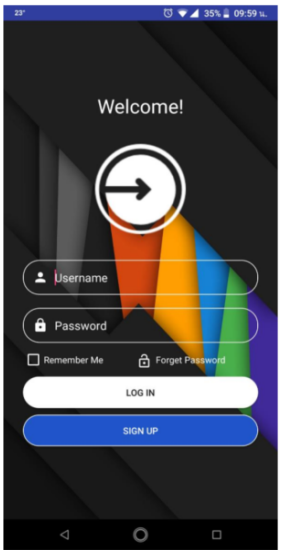
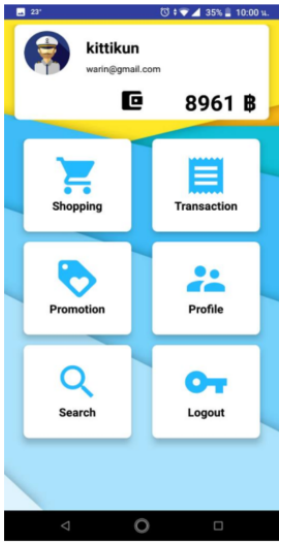


Fig. 5: An example of received weight data of some commodities in research experiments

Fig. 6: The proposed structure of the shopping basket



(a) Login display



(b) A tiger

Fig. 7: Pictures of animals

Fig. 8: The proposed structure of the shopping basket

V. Conclusions and Future Works

This kind of system and application will provide a way for smart shopping. It will be a great way to handle customer inconvenience that are faced during shopping, especially during the festival seasons. Customers, simply by using their own android phone application, can manage everything within the shopping environment in this work. Since the products are scanned quickly as soon as they are placed into the shopping basket and paperless bills are generated and sent to the customers registered E-mail it saves time of waiting in a long queue at the cash counter. Thus it shows a high potential of IoT system to be integrated in supermarkets or shopping malls.

Acknowledgment

This research was supported in part by the School of Information and Communication Technology, University of Phayao, Thailand.

References

[1] P. Castillejo, J. Martinez, J. Rodriguez-Molina, and A. Cuerva, "Integration of wearable devices in a wireless sensor network for an e-health application," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 38–49, August 2013.

[2] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Human activity recognition using triaxial acceleration data from smartphone and ensemble learning," in *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, Dec 2017, pp. 408–412.

[3] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "Smartwatch-based sitting detection with human activity recognition for office workers syndrome," in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, Feb 2018, pp. 160–164.

[4] P. Chandrasekar and T. Sangeetha, "Smart shopping cart with automatic billing system through rfid and zigbee," *2014 International Conference on Information Communication and Embedded Systems, ICICES 2014*, 02 2015.

[5] A. Kumar, A. Gupta, S. Balamurugan, S. Balaji, and R. Marimuthu, "Smart shopping cart," in *2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)*, Aug 2017, pp. 1–4.

[6] P. Aryan, "Smart shopping cart with automatic billing system through rfid and bluetooth," *International Journal Of Emerging Technology and Computer Science*, vol. 1, no. 2, 2016.

[7] J. Suryaprasad, B. O. P. Kumar, D. Roopa, and A. K. Arjun, "A novel low-cost intelligent shopping cart," in *2011 IEEE 2nd International Conference on Networked Embedded Systems for Enterprise Applications*, Dec 2011, pp. 1–4.

[8] B. Wu, S. Yao, L. Hou, P. Chang, W. J. Tseng, C. Huang, Y. Chen, and P. Yang, "Intelligent shopping assistant system," in *2016 International Automatic Control Conference (CACS)*, Nov 2016, pp. 236–241.

Short-circuit and Over Current Notification in Sub-transmission Line by Messege Cellular Network

Sumate Lipirodjanapong
Electrical Technology,
Faculty of Industrial Technology,
Rajabhat Muban Chombueng University
Ratchaburi, Thailand
lsumate@hotmail.com

Pumpat Uthaisiritanon
Provincial Electricity Authority
Phetchaburi, Thailand
lekactivo@gmail.com

Pitipol Duangjinda
Electrical Technology,
Faculty of Industrial Technology
Rajabhat Muban Chombueng University
Ratchaburi, Thailand
ti.pone2531@gmail.com

***Abstract*— This research were studied the preliminary system of short circuit current detection in sub-transmission line. It can utilize rapidly working staff for the Province Electrical Authority (PEA) and PEA's control center to limit the investigation area of contingent circuits. So the short-circuit or over current is the problem in electrical transmission line network, what to effected the damageable electrical protection devices such as fuse etc. it's cause to occur the back-out event. It need several time to repair. For this research have proposed the automatic short-circuit current warning system via a mobile phone network by microcontroller processing. This system will be monitor the current value as an indicator in real time short circuit events using the electrical current measurement device 1 set per 1 phase. When the current value from one of the values exceeds the specified value. The system detector will send short message service (SMS) that measured current value to the mobile phone the specified number. It can check the measured current values at any time by sending SMS to the device. The device will send the current value of the 3 phases measured at that moment back. Finally, the experiment results in 3 area as Petchburi's PEA, PEA-substation and Big-C department store can detect short-circuit current in sub-transmission line any area have accuracy about 98%. It help to repair sub-transmission line rapidly.**

Keywords*— *short circuit current detection, sub-transmission line, short message notification

I. INTRODUCTION

Electric power transmission via transmission line to consumers continuously with electrical quality are important to the reliability and stability of the power distribution system, depends on the design and maintenance of the electrical system [1]. Occurrence of power failure will cause a lot of damage. Electrical power transmission that must cover all areas, Therefore in some electrical power transmission line have to distribute electricity at a long distance and there are a lot of junction points going through the alleyways. When a short circuit occurs in sub transmission line, it is difficult and take a long time to find the short circuit location. Because today Provincial Electricity Authority still using random sampling methods to find the location of the event, by driving a vehicle to find them. This research studied the introduction of short circuit current detection systems can responsibility to the staff and control center to limit the area of detection of the incident point, to be able to find the short circuit location quickly and accurately, make the solution quickly and restore the electrical system faster.

II. BASIC THEORIES

A. Basic knowledge of short-circuit current

The Standard IEC 60909 divides the short-circuit current into two types [2].

- Near-to-generator short-circuit

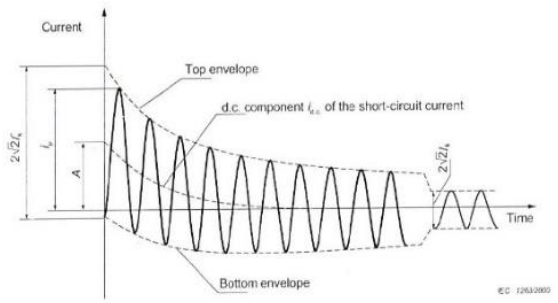


Fig. 1. Near-to-generator short-circuit

- Far-from-generator short-circuit

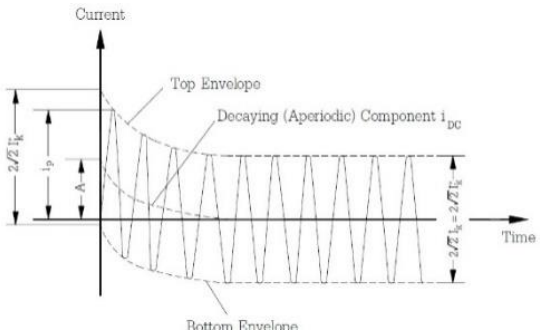


Fig. 2. Far-from-generator short-circuit

B. Calculation of short-circuit current

Determining various of short-circuit current can be calculated from the equation as in Table 1 [2], [3].

TABLE I. EQUATION FOR SHORT-CIRCUIT CURRENT

Fault type	*Equation*
3 - ph	$I_k^{"} = \frac{c(kV)}{\sqrt{3}Z_1}$
1 - ph to ground	$I_k^{"} = \frac{\sqrt{3}c(kV)}{Z_0 + Z_1 + Z_2}$
2 - ph to ground	$I_k^{"} = \frac{\sqrt{3}c(kV)}{Z_1 + Z_0 + Z_0(Z_1 / Z_2)}$
Line - to - line	$I_k^{"} = \frac{c(kV)}{Z_1 + Z_2}$

From the study of the difference of short-circuit current according to IEC60909 and IEEE551 and impact on the overcurrent protection system in substation [4]. The short-circuit current has a very high value. (more than 1,000 A) [5], [6]

III. SCOPE OF RESEARCH

The short-circuit current is higher than 1000 A, it difficult to generated a short-circuit current to test the operation of the equipment and do not know when the short-circuit current will be occur. So the equipment will be tested by reference from normal load range compare with peak load range. When the measured current is greater than the set. The device must send SMS the measured current value to the specified mobile phone immediately, and must be able to send SMS the measured current to the specified mobile phone every time when receiving a request message.

IV. PRINCIPLES OF CREATING A SHORT-CIRCUIT CURRENT DETECTION SYSTEM

The short-circuit current detection is controlled by microcontroller for transmitting short-circuit current, and able to report the results via the mobile phone network. The process of creating the system are as follows.

A. Build a current measurement device

Using current transformers installed at the 3 phase power transmission line to measure the current in each phase, and send the value through the wireless module NRF24L01 in the 2.4 GHz band, controlling by Arduino board. This set will be powered by a 9 V battery as shown in Fig. 3.

Fig. 3. Short-circuit current measurement set

The operation of the current measurement device is based on the flow chart as shown in Fig. 4.

From the flow chart of the current measurement set, start when CT sends the output voltage, the Arduino Nano will process it into the current that occurs in the electrical wire at that time, then combine letters A, B or C with current values as configured, such as A20, B20 or C20, the NRF4L01 module sends that message through the 2.4 GHz frequency band.

Current measurement device will be created 3 sets to install at phase A, B and C, which set which phase is installed will be programmed to enter the text of that phase into the messages to be send out.

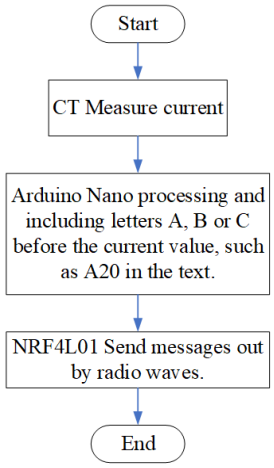


Fig. 4. Flow chart shows the operation of the current measurement device

B. Build a main control box

Build a main control box installed at the bottom of the electricity pole to receive all 3-phases current values through the wireless module NRF24L01 in the 2.4 GHz frequency band, processed with the Aduino board. If any phase current value exceeds the specified value, program will send notify SMS to specified mobile phone numbers immediately. When the device receives the message "chk", it will send all 3 phases current values to the specified mobile phone number as well. This set will be powered by a rechargeable battery. Because when a short-circuit occurs, it will cause the electric current at the location of this device to be set off as well, therefore a backup power source is needed to supply power to the device.

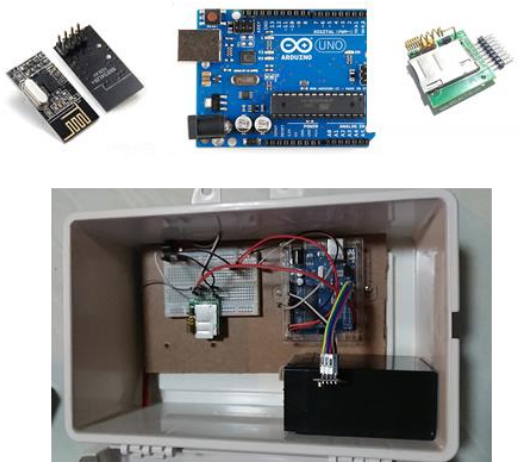


Fig. 5. Main control box

The operation of the main control box is based on the flow chart as shown in Fig. 6.

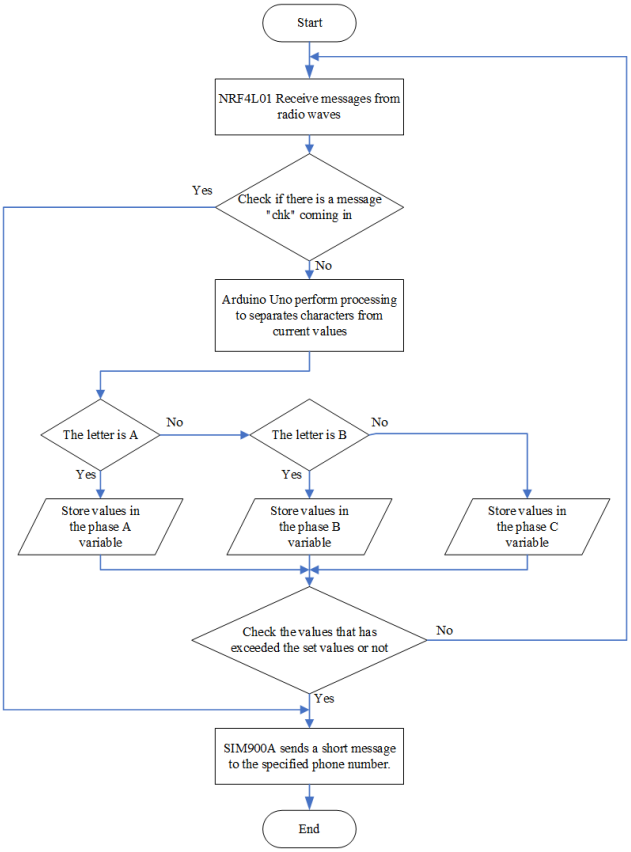


Fig. 6. Flow chart shows the operation of the main control box

From the flow chart of the main control box set, start by receiving the current values from the current measurement device and extracting the first character to specify which phase it came from, then take the values stored in the variable of that phase, then compare the current values with the set values, if the values exceeds the set values program will send the message "Fault !!!" and how much current at which phase to specified mobile phone immediately, if the current values does not exceed the set but the device receives the message "chk", it will send all 3 phases current values to the specified mobile phone as well.

C. *Functional test*

Bring the current measuring device and main control box to install on the electricity pole at various points to test the measurement of current value, test the data transmission from current measuring device to the main control box, test communication via mobile phone network by SMS, and test sending SMS automatically when the current value exceeds the specified value.

Fig. 7. Installing a short-circuit warning device

V. RESULT

A. *Efficiency of the short-circuit current measuring device*

Tested by installing a current measuring device on the power transmission line, as in Fig. 6. Which compares the current value obtained from the device with the current measuring instrument brand fluke model 336. Define test locations in 3 areas, Phetchaburi Provincial Electricity Authority 13.098211"N, 99.943979"E, clear weather, temperature 32°C , Phetchaburi sub station 13.089290 "N, 99.942135"E, clear weather, temperature 32°C and Big C department store 13.079514 "N, 99.949211"E, clear weather, temperature 34°C

Fig. 8. Current measurement with a set of devices compared to current measuring instrument on the power transmission line

Comparative results of current measurement as in Table II.

TABLE II. CURRENT MEASUREMENT RESULTS

Location	***Current values from the test device***	***Current values from measuring instrument (Fluke 336)***	***% Error***
Phetchaburi Provincial Electricity Authority	21	21.32	1.50
Phetchaburi sub station	159	158.53	0.29
Big C department store	19	18.65	1.88

From Table II, show that the current measurement device measure current values 21A, 159A and 19A respectively, compare with electrical measuring instruments that meet EN / IEC 1010-2-032 current measurement range 0-600A accuracy 2% ± 5 counts (10-100 Hz), 6% ± 5 counts (100 - 400 Hz) true RMS measure the current on the power transmission lines at all 3 locations 21.32A, 158.53A and 18.65A. respectively. From this result can calculate the average error percentage is 1.22%

B. Wireless communication between current measurement devices with main control box

Tested by measuring the distance between current measurement devices and the main control box in which the current measurement devices are installed on power distribution line at head of the electricity pole and the main control box is installed at the middle of the same pole as shown in Fig. 9. The first point has a distance of 2 meters and then moves further down along the pole for 1 meter at a time. Check from the LED display screen whether data can be received or not, testing at Phetchaburi substation. Explained in Table III.

Fig. 9. Measuring the distance between the current measurement and the main control box

TABLE III. THE RESULT OF TESTING THE DISTANCE BETWEEN THE CURRENT MEASUREMENT DEVICE WITH MAIN CONTROL BOX

Distance (meter)	***Can receive data***	***Cannot receive data***
2	√	
3	√	
4	√	
5		√

The result from Table III shows that the maximum distance that the main control box will receive data from the current measurement devices installed on the power transmission lines is 4 meters. If installed at a greater distance will not be able to receive data.

C. Sending SMS via mobile phone to the device to request data

Tested by sending a short message "chk" from 3 mobile phones 10 times per device to the telephone number of the device, a total of 30 times. Received a response message from the device complete all 30 times, equivalent to 100 percent.

TABLE IV. THE RESULT OF TESTING SENDING SHORT MESSAGES VIA MOBILE PHONE TO THE DEVICE TO REQUEST DATA

Mobile phone network	***Number of SMS sent from mobile phones to Equipment set (time)***	***Number of SMS received from the device (time)***
DTAC	10	10
TRUEMOVE H	10	10
AIS	10	10
TOTAL	**30**	**30**

D. Sending SMS when the current value is greater than the specified value automatically

Tested by installing the equipment set on power distribution line at the Phetchaburi Provincial Electricity Authority around 7:00 am, clear weather, set up the device to 5 A. At 08:31, the device sent a notification message that there was a 6 A current to the mobile phone.

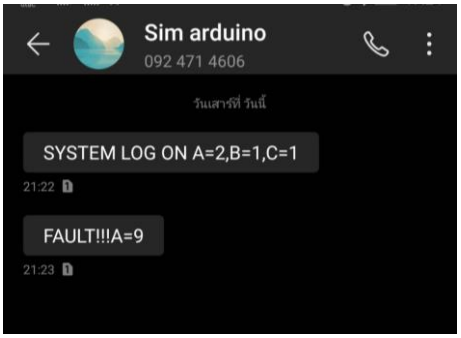


Fig. 10. Show received message

E. Battery life in the equipment set

Alkaline batteries in the current measurement devices has a maximum lifespan up to 276 hours.

Fig. 11. Show alkaline batteries in the current measurement device

Sealed rechargeable battery type 12 V 7.5 AH in the main control box has a maximum lifespan up to 46 hours per 1 full charge.

Fig. 12. Show a Sealed rechargeable battery type 12 V 7.5 AH in the main control box

VI. Cost economic analysis

Considering the investment in the set of built equipment, as in Fig. 13, the price is approximately 6,000 baht. When compared with the Fault indicator device, as in Fig. 14, which is currently in use, that will have a flashing warning lights when the short-circuit current occurs, the price is approximately 36,000 baht per 1 set. (1 set has 3 pieces)

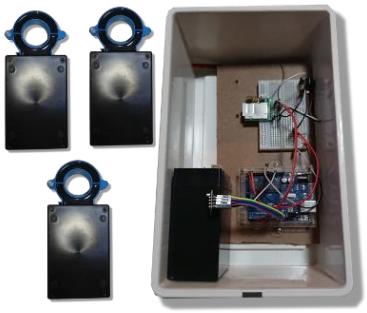


Fig. 13. Show short-circuit current detection and automatically notify to mobile phone set.

Fig. 14. Show Fault Indicator device that PEA currently uses

VII. Summary and discussion of results

Short-circuit current detection and automatically notify to mobile phone system can measure current with the highest percentage error in the test at 1.22%. When the measured current value is greater than the set value, the system will send an SMS to the specified mobile phone numbers automatically, and can also send short messages of current value back when receiving a short message requesting. The current measurement devices and the main control box must be installed not more than 4 meters apart.

References

[1] R. Billinton and R. N. Allan, "Reliability Evaluation of Power Systems," 2nd Ed., Plenum Press and Pitman Publishing Ltd., USA, 1996.

[2] IEC60909 Standard, Short-circuit currents in three-phase a.c. systems. 2001.

[3] IEEE Standard 551, Calculating Short-Circuit Currents in Industrial and Commercial Power systems, 2006.

[4] Worawet Pankrajang, "The study of the difference of the short circuit current according IEC60909 and IEEE551 ,Impact on the overcurrent protection system in substations."

[5] Feng Du, Weigang Chen, Yue Zhuo and Michael Anheuser., "A New Method of Early Short Circuit DetectionJournal of Power and Energy Engineering," 2014, 2, 432-437.

[6] T. Mützel, F. Berger and M. Anheuser, " New Algorithm for Electronic Short-Circuit Detection." 52nd IEEE Holm Conference on Electrical Contacts, Montreal, 42-47, 2006.

[7] Paitoon Ngewtang, "Water level alarm system via mobile phone SMS," Pibulsongkram Rajabhat University., Thailand, 2016.

[8] Aphichai Wonpuan, "Transformer load monitoring systems," King mongkut's university of technology north bangkok. Thailand, 2013.

e-Learning Recommendation Model Based on Multiple Intelligence

Thaksaorn Jommanop
Master of Modern Information Technology Management
School of Information and Communication Technology
University of Phayao, Phayao, Thailand
j.thaksaorn@gmail.com

Sakorn Mekruksavanich
Department of Computer Engineering
School of Information and Communication Technology
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th

***Abstract*—E-learning has offered numerous advantages such as flexibility, remote operability, cost effectiveness, simplicity, consistency and many more. The utilization of smart tools and technologies has provided easy and convenient education in an effective way without barrier of time and place. The purposes of this research were to 1) to synthesize a conceptual framework of collaborative and adaptive e-Learning for student with different multiple intelligences, 2) to develop the adaptive e-Learning, 3) to study the students'learning achievement after using multiple intelligence models and 4) to study the students satisfaction after using multiple intelligence models in innovation and information technology in education subject for student in school of teacher education, Phayao university.**

***Keywords*—collaborative learning, adaptive e-Learning, multiple intelligence, instructional technology**

I. Introduction

In traditional face-to-face teaching, teachers determine the environment of all classes and learning methods. The current system of learning and teaching has evolved from the original, for example the e-learning system, distance learning, or etc. The condition of learning and teaching today provides learners with opportunities to learn by themselves, teachers are only a guide. Generally, each learner has different abilities or aptitudes. Therefore, if learners are not equally skilled in the class, the learning performance becomes unbalanced and some learners might fall behind.

Past research has offered various types of learning and teaching methods, such as learning and teaching by brainstorming [1]. Learning and teaching by brainstorming means that learners will participate in problem solving in the class. If learners brainstorm, the more opportunity to solve problems arises. The advantage of learning and teaching in this way helps groups of learners to solve problems faster than normal learning and teaching. The limitation of this method is that learners do not gain learning contents according to the individual's aptitude.

For the research learning and teaching is a project base learning method (Project Base Learning) which focuses on every learner. From this learning and teaching method, learners can learn from the project at their own speed. The project is divided into appropriate groups of learners and learning contents. But the limitation occurs when learners are given contents that do not match their capabilities.

Researches [2] also use problem base learning (Problem Base Learning). This learning and teaching method focuses on learners by defining problems and solving them. This learning and teaching method is suitable for learners who are experts in analysis. But the restriction remains the same that learners do not get the lesson according to the individuals aptitude. Research [3] focuses on learning and teaching by emphasizing on learning activities base on Multiple Intelligence. This research focused on learning activities. But again was limited as individuals aptitudes were not catered for. In addition, [4] proposed learning activities by using Multiple Intelligences in the environment of e-learning. But the research focused on only learning activities and failed to consider learners with problems during study.

In addition, a group of researchers have used this technology to develop games with multimedia in order to help learning and teaching. Research [5] are learning and teaching methods which use games to display learning material. This learning and teaching method is suitable for learners who enjoy playing games because when they play the game, they are learning at the same time. However, this application is suitable for only a small group of learners.

From the research mentioned above, learning and teaching by employing types of project base learning, problem base learning, activity base learning, game base learning and student center learning were found to be limited. In each method, learners are given learning contents that do not match individual aptitudes. Thus, this research presents the development e-Learning activities based on multiple intelligences for supporting critical thinking. Rule bases are rules which come from the data analysis of learners. Subjects of secondary 1 education in Demonstration school, Phayao university is used to conduct the research experiments.

This research consists of 4 parts; First section, introduction of previous work. Second section covers a presentation of the background for this research. Third section, describes the concept of framework. Lastly, the proposed design of the rule base for learner recommendation.

II. Background and Previous Work

A. e-Learning

e-Learning is basically any educational related activities via internet, network or standalone computer or rather in today's smart world; it is learning activity available on electronic medium at any place, any time for any person on any smart internet enabled device [6]. e-Learning is mostly associated with activities involving computers and interactive networks simultaneously. The computer does not need to be the central element of the activity or provide learning content. However, the computer and the network must hold a significant involvement in the learning activity. e-Learning comprises of several sub types such as web based, online and distance learning activities.

- Web-based learning is associated with learning materials delivered in a Web browser, including when the materials are packaged on CD-ROM or other media.
- Online learning is associated with content readily accessible on a computer. The content may be on the Web or the Internet, or simply installed on a CD-ROM or the computer hard disk.
- Distance learning involves interaction at a distance between instructor and learners, and enables timely instructor reaction to learners. Simply posting or broadcasting learning materials to learners is not distance learning. Instructors must be involved in receiving feedback from learners.

B. Multiple Intelligences

The multiple intelligences is identified nine distinct intelligences. According to this theory which emerged from cognitive research, multiple intelligences document the extent to which students possess different kinds of minds and therefore learn, remember, perform, and understand in different ways.

The Multiple Intelligences theory [7] is divided into 3 groups. 1) Analytic group, this group focuses on analysis and the thinking processes. The analytic group consists of 3 parts; Logical-mathematic intelligence, Musical intelligence, and Naturalist intelligence. 2) Introspective group, this group focuses on imagination and understanding. The introspective group consists of 3 parts; Intrapersonal intelligence, spatial intelligence, and Existential intelligence. 3) Interactive group, this group focuses on communication and interactive. The interactive group consists of 3 parts; Linguistic intelligence, Interpersonal intelligence, and Kinesthetic intelligence. Fig. 1 shows the model of multiple intelligences.

C. Multiple Intelligences-Based Activities in e-Learning

According to Gardner and Hatch [8], there are three key elements that determine a person's intelligence:

1) The ability to create a service or product that will be valued in the person's society of culture.
2) A skill set that allows the person to solve real world problems that they may encounter in life.

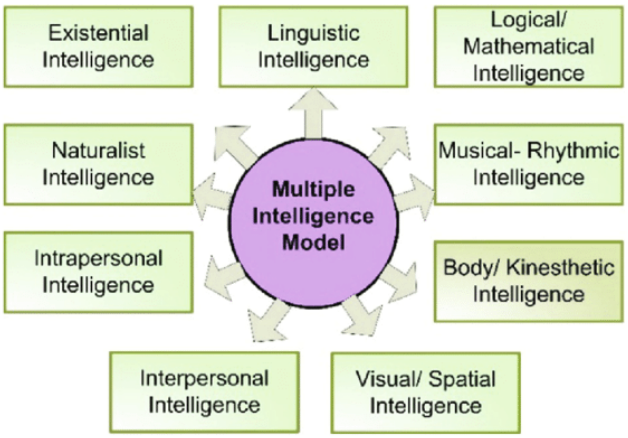


Fig. 1: Multiple intelligences model

3) The ability to potentially create new solutions for problems or to utilize existing solutions. This typically involves the acquisition of new knowledge.

Multiple intelligences theory can be implemented in e-Learning based upon nine multiple intelligences. Each category and which multiple intelligences-based activities can be utilized during instructional design to create the best possible e-Learning experience for the learners as following:

1) *Musical (or Rhythmic) Intelligence.* This intelligence involves the capacity to think and learn in terms of music and rhythm, and to recognize and hear patterns. An activity that would appeal to this type of intelligence is a lesson that includes music or sound, such as a multimedia presentation. Try to use music that emphasizes the subject matter and creates a more immersive experience for the learner. People who demonstrate a high degree of musical intelligence may be ideally suited for musical professions, such as composing or playing an instrument.
2) *Linguistic Intelligence.* This intelligence is associated with expression through language. These people tend to be able to eloquently convey their thoughts and to understand the words of others. Writers and speakers typically display a high degree of this sort of intelligence. Any activities that include discussion, such as online forums or group-based scenarios, are ideal for individuals who lean more toward linguistic intelligence.
3) *Mathematical (or Logical) Intelligence.* This involves the ability to identify principles or structures within a system. This intelligence is often associated with the logic or the manipulation of numbers. Activities ideally suited for this intelligence may include diagrams, charts, or tables. Critical thinking scenarios are also useful with this group. Accountants and researchers often have a high level of mathematical intelligence.
4) *Kinesthetic (or Bodily) Intelligence.* Body Intelligence involves the use of a person's entire body to figure out solutions or to create something. People who demon-

strate a high degree of kinesthetic intelligence may be ideally suited for performing arts professions, such as dancing, or careers that require an innate knowledge of one's own body, such as a doctor or athlete. Activities that are best suited for this sort of intelligence include games that involve hand-eye-coordination or interactive scenarios that require physical involvement. The thing to keep in mind about this group is that they are best able to learn when muscular movement is involved. So, include activities that require movement and physical response.

5) *Spatial Intelligence.* This intelligence pertains to a keen sense of space and how one can navigate those spaces. Activities that involve flow charts and graphics are ideal for this intelligence group, as well as games or multimedia that is visually appealing. Architects, pilots, and sailors often have a high degree of spatial intelligence.
6) *Intrapersonal Intelligence.* This involves an in depth understanding of oneself, such as what you can accomplish and how you react to certain situations. As such, individuals with high intrapersonal intelligence often have a sense of what they should avoid and what they want to achieve in their lives. Professors and philosophers often possess high degrees of intrapersonal intelligence. Activities such as collaborative learning exercises (online forums) and chat programs enable intrapersonal intelligence learners to help others and to share experiences and ideas. This category responds well, first and foremost, to activities, which require introspection.
7) *Interpersonal Intelligence.*This is the capacity to understand and learn from others. People who demonstrate a high degree of intrapersonal intelligence may be ideally suited for service professions, such as teaching or politics. Those who identify more with this category of intelligence may benefit from group discussion activities and in depth questions that make them fully explore the topic. What's important to remember about interpersonal intelligence is that these individuals are sensitive to others' moods and feelings. They work well in-group settings and are often able to learn more effectively when collaborating.
8) *Naturalist Intelligence.* This intelligence involves the capacity to differentiate between living organisms and to view the connection between all natural things. People with a high degree of naturalist intelligence usually have a close bond with nature. Botany and biology are two career fields that closely identify with this sort of intelligence. Activities that involve classification or organization appeal to these individuals.
9) *Existential Intelligence.* This particular intelligence was added later by Gardner, and is not commonly associated with learning environments, as it is geared more toward spiritual and philosophical views. For example, someone who has a high degree of existentialist intelligence may have a tendency to pose questions about life's purpose or death.

III. The Proposed Methodology of Multiple Intelligences in e-Learning

This research proposes the development e-Learning activities based on multiple intelligences for supporting critical thinking. So, the research focuses on the principle of mathematical/logical intelligence part of multiple intelligences to design a rule base for the e-learning recommendation system. Critical thinking is the ability to think clearly and rationally, understanding the logical connection between ideas. This ability is the most important one of mathematical/logical intelligence of multiple intelligences that most student's intelligence should be developed. The design framework of e-Learning in the framework which is based on the multiple intelligences principle is shown in Fig. 2.

A. Critical Thinking Module

The critical thinking module is the module that keeps the learning styles of learners according to their aptitude of critical thinking in mathematical/logical intelligence.

B. Recommendation Module

The recommendation module is the module that introduces learners to the type of learning content learners should learn such as Interactive content, Introspective content, or Analytic content.

C. LMS Module

The Learning Management System (LMS) module is the module that acts as an intermediary between e-Learning systems and learners. As parts of the LMS module are connected to most of the other modules in order to use it in learning and teaching.

D. Learning Content Module

The Learning Content Module is the module that stores learning contents which comes from nine multiple intelligences and divided into 3 groups:

1) *Analytic content* used with learners who have analytical, logical and mathematical aptitude. The critical thinking analysis for supporting critical thinking from Critical Thinking Module is also included in these learning contents.
2) *Introspective content* used with learners who have the imaginative and artistic aptitude.
3) *Interactive content* used with learners who have the skills of interactive and communication aptitude to others.

IV. Activities Design for Critical Thinking Analysis

The section focuses on development e-Learning activities in Critical Thinking Module. These activities base on mathematical/logical intelligence for supporting critical thinking intelligence for student learners. The development framework of e-Learning activities for supporting critical thinking skill is

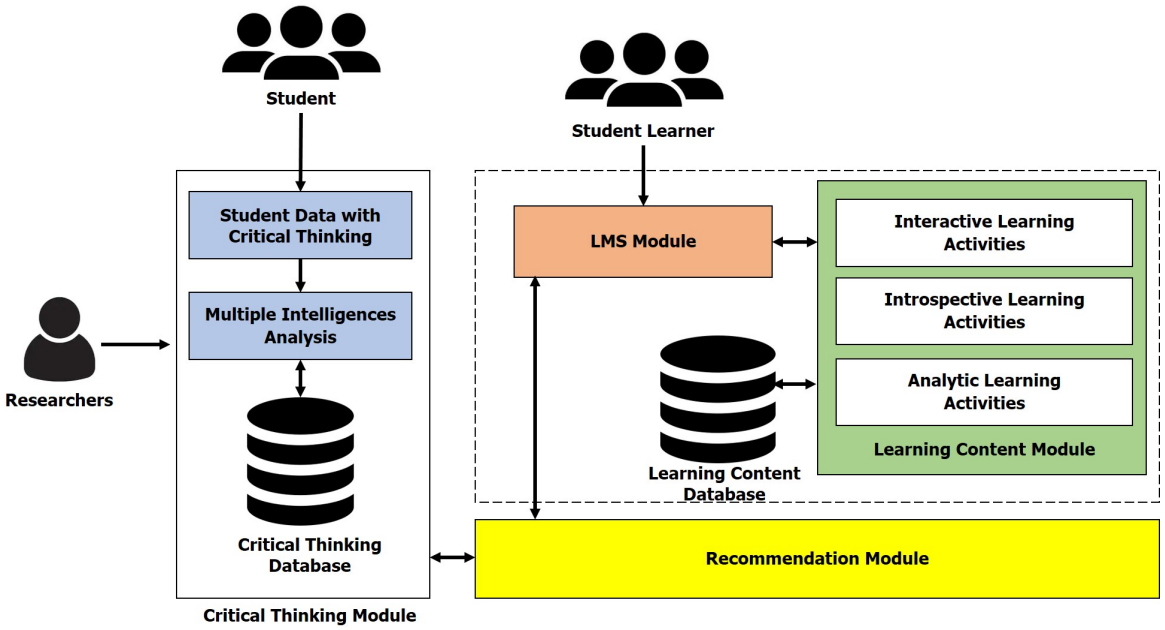


Fig. 2: Conceptual framework of multiple intelligences in e-Learning

shown in Fig. 3. The population of the research is 108 students in secondary 1 education level of The Demonstration School of Phayao University. The *Technology (Design and Technology 1)* subject is chosen with 4 chapters to develop e-Learning that is putted critical thinking intelligence in its chapters. Four chapters in *Technology (Design and Technology 1)* subject as control variables are:

1) Technology system,
2) Mechanical, electricity and principle of electronics,
3) Engineering design process, and
4) Solutions for engineering design process.

The process of development activities consists of 4 steps as following.

1) First step is a survey of variables that affects the ability of critical thinking intelligence of mathematical/logical Intelligence for students. This process studies previous research to study variables that have an impact on the ability of critical thinking intelligence. Moreover, multiple intelligence experts are also interviewed in this step. The students with expected critical thinking intelligence by intelligence examinations is chosen for sample groups. The sample groups is consists of 4 groups with 20 students per each group.
2) The second step is to create the examination activities to classify data from sample groups. The examination activities is focused to considered 4 properties of critical thinking intelligence: 1) Systems analysis, 2) Argument analysis, 3) Creation, and 4) Evaluation. The e-Learning activity process for supporting critical thinking skill can be shown in Fig. 4. The questions of the examinations consist of two parts. One covering general information about the respondents (defined as a main variable) such as name, year, field, faculty, etc. The second part is the question to separate the ability of critical thinking intelligence of mathematical/logical intelligence (defined as a prediction variable).
3) The third step is examinations information from a sample of students to answer the examinations created from the last part. This paper uses the sample of 20 students per each group of 4 properties of critical thinking intelligence.
4) The fourth step analyzes the examinations information from the previous step by using Data Mining technique.

This research uses five algorithms for analysis activities of critical thinking intelligence. All Algorithms are consists of:

- ID3 algorithm
- C4.5 algorithm
- NBTree algorithm
- Naive Bayes algorithm
- Bayes Net algorithm

Then, results from the 5 algorithms will be compared to find the most efficient referring to development the set of rules for supporting critical thinking intelligence. Fig. 5 shows the process of rule base design.

Fig. 6 presents the evaluation of creating a rule base to guide students by considering the percentage of precision from students ability. One can compare the evaluation of the rule base by using 5 methods;

1) ID3 algorithm
2) C4.5 algorithm
3) NBTree algorithm
4) Naive Bayes algorithm
5) Bayes Net algorithm.

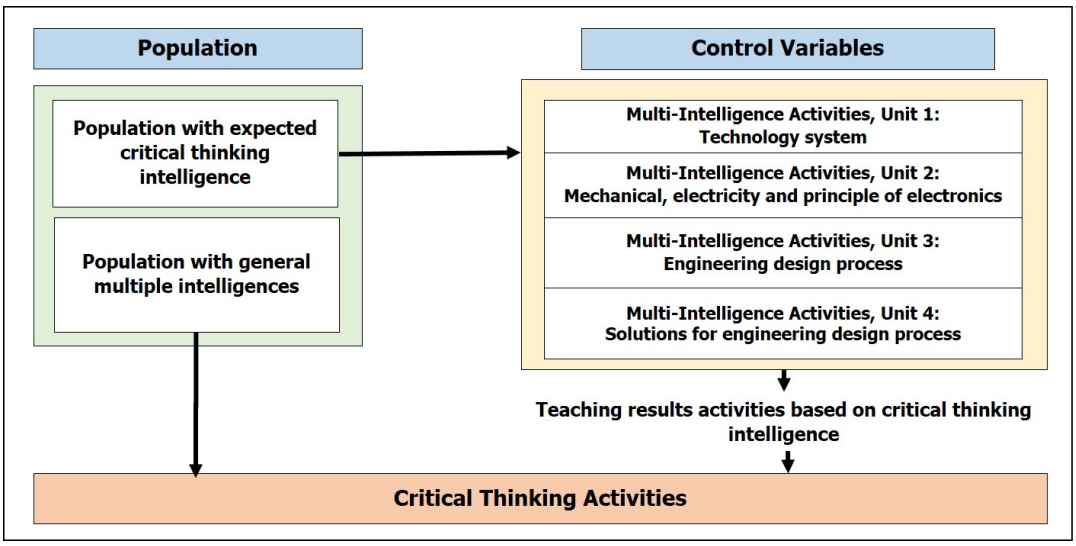


Fig. 3: Development framework of e-Learning activities for supporting critical thinking skill

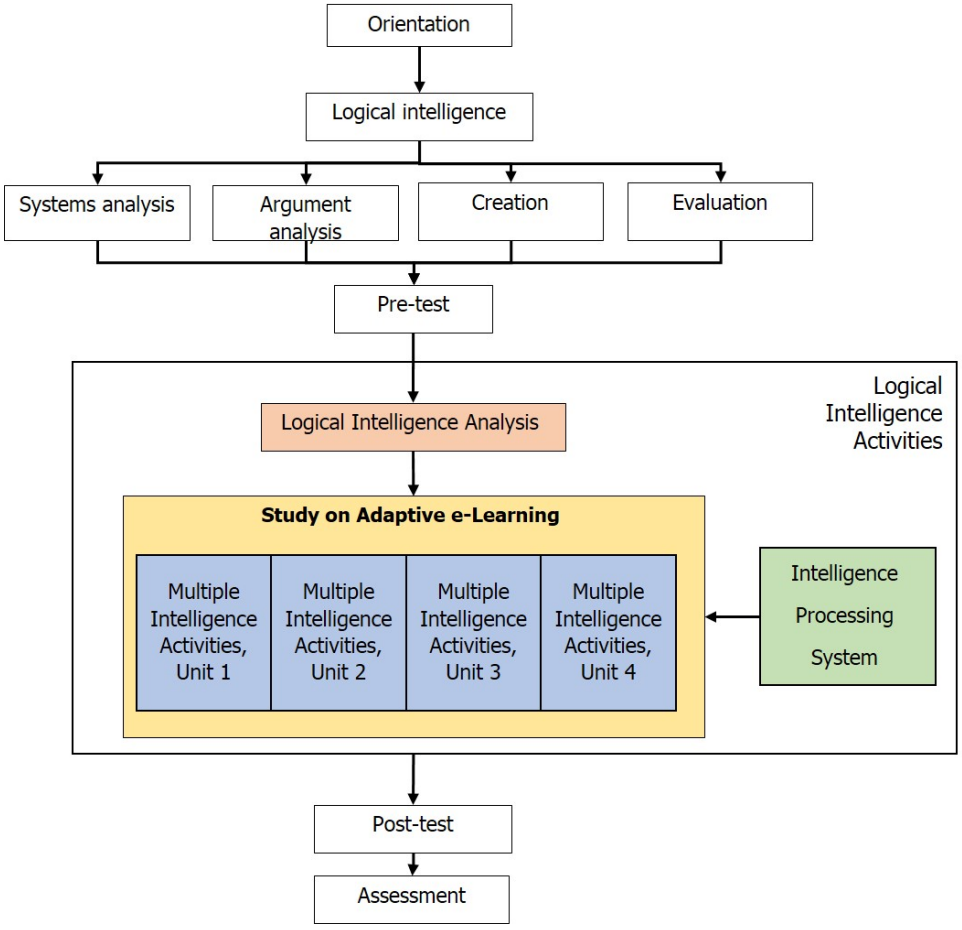


Fig. 4: e-Learning activity process for supporting critical thinking skill

Fig. 6 shows the accuracy comparison of the rule base. The result of each algorithm follows, ID3 algorithm equaled 78.62%, C4.5 algorithm equaled 83.43%, NBTree algorithm equaled 77.00%, Naive Bayes algorithm equaled 60.90%, and Bayes Net algorithm equaled 66.37%. The research focused on percentage of prediction for each algorithm. C4.5 algorithm was found to have the highest percentage of prediction. Percentage of prediction of C4.5 algorithm equaled 83.43%.

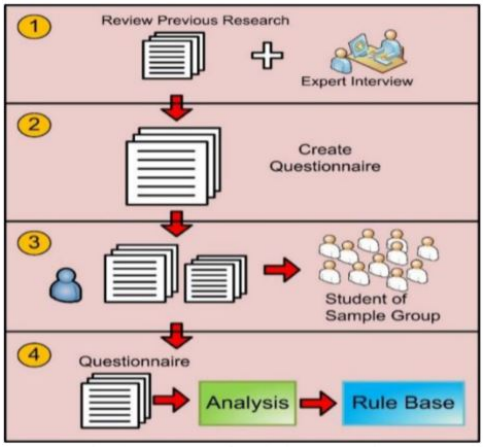


Fig. 5: Rule development process

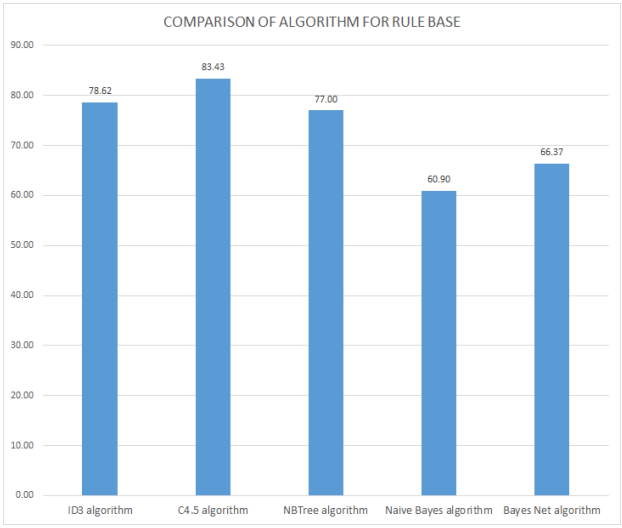


Fig. 6: Algorithm comparison for rule base

V. Conclusion

Multiple Intelligence theory is a classification and conceptualization of human intelligence. In order to make effective e-learning paradigm, MI theory offers a specific pluralistic conceptualization of intelligence. The paper has presented e-learning approach and various communication modes. The objective of this research was to development e-Learning activities based on multiple intelligences for supporting critical thinking. Design of the rule base consists of four parts. The first part was a survey of the variables. Second part was creation of the questionnaires. Third part was a survey using student sample groups. The fourth part was an analysis of data which came from the results of the survey. The process for selection of the rule base was undertaken by comparing 5 algorithms as follows 1) ID3 algorithm 2) C4.5 algorithm 3) NBTree algorithm 4) Naive Bayes algorithm 5) Bayes Net algorithm. The result of each algorithm is, ID3 algorithm are as follows 78.62%, C4.5 algorithm equaled 83.43%, NBTree algorithm equaled 77.00%, Naive Bayes algorithm equaled 60.90%, Bayes Net algorithm equaled 66.37%. When considering percentage of prediction for each algorithm, C4.5 algorithm had the highest percentage of prediction. Percentage of prediction for the C4.5 algorithm equaled 83.43%.

Acknowledgment

This research was supported in part by the School of Information and Communication Technology, University of Phayao, Thailand.

References

[1] S.-T. Yuan and Y.-C. Chen, "Semantic ideation learning for agent-based e-brainstorming," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, pp. 261–275, 03 2008.

[2] F. Aruan, A. S. Prihatmanto, H. Hindersah, and Kuspriyanto, "The designing and implementation of a problem based learning in collaborative virtual environments using mmog technology," in *2012 International Conference on System Engineering and Technology (ICSET)*, Sep. 2012, pp. 1–7.

[3] I. Arroyo, B. P. Woolf, W. Burelson, K. Muldner, D. Rai, and M. Tai, "A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect," *International Journal of Artificial Intelligence in Education*, vol. 24, no. 4, pp. 387–426, Dec 2014. [Online]. Available: https://doi.org/10.1007/s40593-014-0023-y

[4] S. Tangwannawit, N. Sureerattanan, and M. Tiantong, "Multiple intelligences learning activities model in e-learning environment," *International Journal Of The Computer , The Internet and Management*, vol. 16, pp. 25.1–25.4, 12 2008.

[5] R. W. Lau, N. Y. Yen, F. Li, and B. Wah, "Recent development in multimedia e-learning technologies," *World Wide Web*, vol. 17, no. 2, pp. 189–198, Mar. 2014. [Online]. Available: http://dx.doi.org/10.1007/s11280-013-0206-8

[6] M. Montebello, "Next generation e-learning," in *Proceedings of the 5th International Conference on Information and Education Technology*, ser. ICIET '17. New York, NY, USA: ACM, 2017, pp. 150–154. [Online]. Available: http://doi.acm.org/10.1145/3029387.3029405

[7] B. S. Gardner and S. J. Korth, "A framework for learning to work in teams," *Journal of Education for Business*, vol. 74, no. 1, pp. 28–33, 1998. [Online]. Available: https://doi.org/10.1080/08832329809601657

[8] H. Gardner and T. Hatch, "Educational implications of the theory of multiple intelligences," *Educational Researcher*, vol. 18, no. 8, pp. 4–10, 1989. [Online]. Available: https://doi.org/10.3102/0013189X018008004

Statistical Machine Translation between Kachin and Rawang

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

***Abstract*—This paper contributes the first evaluation of the quality of machine translation between Kachin and Rawang. We also developed a Kachin-Rawang parallel corpus (around 10K sentences) based on the Myanmar language of ASEAN MT corpus. The 10 folds cross-validation experiments were carried out using three different statistical machine translation approaches: phrase-based, hierarchical phrase-based, and the operation sequence model (OSM). The results show that all three statistical machine translation approaches give higher and comparable BLEU and RIBES scores for both Kachin to Rawang and Rawang to Kachin machine translations. OSM approach achieved the highest BLEU and RIBES scores among three approaches machine translation.**

***Index Terms*—Statistical Machine Translation, Under-resourced languages, Dialects, Kachin, Rawang**

I. Introduction

Our main motivation for this research is to investigate SMT performance for Kachin and Rawang language pair. The Kachin language is closely related to Rawang language and it is often considered as dialect of Kachin language. The state-of-the-art techniques of statistical machine translation (SMT) [1], [2] demonstrate good performance on translation of languages with relatively similar word orders [3]. To date, there have been some studies on the SMT of Myanmar language. Ye Kyaw Thu et al. (2016) [4] presented the first large-scale study of the translation of the Myanmar language. A total of 40 language pairs were used in the study that included languages both similar and fundamentally different from Myanmar. The results show that the hierarchical phrase-based SMT (HPBSMT) [5] approach gave the highest translation quality in terms of both the BLEU [6] and RIBES scores [7]. Win Pa Pa et al (2016) [8] presented the first comparative study of five major machine translation approaches applied to low-resource languages. PBSMT, HPBSMT, tree-to-string (T2S), string-to-tree (S2T) and OSM translation methods to the translation of limited quantities of travel domain data between English and Thai, Laos, Myanmar in both directions. The experimental results indicate that in terms of adequacy (as measured by BLEU score), the PBSMT approach produced the highest quality translations. Here, the annotated tree is used only for English language for S2T and T2S experiments. This is because there is no publicly available tree parser for Lao, Myanmar and Thai languages. According to our knowledge, there is no publicly available tree parser for both Dawei and Myanmar languages and thus we cannot apply S2T and T2S approaches for Myanmar-Dawei language pair. From their RIBES scores, we noticed that OSM approach achieved best machine translation performance for Myanmar to English translation. Moreover, we learned that OSM approach gave highest translation performance translation between Khmer (the official language of Cambodia) and twenty other languages, in both directions [9].

Relating to Myanmar langauge dialects, Thazin Myint Oo et al. (2018) [25] contributed the first PBSMT, HPBSMT and OSM machine translation evaluations between Myanmar and Rakhine. The experiment was used the 18K Myanmar-Rakhine parallel corpus that constructed to analyze the behavior of a dialectal Myanmar-Rakhine machine translation. The results showed that higher BLEU (57.88 for Myanmar-Rakhine and 60.86 for Rakhine-Myanmar) and RIBES (0.9085 for Myanmar-Rakhine and 0.9239 for Rakhine-Myanmar) scores can be achieved for Rakhine-Myanmar language pair even with the limited data. Based on the experimental results of previous works, in this paper, the machine translation experiments between Myanmar and Dawei were carried out using PBSMT, HPBSMT and OSM.

II. Related Work

Karima Meftouh et al. built PADIC (Parallel Arabic Dialect Corpus) corpus from scratch, then conducted experiments on cross dialect Arabic machine translation [10]. PADIC is composed of dialects from both the Maghreb and the Middle-East. Some interesting results were achieved even with the limited corpora of 6,400 parallel sentences. Using SMT for dialectal varieties usually suffers from data sparsity, but combining word-level and character-level models can yield good results even with small training data by exploiting the relative proximity between the two varieties [11]. Friedrich Neubarth et al. described a specific problem and its solution, arising with the translation between standard Austrian German and Viennese dialect. They used hybrid approach of rule-based preprocessing and PBSMT for getting better performance. Pierre-Edouard Honnet et al. proposed solutions for the machine translation of a family of dialects, Swiss German, for which parallel corpora are scarce [12]. They presented three strategies for normalizing Swiss German input in order to address the regional and spelling diversity. The results show that character-based neural MT was the most promising one for text normalization and that in combination with PBSMT achieved 36% BLEU score.

III. Kachin State and Kachin People

The Kachin State is situated in north Myanmar and is the place where the most of jinghpaw peoples live in. It lies between north latitude 23°27'and 28°25' longitude 96°0' and 98°44'. The area of Kachin State is 89,041km(34,379 sq mi). The capital of Kachin State is Myitkyina and Bhamo is the second largest historic city of jinghpaw peoples. About 2 millions of jinghpaw peoples live in Myanmar. In jinghpaw, there are two group called Bhamo jinghpaw and Myitkyina jinghpaw. The following pictures show the two groups of jinghpaw. Jinghpaw peoples live in Shan State. Jinghpaw peoples are on of the ethnic groups in Myanmar. Jinghpaw comprises six tribes or subdivisions: Lisu, Lashi, Rawang, Zaiwa, Lhao Vo. All have their own language and literature. As they all comes from Jinghpaw, they can also speak or use the Jinghpaw language . Firstly, the Jinghpaw literature want to explaine. Jinghpaw alphabet is based on Lathin script. The Jinghpaw literature was strated using in the era of "King Min Done Min"(1853-1878) . But the literature that Kachin peoples are using now was written on May 5,1895 by Dr. Ola Hanson, in the era of "King Thi Paul" (1878-1885) .

IV. Kachin Language

Jingpho (Jinghpaw, Chingp'o) or Kachin (Burmese: ကချင်ဘာသာ [kət` bàðà]) is a Tibeto-Burman language of the Sal branch mainly spoken in Kachin State, Burma and Yunnan, China. There are a lot of meanings for Jingpho. In the Jingpho language, Jingpho means people. The term "Kachin language" can refer either to the Jingpho language or to a group of languages spoken by various ethnic groups in the same region as Jingpo: Lisu, Lashi, Rawang, Zaiwa, Lhao Vo, Achang and Jingpho. These languages are from distinct branches of the highest level of the Tibeto-Burman family. The Jingpho alphabet is based on the Latin script. The ethnic Jingpho (or Kachin) are the primary speakers of Jingpho language, numbering approximately 900,000 speakers.The Turung of Assam in India speak a Jingpho dialect with many Assamese loanwords, called Singpho.

V. Rawang Language

Rawang peoples is the sub-group of Kachin(Jinghpaw) in Myanmar. Rawang peoples live in northern Kachin state: Puato, Machanbaw, Naungmaw, Kawnglangphu, and Pannandin townships. 70,000 of Rawang people live in Myanmar. There are four enthic roup in Rawang. They are Lungmi, Matwang, Daru and Tangsar. The Normal (-) , High (´) and Low (‘) symbols:

Example

VI. Grammar For Myanmar, Kachin and Rawang

Kachin and Rawang sentences use : . / , / ? / “ ” / ‘’ / ! / : / ; / -

Myanmar(my), Kachin(kc) and Rawang(rw) are in the same word order. Example of parallel sentences in Myanmar(my), Kachin(kc) and Rawang(rw) are given as follow :

my: နေကောင်းလား ။
kc: Hkam kaja ai i .
rw: PÀMVRÀ ÍÈ MÁ .

my:လုံချည် တစ်ထည် ဘယ်လောက်လဲ ။
kc: Ba hkang langai kade rai ?
rw: SHVRØM TÌQ DUNG KÀDVNGTĒ ÍÈ lĒ .

my: ထမင်းစား ပြီးပြီ လား ။
kc: Shat sha sai i ?
rw: VMPÀ VM BØĪ MÁ .

my: ကလေးများ ကစား နေကြသည် ။
kc: Ma ni gasup taw nga ma ai .
rw: CVMRE RĪ GVSØP MĒ .

VII. Methodology

In this section, we describe the methodology used in the machine translation experiments for this paper.

A. Phrase-Based Statistical Machine Translation

A PBSMT translation model is based on phrasal units [1]. Here, a phrase is simply a contiguous sequence of words and generally, not a linguistically motivated phrase. A phrase-based translation model typically gives better translation performance than word-based models. We can describe a simple phrase-based translation model consisting of phrase-pair probabilities extracted from corpus and a basic reordering model, and an algorithm to extract the

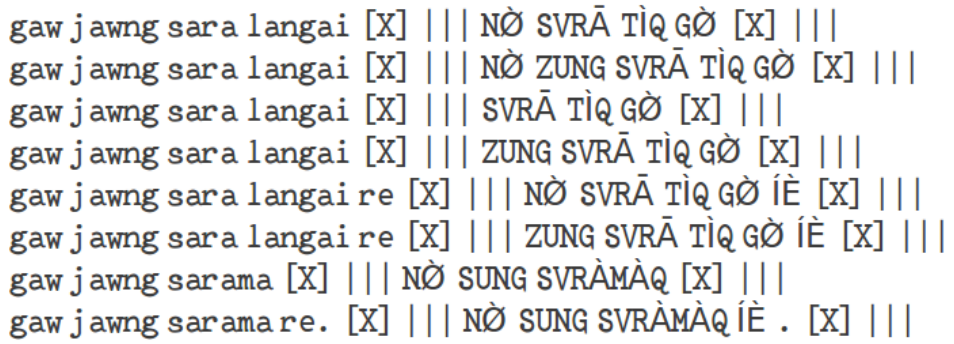


Fig. 1: Some examples of hierarchical phrase-based grammar between Kachin and RaWang phrases

phrases to build a phrase-table [14]. The phrase translation model is based on noisy channel model. To find best translation $\hat{e}$ that maximizes the translation probability $\mathbf{P}(f)$ given the source sentences; mathematically. Here, the source language is French and the target language is an English. The translation of a French sentence into an English sentence is modeled as equation 1.

$$\hat{e} = argmax_e \mathbf{P}(e|f) \tag{1}$$

Applying the Bayes' rule, we can factorized into three parts.

$$P(e|f) = \frac{\mathbf{P}(e)}{\mathbf{P}(f)} \mathbf{P}(f|e) \tag{2}$$

The final mathematical formulation of phrase-based model is as follows:

$$argmax_e \mathbf{P}(e|f) = argmax_e \mathbf{P}(f|e)\mathbf{P}(e) \tag{3}$$

We note that denominator $\mathbf{P}(f)$ can be dropped because for all translations the probability of the source sentence remains the same. The $\mathbf{P}(e|f)$ variable can be viewed as the bilingual dictionary with probabilities attached to each entry to the dictionary (phrase table). The $\mathbf{P}(e)$ variable governs the grammaticality of the translation and we model it using n-gram language model under the PBMT paradigm.

B. Hierarchical Phrase-Based Statistical Machine Translation

The hierarchical phrase-based SMT approach is a model based on synchronous context-free grammar [14]. The model is able to be learned from a corpus of unannotated parallel text. The advantage this technique offers over the phrase-based approach is that the hierarchical structure is able to represent the word re-ordering process. The re-ordering is represented explicitly rather than encoded into a lexicalized re-ordering model (commonly used in purely phrase-based approaches). This makes the approach particularly applicable to language pairs that require long-distance re-ordering during the translation process [15]. Some examples of hierarchical phrase based grammar between Dawei and Myanmar phrases are shown in Figure 1.

C. Operation Sequence Model

The operation sequence model which combines the benefits of two state-of-the-art SMT frameworks named n-gram-based SMT and phrase-based SMT. This model simultaneously generate source and target units and does not have spurious ambiguity that is based on minimal translation units [16] [17]. It is a bilingual language model that also integrates reordering information. OSM motivates better reordering mechanism that uniformly handles local and non-local reordering and strong coupling of lexical generation and reordering. It means that OSM can handle both short and long distance reordering. The operation types are such as generate, insert gap, jump back and jump forward which perform the actual reordering. The following shows an example translation process of English sentence Please sit here into Myanmar language with the OSM.

Source: Please sit here
Target: ကျေးဇူးပြုပြီး ဒီမှာ ထိုင်

Operation 1: Generate (Please, ကျေးဇူးပြုပြီး)
Operation 2: Insert Gap
Operation 3: Generate (here, ကျေးဇူးပြုပြီး ဒီမှာ)
Operation 4: Jump Back (1)
Operation 5: Generate (sit, ကျေးဇူးပြုပြီး ဒီမှာ ထိုင်)

VIII. Experiment

A. Corpus Statistics

We used 10K Myanmar sentences (without name entity tags) of the ASEAN-MT Parallel Corpus [18], which is a parallel corpus in the travel domain. It contains six main categories and they are people (greeting, introduction and communication), survival (transportation, accommodation and finance), food (food, beverage and restaurant), fun (recreation, traveling, shopping and nightlife), resource (number, time and accuracy), special needs (emergency and health). Word segmentation for Rawang was done manually. We held 10-fold cross-validation experiments and used 8,468 to 8,519 sentences for training,

500sentences for development and 985 to 1,026 sentences for evaluation respectively.

B. Moses SMT System

We used the PBSMT, HPBSMT and OSM system provided by the Moses toolkit [19] for training the PB-SMT, HPBSMT and OSM statistical machine translation systems. The word segmented source language was aligned with the word segmented target language using GIZA++ [20]. The alignment was symmetrized by grow-diag-final and heuristic [1]. The lexicalized reordering model was trained with the msd-bidirectional-fe option [21]. We use KenLM [22] for training the 5-gram language model with modified Kneser-Ney discounting [23]. Minimum error rate training (MERT) [24] was used to tune the decoder parameters and the decoding was done using the Moses decoder (version 2.1.1) . We used default settings of Moses for all experiments.

IX. Evaluation

We used two automatic criteria for the evaluation of the machine translation output. One was the de facto standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [6] and the other was the Rank-based Intuitive Bilingual Evaluation Measure (RIBES) [7]. The BLEU score measures the precision of n-gram (over all n 4 in our case) with respect to a reference translation with a penalty for short translations [6]. Intuitively, the BLEU score measures the adequacy of the translation and large BLEU scores are better. RIBES is an automatic evaluation metric based on rank correlation coefficients modified with precision and special care is paid to word order of the translation results. The RIBES score is suitable for distance language pairs such as Myanmar and English. Large RIBES scores are better.

X. Results and Discussion

The BLEU and RIBES score results for machine translation experiments with PBSMT, HPBSMT and OSM are shown in Table 1. Bold numbers indicate the highest scores among three SMT approaches. The RIBES scores are inside the round brackets. Here, "kc" stands for Kachin, "rw" stands for Rawang, "src" stands for source language and "tgt" stands for target language respectively.

The BLEU and RIBES score results for machine translation experiments with PBSMT, HPBSMT and OSM between Kachin and Rawang languages are shown in Table 1. From the results, OSM method achieved the highest BLEU and RIBES score for both Kachin to Rawang and Rawang to Kachin bi-directional machine translations. Interestingly, the BLEU and RIBES score of all three methods are comparable performance. Our results with current parallel corpus indicate that Rawang to Kachin machine translation is better performance (around 3 BLEU and 0.02 RIBES scores higher) than Kachin to Rawang machine translation direction.

As we expected, generally, machine translation performance of all three SMT approaches between Kachin and Rawang languages achieved good scores for both BLEU and RIBES. The reason is that as we mentioned in Section 3, the two languages, Kachin and Rawang are close languages. We assume that long distance reordering is relatively rare and only local reordering is enough for the Kachin-Rawang language pair.

XI. Conclusion

This paper contributes the first PBSMT, HPBSMT and OSM machine translation evaluations from Kachin to Rawen and Rawen to Kachin. We used the 10,000 Kachin-Rawang parallel corpus that we constructed to analyze the behavior of a dialectal Kachin-Rawang machine translation. We showed that higher BLEU and RIBES scores can be achieved for Kachin-Rawen language pair even with the limited data. In fucture work , we would like to improve SMT approach for Jinghpaw dialect language . Such as Jinghpaw, Lisu, Lashi, Rawang, Zaiwa, Lhao Vo.

Acknowledgment

We would like to thank

References

[1] Koehn, Philipp and Och, Franz Josef and Marcu, Daniel, "Statistical phrase-based translation," Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, 2003, pp. 48–54.

[2] Koehn, Philipp and Hoang, Hieu and Birch, Alexandra and Callison-Burch, Chris and Federico, Marcello and Bertoldi, Nicola and Cowan, Brooke and Shen, Wade and Moran, Christine and Zens, Richard and Dyer, Chris and Bojar, Ondřej and Constantin, Alexandra and Herbst, Evan, A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 2007, pp. 177–180.

[3] Koehn, Philipp, "Europarl: A parallel corpus for statistical machine translation," Conference Proceedings: the tenth Machine Translation Summit, 2005, pp. 79–86.

[4] Ye Kyaw Thu, Andrew Finch, Win Pa Pa, and Eiichiro Sumita, "A Large-scale Study of Statistical Machine Translation Methods for Myanmar Language," in Proceeding of SNLP2016, February 10-12, 2016.

[5] Chiang, David, "Hierarchical phrase-based translation," Computational Linguistics 33(2), 2007, pp. 201-228.

[6] Papineni, Kishore and Roukos, Salim and Ward, Todd and Zhu, Wei-Jing, "BLEU: a Method for Automatic Evaluation of Machine Translation," Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Philadelphia, Pennsylvania, 2002, pp. 311–318

[7] Isozaki, Hideki and Hirao, Tsutomu and Duh, Kevin and Sudoh, Katsuhito and Tsukada, Hajime, "Automatic evaluation of translation quality for distant language pairs," Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010, pp. 944-952.

[8] Win Pa Pa,Ye Kyaw Thu, Andrew Finch and Eiichiro Sumita, "A Study of Statistical Machine Translation Methods for Under Resourced Languages," 5th Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU Workshop), 09-12 May, 2016, Yogyakarta, Indonesia, Procedia Computer Science, Volume 81, 2016,pp. 250–257.

TABLE I: Average BLEU and RIBES scores for PBSMT, HPBSMT and OSM

src-tgt	PBSMT	HPBSMT	OSM
kc-rw	43.071 (0.79964)	43.197 (0.80064)	**43.973 (0.80079)**
rw-kc	46.281 (0.81064)	46.129 (0.81224)	**46.597 (0.81180)**

[9] Ye Kyaw Thu, Vichet Chea, Andrew Finch,Masao Utiyama and Eiichiro Sumita, "A Large-scale Study of Statistical Machine Translation Methods for Khmer Language" 29th Pacific Asia Conference on Language, Information and Computation,October 30 - November 1, 2015,Shanghai, China,pp. 259-269.

[10] Karima Meftouh, Salima Harrat, Salma Jamoussi, Mourad Abbas and Kamel Smaili, "Machine Translation Experiments on PADIC: A Parallel Arabic DIalect Corpus," oin Proc. of the 29th Pacific Asia Conference on Language, Information and Computation, PACLIC 29, Shanghai, China, October 30 - November 1, 2015, pp. 26-34.

[11] Neubarth Friedrich, Haddow Barry, Huerta Adolfo Hernandez and Trost Harald, "A Hybrid Approach to Statistical Machine Translation Between Standard and Dialectal Varieties," Human Language Technology, Challenges for Computer Science and Linguistics: 6th Language and Technology Conference, LTC 2013, Poznan, Poland, December 7-9, 2013, Revised Selected Papers, pp .341–353.

[12] Pierre-Edouard Honnet, Andrei Popescu-Belis, Claudiu Musat and Michael Baeriswyl, "Machine Translation of Low-Resource Spoken Dialects: Strategies for Normalizing Swiss German," CoRR journal, volume (abs/1710.11035), 2017.

[13] John Okell, "Three Burmese Dialects," 1981, London Oxford University press, Univeristy of London.

[14] Lucia Specia„ "Tutorial, Fundamental and New Approaches to Statistical Machine Translation," International Conference Recent Advances in Natural Language Processing, 2011

[15] Braune, Fabienne and Gojun, Anita and Fraser, Alexander, "Long-distance reordering during search for hierarchical phrase-based SMT," n Proc. of the 16th Annual Conference of the European Association for Machine Translation, 2012, Trento, Italy, pp. 177-184.

[16] Durrani, Nadir and Schmid, Helmut and Fraser, Alexander, "A Joint Sequence Translation Model with Integrated Reordering," in Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, 2011, Portland, Oregon, pp. 1045-1054.

[17] Nadir Durrani, Helmut Schmid, Alexander M. Fraser, Philipp Koehn and Hinrich Schutze "The Operation Sequence Model - Combining N-Gram-Based and Phrase-Based Statistical Machine Translation," Computational Linguistics, Volume 41, No. 2, 2015, pp. 185-214.

[18] Prachya, Boonkwan and Thepchai, Supnithi, "Technical Report for The Network-based ASEAN Language Translation Public Service Project," Online Materials of Network-based ASEAN Languages Translation Public Service for Members, NECTEC, 2013.

[19] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation," Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic, June 2007.

[20] Och Franz Josef and Ney Hermann, "Improved Statistical Alignment Models," in Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong, China, 2000, pp. 440-447.

[21] Tillmann Christoph, "A Unigram Orientation Model for Statistical Machine Translation," in Proceedings of HLT-NAACL 2004: Short Papers, Stroudsburg, PA, USA, 2004, pp. 101-104.

[22] Heafield, Kenneth, "KenLM: Faster and Smaller Language Model Queries," in Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT '11, Edinburgh, Scotland, 2011, pp. 187-197.

[23] Chen Stanley F and Goodman Joshua, "An empirical study of smoothing techniques for language modeling," in Proceedings of the 34th annual meeting on Association for Computational Linguistics, 1996, pp. 310-318.

[24] Och Franz J., "Minimum error rate training in statistical machine translation," in Proceedings of the 41st Annual Meeting n Association for Computational Linguistics – Volume 1,Association for Computer Linguistics, Sapporo, Japan, July, 2003, pp.160-167.

[25] Thazin Myint Oo, Ye Kyaw Thu, Khin Mar Soe, "Statistical Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)", In Proceedings of ICCA2018, February 22-23, 2018, Yangon, Myanmar, pp. 304-311

[26] (NIST) The National Institute of Stan- dards and Technology, Speech recognition scoring toolkit (sctk), version: 2.4.10, 2015

[27] Miller, Frederic P. and Vandome, Agnes F. and McBrewster, John, Levenshtein Distance: Information Theory, Computer Science, String (Computer Science), "String Metric, Damerau Levenshtein Distance, Spell Checker, Hamming Distance", ISBN: 6130216904, 9786130216900, Alpha Press, 2009.

Predicting learning organization factors that affect performance by data mining techniques.

Wiwit Suksangaram
Department of Business Information Management
Phetchaburi Rajabhat University
Thailand
Phetchaburi , Thailand
suksangaram@hotmail.com

Waratta Hemtong
Department of thai traditional medicine
Phetchaburi Rajabhat University
Thailand
Phetchaburi , Thailand
Waratta.hem@mail.pbru.ac.th

Sopaporn Klamsakul
Department of Business administration
Phetchaburi Rajabhat University
Thailand
Phetchaburi , Thailand
Sopapon_s@yahoo.com

Abstract— **This research proposed factors and model affecting performance in learning organization prediction by using the classification techniques. Results show learning organization factors affecting performance composed of 8 factors including personality factors focusing on excellence, learning dynamics, common vision factors, team learning factors, workload factor, and performance factor. Comparison of classification models showed that SVM technique was the most suitable technique in prediction of learning organization affecting employees performance in the Bank for agriculture and Agricultural Cooperatives in the western region with 98.33% of accuracy, 0.025 of precision, and 0.984 of recall values.**

Keywords—learning organization, Data Classification, Decision tree, Naïve Bayes, support Vector Machine

I. INTRODUCTION

Bank for Agriculture and Agricultural Cooperatives Financial institutions for rural development by providing financial assistance and supporting development to rural target groups. The Bank has a strong need for successful corporate development. The focus is on human resources development. It is hoped that the organization will develop. Being a learning organization to develop people and organizations, it is also to maintain personnel to stay with the organization and dedicated to work. This is an important factor that affects the effectiveness of employees.

This research uses organizational learning factors that influence the effectiveness of employees in the Bank for Agriculture and Agricultural Cooperatives in the western region. In the data analysis. Create modeling the effectiveness of bank farm workers using data mining techniques, including data classification. The use of data mining techniques to predict the effectiveness of the work done by the learning organization. The results will lead to the development of learning organizations that affect the effectiveness of the work of personnel in the organization. Make better.

II. PRELIMINARY

A. Learning Organization

Organization of learning It is an organization where executives and members of the organization are continually applying their knowledge to develop themselves. Can exchange each other's learning methods. By the features of the learning organization. Of the Bank for Agriculture and Agricultural Cooperatives Individuals who focus on excellence. The dynamics of learning. The vision is shared. Teamwork Learning and technology.

1.Being a person who focuses on excellence being enthusiastic. Always try and learn new things by believing that the potential, knowledge, and ability are all things that people seek and gain. They can learn and increase their potential to achieve their goals and achieve their goals. The promotion of the agency to learn new things. And self development and like to spend time in creating new things. To develop the work done. In the present

2.Learning dynamics refers to a person's ability to improve. Develop ways to work in the face of the situation. By solving problems in a systematic and systematic way, rather than solving problems, one can see the relationships of elements in the work that affect each other in a chain. Organize the database in order to make decisions.because in developing the learning organization, it is necessary to apply the system in every step.

3.The shared vision means the organization is open minded. Be willing to listen to the opinions or feelings of employees. And in the same direction throughout the organization. The vision of the organization can be applied in real work. Can bring the vision of the organization to apply in the actual work, accept and be ready to comply with the agency agreement. It recognizes that the future and success of the agency. It is a shared responsibility of all personnel.

4.Learning together as a team means acquiring new knowledge. To learn new knowledge to exchange knowledge as a new knowledge. And systematically store knowledge. To apply knowledge to work. And can bring knowledge to exchange each other.

5.Technology refers to the use of learning tools in the organization. And to share knowledge with each other. The organization should train computer personnel to use it effectively. It also provides knowledge about technology in operation and can bring modern technology to use in the operation.

B. Work efficiency

Work efficiency The ability of employees to bring existing knowledge into the work to achieve the goal. Based on individual production. The quality of work, workload, achievement in work.

1. Quality of work Job satisfaction assigned by the supervisor. By understanding the content of your assigned work, the assigned task is accomplished. The quality of work is determined by the supervisor.

2.Workload Ability to complete tasks assigned. Complies with standards. Able to work under pressure and suitability of staff.

3.Achievement in work Will willingness to perform work assigned by the rules and regulations to achieve the achievement of the task. By colleagues and supervisors to assist you. When problems are encountered. Can manage time in operation.

C. Data mining

Data mining is the process of finding information or knowledge in a large, complex database and bring knowledge to use in decision making. Information that can be generated can be predicted. For classifying or expressing relationships between properties. Or provide a summary of the essence of the information. It consists of statistical process and computer-based learning. To create models, rules, models, forecasts. And information from large databases and to be able to extract knowledge from a large database.

D. Data Classification

Data classification techniques are techniques for searching knowledge on large databases. It is a technique of data modeling to manage the data in a given group from the data set called data system. The data consists of many attributes. The attribute may be a continuation or a group, with a break attribute. This is the data class identifier. The purpose of data classification is to model the separation of one attribute based on the other attributes. Models derived from classifying data will make it possible to classify data into unpartitioned data in the future.

E. Decision tree

Decision tree [1] is a data model for predictive modeling. And supervised learning is the learning of a teacher model. Modeling can be made from predefined data samples. And the group of items that have not been categorized by the tree.

F. Naïve Bayes

Naïve Bayes [2] is a clustering model based on probability based on Bayes'Theorem And assumptions that determine the occurrence of events. The grouping is independent of each other. This learning by Naïve Bayes process has been widely used in machine learning. Naïve Bayes has gained popularity due to its uncomplicated work. It works effectively, such as the research on document classification, etc.

G. Support Vector Machine

Support vector machine [3] It's a matter of putting the value of the data into the Feature Space and then finding the lines that divide the two data together by creating a straight line hyper plane. And to know which straight lines divide two groups together, which one is the best line. The original support vector machine was used for linear data. In fact, the information that is used in the teaching system, most of the information is usually nonlinear. This can be solved by introducing a Kernel Function to the multidimensional data classification. Use the most appropriate selection. The feature selection is based on the information that is taught.

H. Review literature and related research

Sakkarin phupanna, paweena wanchai,wararat songpan and nunnapus benjamas [4] predictive analytic for student dropout in high vocational certificate using data mining technique. The aim of paper is to compare the performance of different classification techniques included c4.5, naïve Bayesian learning and multilayer perception algorithms. To predict and analyze factors influencing student's decision to drop out. The results show that the performance of multilayer perception is 94.45% which higher than the c4.5 and Naïve Bayesian learning, with 93.74% and 83.74% accordingly. The findings also indicate that student's decision to drop out significantly influenced by type of school, type of course, first grade, student loan and major.

Narroup rakngam, Narongdech Keeratipranon [5] The suspicious Transactions forecasting using Neural Network. This study was presented the model for identify the transaction that could related to money laundering.in this paper present the process to forecast the suspicious transactions by considering the transactions with other components such as behavioral characteristics and personal information. We use the artificial neural network system to identify the suspicious transactions. And found that the system could identify the suspicious transactions is highly effective at a good level.

Konstantions Pliakos, Celine Vens [6] Mining features for biomedical data using clustering tree ensembles. The volume of biomedical data available to the machine learning community grows very rapidly. In this article, we emphasize on the aforementioned problem and propose a target-informed feature induction method based on tree ensemble learning. The contribution of this article is twofold. Firstly, a problem affecting the quality of biomedical data is highlighted, and secondly, a method to handle that problem is proposed. The efficiency of the presented approach is validated on multi-target prediction tasks. The obtained results indicate that the proposed approach is able to boost the discrimination between the data instances and increase the predictive performance.

III. PROPOSED METHOD

In this research. Data mining was done using the Weka 3.82 program, with the decision tree C4.5 naïve Bayes algorithm and the support vector machine to find the appropriate model for determining the learning organization factors that affect the effectiveness of the work.

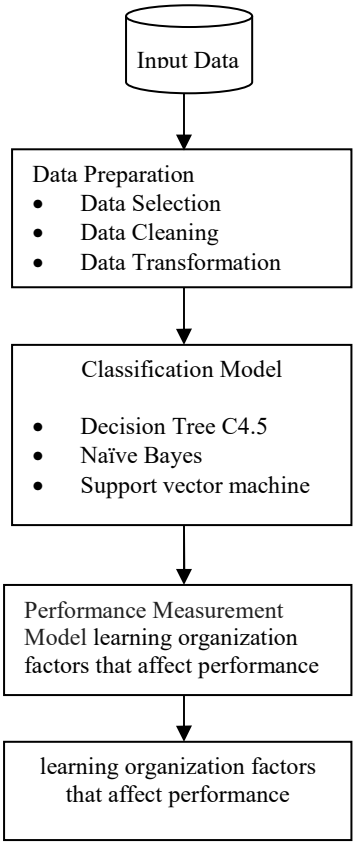


Figure 1. Framework for learning organization factors and creating model for predicting performance.

A. Data Preprocessing

It prepares data before producing data mining. This step includes, Data selection, Data Cleaning, Data Transformation and Data Reduction to have accurate and appropriate data for clustering and classification.

B. Feature Selection

It selects data influencing learning organization factors that affect performance using personality factors that focus on excellence, The dynamics of learning, common vision factors, team learning factors, technology factor, quality factor, workload factor, performance factor to seek factor relationships in different dimensions. Data is clustered by DBScan Algorithm to find factor relationship learning organization factors that affect performance.

Table 1 Learning organization factors that affect performance attributes

variable	Description	Variable type
L_focus_excellence	personality factors that focus on excellence	Number
LD_learning,	The dynamics of learning	Number
LC_vision	common vision factors	Number
LT_learning	team learning factors	Number
L_technology	technology factor	Number
L_quality	quality factor	Number
L_workload	workload factor	Number
L_performance	performance factor	Number

C. *Comparison Classification Model*

The researcher designed an experiment to test efficiency by considering Accuracy, Precision and Recall as illustrated in Equation (1) (2) and (3) as follows;

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

When TP = True Positive, FP = False Positive, FT = True Negative และ FN = False Negative

IV. Experimental Results

To study the efficiency of this model, we used WEKA version 3.82 as a tool to model the C4.5, Naïve Bayes and support vector machine using the default parameters of the WEKA program. In this research, 10 fold Cross-validation The results are shown in Figure 2, Figure 3 and Figure 4.

A. *The Results of General Model Efficiency of General Model*

The results of creating a model to classify academic achievement through the 3models give the efficiency value consisting of accuracy, precision and recall as illustrated in Table 1.

TABLE I. ***A COMPARISON OF GENERAL MODEL EFFICIENCY***

Model	10-fold cross validation		
	Accuracy	Recall	Precision
C4.5	98.33	0.984	0.009
Naïve Bayes	78.66	0.836	0.104
SVM	98.33	0.984	0.025

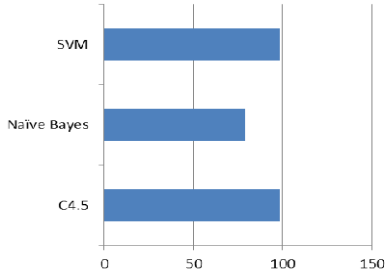


Figure 2. Model Performance Measurement by accuracy

Figure 2 The Accuracy of the model was C4.5 and SVM with the highest Accuracy of 98.33% and the Bayesian Naïve technique was 78.66% Accuracy.

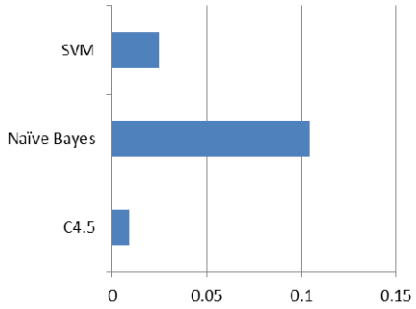


Figure 3 Model Performance Measurement by Precision

Figure 3 The precision values of the model were naive Bayes technique 0.104 precision, SVM precision 0.025 and C4.5 precision 0.009.

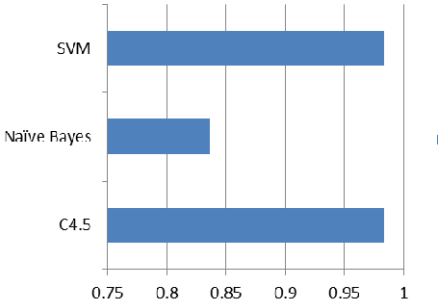


Figure 4 Model performance measurement by Recall

Figure 4 The Recall values of the model were C4.5, Recall 0.984, SVM 0.984 and Naïve Bayes for Recall 0.836 respectively.

V. Conclusion

This study aimed to study and compare the appropriate models. The predictions of organizational factors affecting the working ability of employees of the Bank for Agriculture and Agricultural Cooperatives in the western region. The model was used to compare 3 techniques: C4.5, Naïve Bayes and SVM. The results showed that SVM technique was the most suitable for predicting organizational learning factors affecting the performance of employees of the Bank for Agriculture and Agricultural Cooperatives in the western region. By measuring the performance of the model with Accuracy 98.33% Precision 0.025 and Recall 0.984.

References

[1] J.Han and M.Kamber, Data Mining Concepts and Techniques, Second Edition ed.:Morgan Kaufmann, 2006.

[2] Joachims, “Text categorization with support vector machines: Learning with many relevant features”. Proc. of the 10 th European Conf. on Machine Learning, pp.137-142, 1998.

[3] Pornpon Thamrongrat, Ladda Preechaveerakul, and wiphada wettayaprasit, Web Page Classification Using Feature Reduction and support Vector Machine, The 12th National Computer Science and Engineering Conference, 2008.

[4] Sakkarin phupanna,paweena wanchai,waraat songpan and nunnapus benjamas, Predictive analytic for student dropout in high vovational certificate using data mining technique. The Thirteenth National Conference on Computing and Information Technology 2017, pp. 51-56.

[5] Narroup Rakngam, Narongdech Keeratipranon, The suspicious Transactions forecasting using Neural Network. The Thirteenth National Conference on Computing and Information Technology 2017, pp.14-19.

[6] Konstantions Pliakos, celine Vens, Mining features for biomedical data using clustering tree ensembles. Jornal of Biomedical Informatics (vol.85), pp.40-48.

A Supportive Environment for Knowledge Construction based on Semantic Web Technology: A Case Study in a Cultural Domain

Akkharawoot Takhom[1,2], Dhanon Leenoi[1], Pitchaya Soomjinda[3], Sasiporn Usanavasin[2], Prachya Boonkwan[1], Thepchai Supnithi[1]
[1]*Language and Semantic Technology Research Team*
National Electronics and Computer Technology Center Pathumthani, Thailand
[2]*School of Information, Computer and Communication Technology*
Sirindhorn International Institute of Technology, Pathumthani, Thailand
[3]*Faculty of Fine Arts, Chiang Mai University, Chiang Mai, Thailand*
akkharawoot.tak@ncr.nstda.or.th, dhanon.leenoi@nectec.or.th, pitchaya.soomjinda@cmu.ac.th
sasiporn@siit.tu.ac.th, prachya.boonkwan@nectec.or.th, thepchai.supnithi@nectec.or.th

***Abstract*—It is indisputable that the process of knowledge construction and knowledge management require a lot of collaborations from many stakeholders such as domain experts and knowledge engineers. To effectively model the knowledge in ontology representation, which can serve business applications and organization's goals, we need tools to support and to facilitate the collaborative works and activities among many stakeholders. In this paper, we propose a CD-OAM framework that provides a supportive environment based on the collaborative development (COD) approach. In this framework, stakeholders can communicate and share their understandings and comments during the process of ontology design in order to improve the quality of the knowledge model. To demonstrate how our approach and framework can support the collaborations and knowledge integrations among multiple domain experts and knowledge engineers, we selected a case study in cultural domain because the cultural knowledge is complex and requires various experts to design the ontology. The key contributions of our framework are: (1) presenting an improvement of collaborative activities through a supportive environment based on the COD approach, and (2) demonstrating a collaborative situation to overcome limitations of a communication between domain exports and knowledge engineers.**

Keywords—Knowledge Construction, Collaboration, Ontology Engineering, Semantic Web Technology

I. INTRODUCTION

Semantic Web technology [1] has a crucial role in sharing knowledge within the ontological engineering and a specific-domain community. Towards knowledge construction in particular knowledge has motivated many research works and modeling efforts of stakeholders such as domain experts. Ontology-based applications have been developed to support analysis and insights at a particular goal and the core resource is ontologies which require knowledge engineering and management.

Whereas, only a person cannot develop an effective ontology. Then, an approach of *collaborative ontology development* (*COD*) [2] was applied in most of the supporting tools that support various requirements, such as discussions, a process of improving ontology.

Building a consensus in ontology designing is one of the challenge tasks that require many collaborative activities from various stakeholders through the COD cycle.

In this paper, we provide a supportive environment based on the COD approach to encourage participants to communicate and share their understanding for improving domain ontology. A case study demonstrates how the framework can employ Semantic Web technology to construct ontology in a cultural domain. Supportive features allow stakeholder linking up communication and integrate knowledge for improving Buddha Image ontology.

The key contributions of our working approach are: (1) presenting an improvement of collaborative activities through a supportive environment based on the COD approach, and (2) demonstrating a collaborative situation to overcome limitations of a communication between domain exports and knowledge engineers.

The remainder of the paper is organized as follows: Section II explains a collaborative framework. In Section III, a case study demonstrates how the framework can support collaboration of stakeholders in ontology construction. Section IV distinguish our approach with related work and discuss the contribution and important issues of this paper. Finally, Section V concludes the main findings of the paper and gives an outlook on further work.

II. FRAMEWORK

A. Simplifying Ontology-based Application Development

To encourage stakeholders in various knowledge fields to develop an ontology, ontology-based applications have been developed to represent benefits of their knowledge. However, a high learning curve and efforts demanded are obstacles in building Semantic Web ontology and the ontology-based applications. A supportive framework, called an ontology application management (OAM) [3], purposed simplifying application development based on ontology. The framework provides a graphical ontology editor, Hozo [4] that allows stakeholders to design and visualize knowledge in forms of domain concepts and concept properties.

In this paper, the OAM framework was the first supportive environment to work with stakeholder to understand their requirements in part of ontology development. Therefore, their usage experiences and feedbacks are analyzed for improving our COD approach.

B. Collaborative Activities for Ontology Development

One necessary service of the COD approach is to provide a communication service that allows stakeholders to discuss and share knowledge and their understanding with others. *Community-driven ontology-based application management* (*CD-OAM*) [5], is a supportive environment enhancing a capability of collaboration for the OAM framework. The framework aimed to support stakeholders in knowledge sharing collaboration, especially knowledge co-creation of multidisciplinary knowledge. As illustrated in Fig. 1, a

multitier architecture of the framework is elucidated its details as follows:

1) User tier is the first access web application for stakeholders who want to work with ontology construction.

2) Application tier composes of three key management services: community, knowledgebase, and database. A structure of domain knowledge can represent through ontology editor and visualization in this tier.

3) Service tier provides supportive services following the COD approach: community consensus, community forum, post generator misunderstanding diagnosis, and instantiation.

4) Knowledge tier stores resources of knowledge that are created and uploaded by system users.

5) Data tier is handled data by the application tier requests.

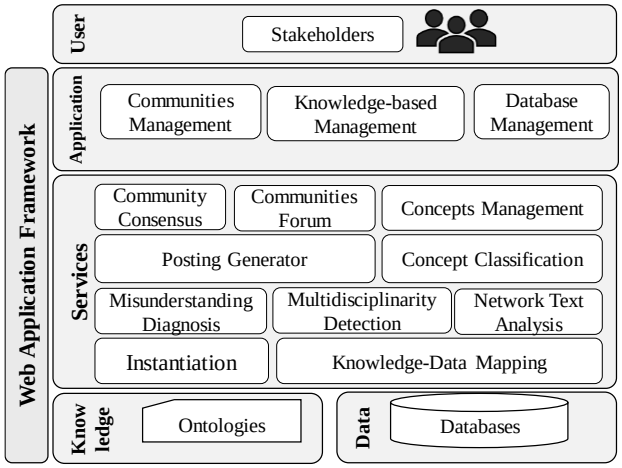


Fig. 1. Multitier architecture of a community-driven ontology-based application management (CD-OAM) framework [5].

III. Case Study on a Collaborative Workshop

A collaborative workshop firstly explains how we can work with stakeholder in a cultural domain to develop ontologies. Then we present our working stages following the COD approach for ontology construction, and a case study of constructing Buddha Image's ontology was selected to collect usage experience and feedback from stakeholders. Lastly, we demonstrate how features of the framework can support their collaborative activities.

A. Development of Cultural Domain Ontology

We organized a collaborative workshop for ontology construction at Chiang Mai International Convention and Exhibition Center on February 2nd, 2019. The workshop intended to encourage domain experts in a cultural domain to understand and to work with the ontology-based Semantic Web technology. As shown in the left of Fig.2, we let stakeholder to select and to explain their domain knowledge through a paper-based design and we represented their domain knowledge in 'Hozo', a visualization of the ontology editor. Then, in the right of Fig.2, the participants presented their expertise ontological-design, mostly in a cultural domain. Consequently, the knowledge engineers analyzed and revised ontological design corresponding the designing theory.

Fig. 2. A collaborative workshop for encouraging domain experts in a cultural domain to develop ontology at Chiang Mai international convention and exhibition center on February 2, 2019.

Different ontologies in the workshop underlying a cultural domain have been developed as follows: Thai tribe ontology, Thai traditional vehicle ontology, worship ontology and buddha image ontology.

For a case study in this paper, we selected "Buddha Image" which is a sacred statue representing the Lord Buddha. The Indian Buddha images had been found in Thailand since the 3rd - 4th century, then they have been created unique styles from the 7th century until now. Each style of statues reflects craftsman uniqueness, artistic development, prosperity, and decline. To indicate the style of unknown Buddha statues could be the key to open the mysterious or unrecorded history.

B. Development of Domain Ontology

In this workshop, an instruction of ontology development follows Noy and McGuinness methodology [6], and consider a theory of role modeling [7]. To facilitate the workshop's participants, we adjusted some part of the working stages into seven working stages, and we explain with the case study in each stage as follows.

First, the stakeholder *determines the domain and scope* the ontology (*Stage 1*). In this workshop, the domain experts define a scope that they need to define the Buddhist era or time of the creation by considering typically components of Buddha image. Then, we looked for the *reusing existing ontologies* (*Stage 2*), but it difficult to reuse in this purpose.

After that, we defined the components of Buddha image to *enumerate important terms* (*Stage 3*). As shown in Fig.3, we selected the bronze Buddha statue in Sukhothai style, invented in 21st century as a sample. The statue composes of posture, head, body, robes, hand-gestures, feet and legs, and time of creation.

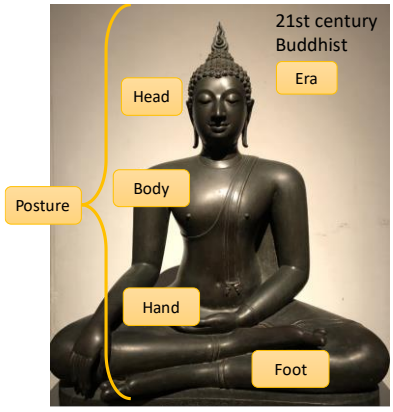


Fig. 3. A sample of the bronze Buddha statue in Sukhothai style, invented in 21st century. Yellow rectangles identify components of the statue composing of posture, head, body, robes, hand-gestures, feet and legs, and time of creation.

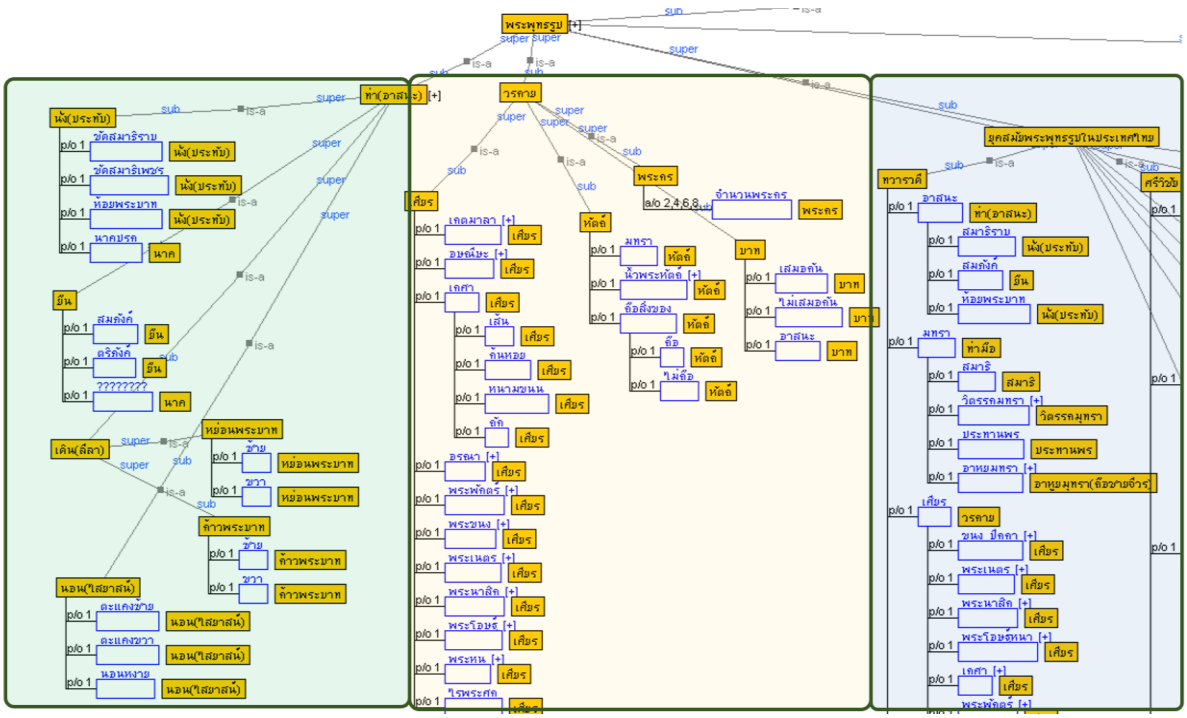


Fig. 4 Excerpt of "Buddha Image" ontology created by a domain expert using Hozo [4], the ontology editor in by an ontology application management (OAM) [3] framework. Note that, the first green rectangle represents concepts of Buddha posture, the second yellow rectangle represents main concepts of Buddha image, and the third blue rectangle represents concepts of Buddhist era.

The terms of Buddha Image's component were defined as *concept and concept's hierarchy* (*Stage 4*). As shown in Fig.4 each concept is *defined its properties and role* (*Stage 5*) by using Hozo [4]: the first green rectangle represents concepts of Buddha posture, the second yellow rectangle represents main concepts of Buddha image, and the third blue rectangle represents concepts of Buddhist era. Each concept is defined slot, for example, Buddha has two hands and we defined slots (*Stage 6*) for Part-Of properties "Hand" in "Buddha Image" concept as 2. Lastly, we let them gave examples of Thai Buddhist art's periods, such as Lanna, and Sukhothai (*Stage 7*).

C. Supportive Features based on the COD approach

In this workshop, we had opportunities to understand the domain knowledge of Buddha Image. However, we found that some part of the ontology in Fig.4 was not defined following the role-concept modeling, such as multiple roles. With limitations of time and budget, we selected an online supportive environment to communicate with the domain experts. Thus, the CD-OAM framework was used for achieving these collaborative activities. Details of supportive features based on the COD approach are explained as follows.

We first invited all participants to register their account into the CD-OAM framework. As mentioned in Fig.1, the framework provides a communication space in an application of communities' management. As illustrated in the Fig5., a "misunderstanding diagnosis" question was automatically available a member in a cultural domain of a discussion forum. The knowledge-based management module notified relevant stakeholder. Then, they were allowed to give comments or suggestion to members of the cultural community. The participant from culture posted the question about designing of the Buddha Image's ontology. The question inquired what kind of concepts to define Buddha posture. Finally, in part of supportive features, the system is automatically tagging the concept "Buddha Image" that related to the terminologies of the cultural domain. The framework also provided supportive functions: replying this post for giving suggestion or comments, voting an understanding about ontology, and (3) checking a statistical of domain terminologies within this post.

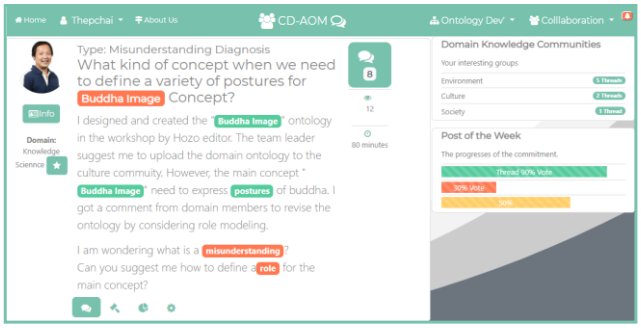


Fig. 5. Excerpt of a question of "misunderstanding diagnosis" that is automatically available a member in culture domain of a discussion forum in the CD-OAM Framework [5].

As a result of communication and collaborative activities, Fig.6 represents the revision of Buddha Image's ontology. The yellow circles represent the main concepts and details of each concept is as follows.

- "Buddha Image" concept has six important properties with roles including "Name", "Era", "Head", "Hand", "Foot", "Posture"
- "Era" concept has three important properties with roles including "Name", "Country", "Duration". This concepts also represented instances of Thai Buddhist art's periods including "Khmer Art found in Thailand", "Lopburi", "Ayutthaya", "Lanna", "Sukhothai", "U-Thong", "Rattanakosin" "Srivijaya", "Dvaravati"

Therefore, exploitation of the framework can support the domain expert in designing of Buddha Image's ontology, and also support knowledge engineer to communicate with the domain expert for revising the ontology.

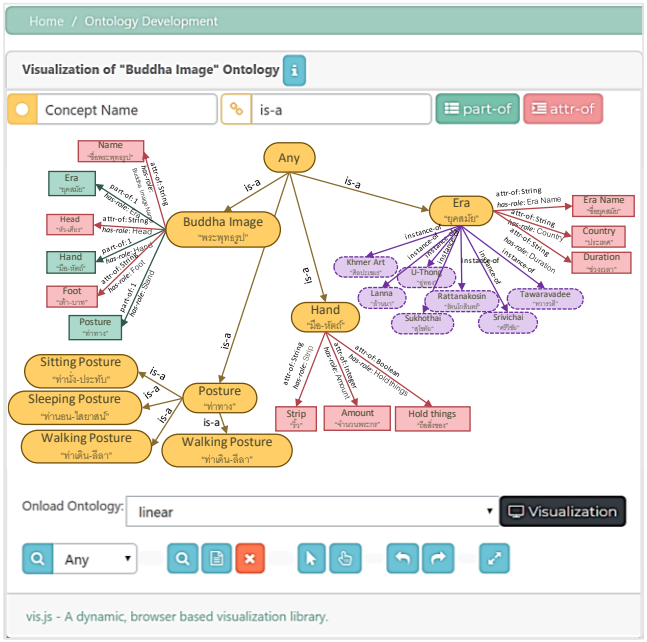


Fig. 6. Excerpt of Buddha Image's ontology that is visualized in a web application provided by a module of knowledge-based management of the CD-OAM Framework [5].

IV. Related Work and Discussion

Many supportive environments have been designed for archiving particular research goals and activities. Several supportive tools based on the COD approach have been designed and implemented. We distinguish our approach with others by comparing environments [2]: (1) management of development processes, (2) organizing of collaborative activities, (3) theory-awareness, (4) architecture, and (5) interoperability.

This paper presents the necessary features [8] following a collaborative approach in order to support ontology design and collaborative activities. A criterion of collaboration was applied to build multiple modules to support collaboration [2] allow in the COD approaches.

V. Conclusion

In this paper, we presented an improvement of collaborative activities through a supportive environment based on the collaborative ontology development (COD) approach, called a community-driven ontology-based application management (CD-OAM) framework. Our working approach contributed a significant improvement in supporting collaborative activities of stakeholder in ontology development and demonstrating a collaborative situation to overcome limitations of communication between domain experts and knowledge engineers.

For the future, our research direction aims to work with tagged concept and instances in order to build the knowledge graph [9]. The expected results will be to comprehend the insight and motivate stakeholder to work with a supportive environment.

Acknowledgment

This research is partially supported by National Electronics and Computer Technology Center (NECTEC), Thailand. Buddha image's materials and data are kindly provided by Faculty of Fine Arts, Chiang Mai University, Thailand.

References

[1] G. Stephan, H. Pascal, and A. Andreas, "Knowledge representation and ontologies," *Semant. Web Serv. Concepts, Technol. Appl.*, pp. 51–105, 2007.

[2] R. Mizoguchi and K. Kozaki, "Ontology Engineering Environments," in *Handbook on Ontologies*, Springer, 2009, pp. 315–336.

[3] M. Buranarach *et al.*, "OAM: An Ontology Application Management Framework for Simplifying Ontology-Based Semantic Web Application Development," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 26, no. 01, pp. 115–145, 2016.

[4] K. Kozaki, Y. Kitamura, M. Ikeda, and R. Mizoguchi, "Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of 'Role' and 'Relationship,'" in *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, 2002, pp. 213–218.

[5] A. Takhom, S. Usanavasin, T. Supnithi, and M. Ikeda, "Collaborative ontology development approach for multidisciplinary knowledge: A scenario-based knowledge construction system in life cycle assessment," *IEICE Trans. Inf. Syst.*, vol. E101D, no. 4, pp. 892–900, 2018.

[6] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Stanford Knowledge Systems Laboratory*. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA, p. 25, 2001.

[7] E. Sunagawa, K. Kozaki, Y. Kitamura, and R. Mizoguchi, "Organizing role-concepts in ontology development environment: Hozo," *Hozo, AI Tech. Rep. (Artificial Intell. Res. Group, ISIR, Osaka Univ, p 2004*, 2004.

[8] T. Slimani, "Ontology development: A comparing study on tools, languages and formalisms," *Indian J. Sci. Technol.*, vol. 8, no. 24, pp. 1–12, 2015.

[9] R. Popping, "Knowledge graphs and network text analysis," *Soc. Sci. Inf.*, vol. 42, no. 1, pp. 91–106, 2003.

Voltage Failure Warning Device for 3-Phase Transformer

LUECHAI PROMRATRAK

Abstract—**The purpose of this research is firstly to study the operation of the Arduino micro device and then to create a prototype of a 3-phase low voltage failure warning device. Second is to test effectiveness of the device with a 3-phase variable transformer via simulation. Finally, is to install the prototype in the low voltage distribution system of the Provincial Electricity Authority. The device has an Arduino micro R3 for data processing; ESP8266 expansion board for WIFI receiver/transmitter and EFDV434 board for interfacing with voltage sensing while displaying status message via line application when failure occurs. The experimental test results indicate that the prototype device has low voltage variation of ± 5.42% with actual accuracy greater than 90%. The warning can operate with the distance of 15 meters with sound levels between 80 to 87 dB.**

Index Terms— Low Voltage detection, 3-Phase transformer

I. INTRODUCTION

Hight temperature can cause electrical issues; i.e., power outages, voltage drop blow the damaging the system. These issues must be detected or corrected before equipment.

In order to detect low voltage in a timely manner, creating a device that could alert when failure occurs with some notification sound, light or messages should be considered.

II. RESEARCH OBJECTIVES

1. To study the operation of the Arduino device and to create a 3-phase voltage failure alarm prototype.
2. Test the effectiveness of the proposed prototype with a 3-phase variable transformer.
3. To implement the prototype in the low voltage distribution system of the Provincial Electricity Authority (PEA).

III. RESEARCH METHODOLOGY

The construction procedures would be to setup Arduino micro R3 board for data transmission. The Wi-Fi receiver/transmitter would be ESP8266 expansion board. The voltage sensing was using the EFDV434 interfaced to the Arduino micro, the sounder prompt, status lights and message work through application line software, when a voltage condition was a failure (low voltage). The device in test run mode the display would be show the voltage reading and set the alarms, when you have power outages through application line software you would be receive a message, a sound prompts(bell) and notification lights. The Arduino was still collecting data for analysis.

IV. CONCEPTS AND OPERATIONS

There are 2 main concepts and operations for this research: research concept and design; details are as follows:

1. Research concepts.

Study literature and other information from the knowledge organization in the form of a block.

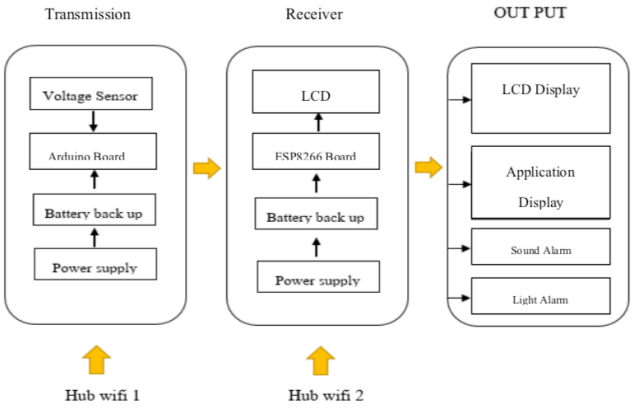


Figure 1. A Diagram shows components of the proposed warning device

From the diagram, the device has 3 parts, Part 1 is the section of the voltage sensor and signal sector that measures and then 3-phase voltage sends the measured data to the receiver. Part 2 is the receiver that contains an analog to display data obtained on the LCD screen and sends the data obtained through, and notify the voltage status. Part 3 is the display of voltage status, by the screen, line apps, sound and light various results.

2. Design and construction of the model.
2.1 Create a signal sector model by used a CAD software design and simulation as shown in Figure 2.

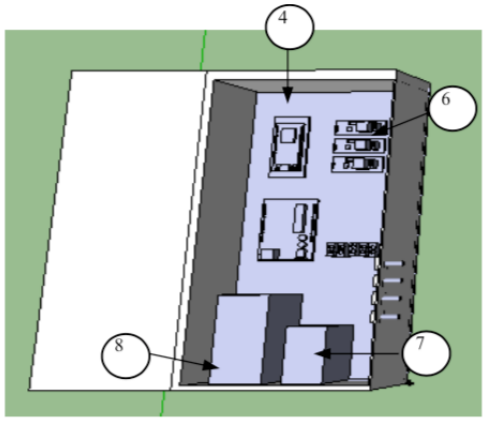


Figure 2. 3D transmitter signal model.

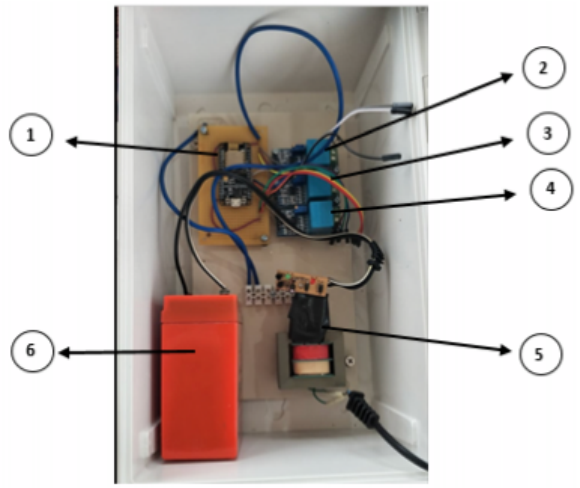


Figure 3. Installing the signal transmission model

1. The Arduino R3 board that receive data from the voltage sensor and transmits data to the receiver.
2. The voltage sensor1 which monitors the voltage phase 1.
3. The voltage sensor2 which monitors the voltage phase 2.
4. The voltage sensor3 which monitors the voltage phase 3.
5. 6 V. DC. power supply.
6. 6 V. - 4.5 AH NV. Battery in Figure 4.

2.2 Create a receiver sector model by used a CAD software design and simulation as shown in Figure 4.

2.

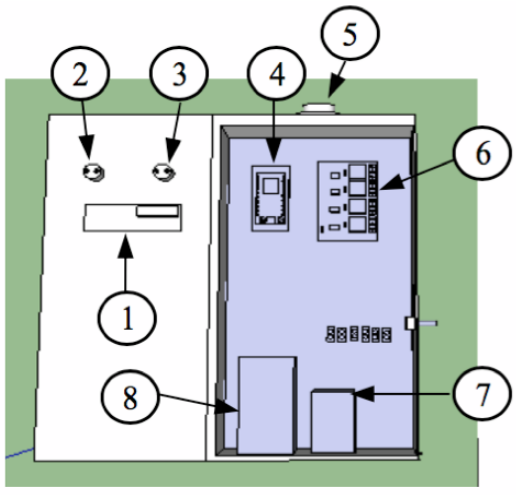


Figure 4. 3D signal receiver model.

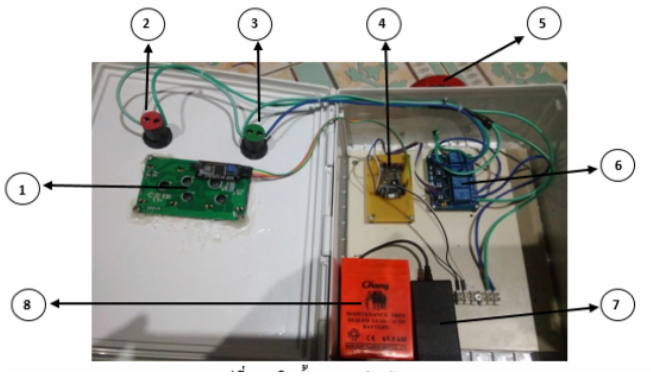


Figure 5. Installing the receiver equipment.

1. LCD screen: display 4 x 20
2. Red alert light: show operation voltage, failure occurs)
3. The green alert light: show operation voltage, normal conditions
4. The ESP8266 board: receiver board
5. The electric bell: notification sound show operating conditions
6. The relay 5V: on-off signal
7. Power supply DC. 6V: supply ESP8266 board
8. The battery is DC. 6V: Backup

V. Testing and results

The tests were divided into 3 tests:
1. Voltage measurement and display
2. Warning tests when failure occurs through the application line message
3. Sound and light warning tests

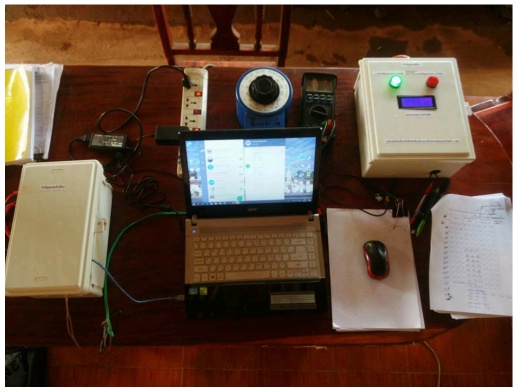


Figure 6. Experiment test-ring

VI. Test Results

The results from the experiment, shown in figure 7, which is the comparison of the voltage measured from the multimeter meter and device.

The low voltage power failure warning device of the 3-phase transformer to calculate the error percentage (% error value) in all 3-phases 15 voltage ranges from 100 - 240 V. The distance was 10 V. Each area could be summarized as the line graph.

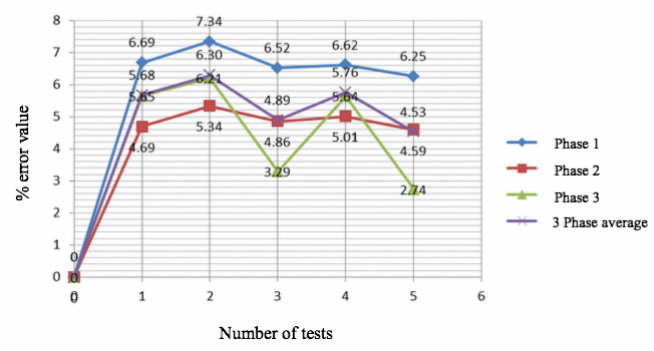


Figure 7. % errors of voltage measurement of the proposed device

The curves in figure 7 showed that phase 2 has lowest error with ± 4.8%; following by phase 3 and 2 of ± 5.2% and ± 6.4% respectively.

Figure 8 shows testing the notification when failure occurs through the application message. Test the transmission of the message through the 15 voltage neighborhoods, ranging from the voltage of 100 - 240 V. Using a voltage regulator, the regulator to adjust the voltage when a voltage failure occurs. The message would be send in the form of low voltage and returning to normal condition, the message would be send in the form of normal voltage as shown in figure 8.

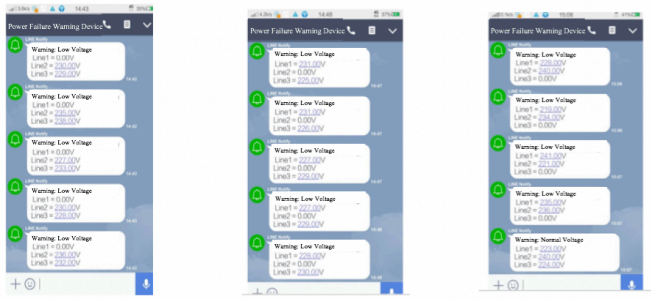


Figure 8. Messages in the line application when failure occurs

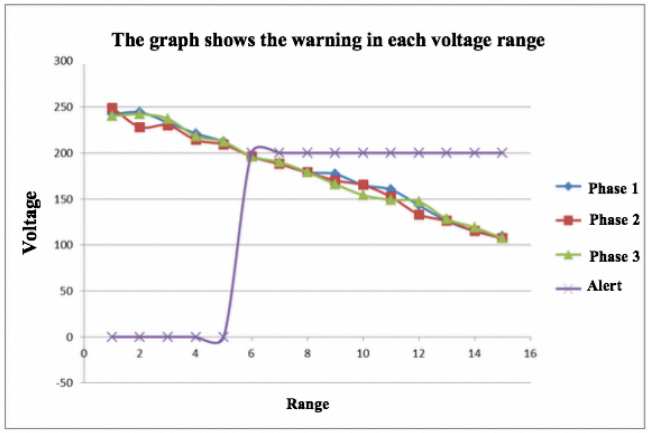


Figure 9. Graph of the test when the power failure occurs through the application message.

The purple line graph was the notification condition of the low voltage failure warning device of the 3-phase transformer. It could be seen from the range 0 - 200 V. When a power failure occurs low voltage failure warning device of the 3-phase transformer would be send a message to notify via the application line in according to the conditions. (Provincial Electricity Authority standard voltage, Emergency voltage is lower than 200 V). As a result, the operator corrects the voltage failure, knowing the voltage data from the application line group.

Test the notification sound and the notification light.

Table 1. Test for notification sounds and notification lights.

Time (Clock)	Voltage from the 3 phase electrical alarm device			Electric bell alarm sound (dB)			Notification via Line application	Notification light
	Phase 1	Phase 2	Phase 3	5 meters	10 meters	15 meters		
14.00	221	231	228	-	-	-	✗	
14.10	234	225	230	-	-	-	✗	
14.20	226	233	228	-	-	-	✗	
14.30	228	230	229	-	-	-	✗	
14.40	0	230	229	102.3	95.8	84.5	✓	
14.50	231	0	225	101.6	96.1	80.2	✓	
15.00	228	240	0	103.1	96.4	86.8	✓	
15.10	0	0	0	102.9	97	87.2	✓	
15.20	233	235	226	-	-	-	✗	
15.30	237	226	233	-	-	-	✗	
15.40	236	237	228	-	-	-	✗	
15.50	231	230	229	-	-	-	✗	
16.00	223	240	224	-	-	-	✗	
16.10	226	233	228	-	-	-	✗	
16.20	228	230	229	-	-	-	✗	
16.30	230	228	227	-	-	-	✗	

Random tests of the operation of the notification sound could work according to the conditions specified was alert when the voltage was lower than 200 V. made the electric bell work. The warning sound was in the distance of 15 meters. The volume were at 80 - 87 dB. fast notification.

Tests of the operation of the notification light notification from the notification light could work according to the conditions specified was notify when the voltage was lower than 200 V. The warning light would be displayed in red and alert the green light when the voltage was over 200 V.

VII. Results

The 3-phase voltage reading has an average error value of ± 5.42% according to the set-up conditions 200 V. There is error rang of 194.58-205.42 V. The device can alert the operator. Can fix the voltage failure.

Notification of 3 phase voltage messages at 205.42 V via the Application Line. Operation, correcting the voltage failure can be done in phase. The work is faster and more efficient.

VIII. References

Komdech Phayuephut. (2012). Iot NodeMCU Development Board, 7 September 2017.https://www.google.com

Luechai Phromratrak. (2015). Handbook for printing thesis. Bachelor of Engineering, Udon Thani Rajabhat University.

Nawapong Nutadei. (2013). Analysis of electromagnetic field strength in distribution system transformers. From external short circuits. Master of Engineering, Technology University Rajamangala.

Natthicha Wanta. (2017). Wireless network system, 12 September 2017. https://www.google.com/searc-h / biw + 1366 & bih

Natthawut Kwankaew. (2013). Voltage sensor, 7 September 2017. http: //www.thaieasyelec.com.down-loads / EFDV434 / single% 20phase% 20voltage% 20sensor.

Padungkit Sawang. (2016). Hazard warning system from electrical leakage in the distribution transformer real time digital processing. Electrical and Electronics Engineering, Engineering, Ubon Ratchathani Rajabhat University.

Padung Kitsawang. (2016). Real-time Alarming System of Dangerous Leakage Current at Distribution Transformer via Digital Signal Processing. The Journal of Industrial Technology, Vol. 12, No. 3 September – December 2016.

Panupong Pattatasing. (2015). NodeMcu ESP Module 8266, 7 September 2017. http://sat2you.com-/ web / 2017/0.

Pha Pha Phaeng-Ngai. (2013). Arduino R3, 9 September 2017. http://www.robotsiam.com/product/2/-ardu ino- uno-r3.

Warawut Khamin. (2008). Power Supply (Power Supply), 14 September 2017. http: //www.telep-art.net.

Sasikan Sutthinantarat. (2015). Hub and Switch, 7 September 2017. https://sites.google.com/site/kangtan75.

Somsak Meenakorn. (2011). Battery, 14 September 2017. https://www.amazon.com/APC-Back-UPS-Protector.

Suwit Boonkua. (2015). Arduino, 7 September 2017. http://morasweb.org/2016/06/15.

Asst.Prof.Dr. Luechai Promratrak

Department of Electrical Engineering,
Faculty of Technology,
64 UdonThani Rajabhat University,
Thahan Rd., Muang, UdonThani,
Thailand 41000

Gender Recognition from Facial Images using Local Gradient Feature Descriptors

Olarik Surinta
Multi-agent Inelligent Simulation Laboratory (MISL)
Faculty of Informatics, Mahasarakham University
Maha Sarakham, Thailand
olarik.s@msu.ac.th

Thananchai Khamket
Applied Informatics Group, Department of Information Technology, Faculty of Informatics, Mahasarakham University
Maha Sarakahm, Thailand
thananchai.k@msu.ac.th

***Abstract*—Local gradient feature descriptors have been proposed to calculate the invariant feature vector. These local gradient methods are very fast to compute the feature vector and achieved very high recognition accuracy when combined with the support vector machine (SVM) classifier. Hence, they have been proposed to solve many problems in image recognition, such as the human face, object, plant, and animal recognition. In this paper, we propose the use of the Haar-cascade classifier for the face detection and the local gradient feature descriptors combined with the SVM classifier to solve the gender recognition problem. We detected 4,624 face images from the ColorFERET dataset. The face images data used in gender recognition included 2,854 male and 1,770 female images, respectively. We divided the dataset into train and test set using 2-fold and 10-fold cross-validation. First, we experimented on 2-fold cross-validation, the results showed that the histogram of oriented gradient (HOG) descriptor outperforms the scale-invariant feature transform (SIFT) descriptor when combined with the support vector machine (SVM) algorithm. The accuracy of the HOG+SVM and the SIFT+SVM were 96.50% and 95.98%. Second, we experimented on 10-fold cross-validation and the SIFT+SVM showed high performance with an accuracy of 99.20%. We discovered that the SIFT+SVM method needed more training data when creating the model. On the other hand, the HOG+SVM method provided better accuracy when the training data was insufficient.**

Keywords—gender recognition, face detection, local gradient feature descriptor, support vector machine

I. Introduction

Gender recognition can be used to improve the efficiency of surveillance and security systems, authentication systems, and face recognition systems [1]. Moreover, it can also be developed into a variety of applications. Research in gender recognition involves with three major tasks; face detection, feature extraction (called *face encoding*) and recognition system [2]–[6].

Related work. In [7], the deep convolutional neural networks and support vector machines were proposed for gender recognition and tested on the ColorFERET dataset. The pre-processing step consists of detecting and cropping the face image. The face images after the detection stage consisted of 8,364 face images and stored at 256x256 pixel resolution. After that, the data augmentation technique is implemented to generate new face images. A pre-trained model of the AlexNet architecture was used to train the face images. The linear support vector machine is attached to the last fully connected layer. Using this method, the best accuracy was 97.3%.

In [3] a local feature descriptor called pyramid histogram of oriented gradients (PHOG) was proposed to represent a local gradient of the image. For the HOG descriptor, the feature vector is calculated according to Equation (1). Additionally, the PHOG descriptor allows dividing an image into a small block at several pyramid levels [8]. The gradient orientations in every level are stored into orientation bins. Then, all of the orientation bins in each pyramid levels are combined. The feature vector is then classified using the SVM classifier with the RBF kernel. The proposed method achieved an accuracy of 88.5% on the labeled faces in the wild (LFW) database.

Also, in [5] proposed multiscale facial fusion feature; however, the multiscale method is related to the pyramid technique [3], [8]. The fusion features used in the experimented, including local phase quantization (LPQ) and local binary pattern (LBP) descriptors. The combination of the feature vector is extracted from two descriptor methods and sent to the SVM classifier to classify the face image. The multiscale facial fusion method obtained an accuracy of 86.11% on the images of groups (IoG) dataset.

In this paper, we first applied a well-known Haar-cascade classifier, which was invented by Viola and Jones [9], [10] that proposed for object and pedestrian detection, to first find the exact location of a face from the complete image. Note that we focus only on the frontal face, and we ignore the profile face if the head of the people is turned to left or to right. Due to the challenge of the ColorFERET dataset [11], [12], we can extract only 4,624 face images from the 11,119 images. After that, all face images were resized to the same size. The face image resolution used in the experiments was 88x80 pixels.

Secondly, two local gradient feature descriptors called the histogram of oriented gradients (HOG) [13] and the scale-invariant feature transform (SIFT) [14] descriptors are proposed to extract the gradient feature from the face image. We experimented with the performance of the local gradient descriptors using several parameters. We set up the parameters of the HOG descriptor; orientations, pixels per cell, and cell per block and the SIFT descriptor; patch size.

Finally, the support vector machine (SVM) [15] with the radial basis function (RBF) kernel is proposed to create a model of the gender feature vector from the training data. We implemented the grid-search method to discover the hyper-parameters (C and γ) [16] until obtaining the best optimize parameters were obtained. Also, the average accuracies and

the standard deviation were used to compare the experimental results.

Contributions. In this paper, we proposed two well-known local gradient feature descriptors; the HOG and SIFT descriptors, to compute the invariant feature vector from the face images. These local gradient feature descriptors are designed to extract features from the gradient image for object detection purposes. The feature descriptor combined with the SVM with the RBF kernel is presented to address the gender recognition problems. The results show that our proposed method achieves very high recognition accuracy.

Paper Outline. The rest of the paper is presented as follows: In Section II, the gender recognition method, which is proposed is explained. In Section III, experimental settings and the results are presented. The conclusion and future work are given in Section IV.

II. Gender Recognition Methods

A. Face Detection

For face detection, the Haar-cascade classifier was proposed by Viola and Jones [10] in 2004. This method, the Haar features were used to compute the feature vector. The sub-window scans within the image to capture the small image. Then send the data of the sub-window to calculate the feature vector. Consequently, the feature vector was trained and predicted with the AdaBoost algorithm.

B. Local Gradient Feature Descriptors

To study the effectiveness of local gradient feature descriptors for gender recognition, we compare two well-known gradient features, called *the histogram of oriented gradients* and *the scale-invariant feature transform*. In this study, the face images are resized to 88x80 pixel resolutions.

1) Histogram of Oriented Gradients (HOG): The HOG descriptor was invented by Dalal and Triggs [13] in 2005 for detecting a pedestrian in an image. The basis of this technique is to compute the gradient orientation from small connected regions of an image. The features that calculated from this technique are robust to the light and geometric changes [17].

The notation of the HOG method [18] can be written as follows:

$$\Phi_f(\mathrm{X}) = Db * \left[\left(g_f * \mathrm{X}\right) \odot \left(g_f * \mathrm{X}\right)\right] \qquad (1)$$

where X is an input image and $\mathrm{X} \in \mathbb{R}^D$.

First, X is convolved with the simple convolution kernel g_f in the horizontal and vertical directions.

Second, blurred with b and the nonlinear transform ($\odot$) is applied to removes sensitivity to edge contrast and increases edge bandwidth. Third, the gradient orientations are weighted and stored into orientation bins Db.

Finally, The histograms from each block are describes as the feature descriptor. Then, the L2-Normalization is used to normalize the feature descriptors. The equation of L2-Normalization [16] can be written as follows:

$$V_k' = \frac{V_k}{\sqrt{\|V_k\|^2 + \varepsilon}} \qquad (2)$$

where

V_k is the histograms from all block regions
ε is a very small value and close to zero
V_k' is the normalized feature descriptor.

2) Scale-Invariant Feature Transform (SIFT): The SIFT method was proposed by Lowe [14] in 2004 for extracting invariant features from images. The features are invariant to image scale and rotation. The complete process of the SIFT method consists of scale-space extrema detection, keypoint localization, orientation assignment, and the local image descriptor. The complete SIFT method is applied to localization of the object in the target image.

3) In this paper, we focus only on computing the invariant feature, called *the SIFT descriptor*. First, the Gaussian kernel is used to convolution the image, I.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (3)$$

where

$I(x, y)$ is the pixel at location x, y of image I
$G(x, y, \sigma)$ is the Gaussian kernel, and σ determines the width of the Gaussian kernel.

Second, the gradient orientation $\theta(x, y)$ and magnitude $m(x, y)$ are computed from the image, $L(x, y)$:

$$G_x = L(x+1, y, \sigma) - L(x-1, y, \sigma) \qquad (4)$$
$$G_y = L(x, y+1, \sigma) - L(x, y-1, \sigma)$$

where G_x is the horizontal and G_y is the vertical components of the gradients.

Third, a sliding window method is used to slide through the whole image to capture a small region. Then all regions are sent to the SIFT descriptor to extract the gradient orientations and magnitudes.

Finally, each region is divided into 4x4 equal blocks. Then an orientation histogram is created for each block. Each histogram uses 8 bins to store the orientation values, which results in 128 dimensions for each region.

C. Classifier Method

The *support vector machine (SVM)* algorithm is a supervised learning algorithm employed for recognizing the feature that is extracted from data. The SVM algorithm,

invented by Vapnik [15], it has been successfully applied to many pattern recognition problems [16], [19] such as image classification, handwritten recognition, face detection, and face recognition.

The SVM algorithm is first created for binary classification problems [20]. This technique finds the function $g(\cdot)$ that is the best separation to the pattern data, called *hyperplane*.

The training set is (x_i, y_i) where $x_i \in \mathbb{R}^n$ and the output label are either +1 or -1, $y_i \in \{+1, -1\}$. It can be split by the following:

$$g(\mathbf{x}) = \boldsymbol{w}^T \cdot \mathbf{x} + b \quad (5)$$

where $\mathbf{x}$ is the weight vector and b is the bias value.

The largest distance between the nearest positives $\boldsymbol{w}^T \cdot \mathbf{x} + b = +1$ and negatives $\boldsymbol{w}^T \cdot \mathbf{x} + b = -1$ is the optimal separating hyperplane.

Moreover, the SVM can be extended to deal with non-linear data. Then, the soft constraint is proposed:

$$y_i(\boldsymbol{w}^T \cdot \mathbf{x} + b) \geq 1 - \xi_i \quad (6)$$

where ξ_i is the slack variable for data $\mathbf{x}_i$

The radial basis function (RBF) kernel is a non-linear similarity function and employed in the SVM classifier. In this kernel, the similarity value between the two input vectors are computed as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (7)$$

where γ is the parameter of the RBF kernel, note that the model can be overfitting when the γ parameter is too large because it increases the number of support vectors.

III. Experimental Settings and Results

In this section, we concisely describe the face image dataset used in the experiments. The experimental results consisting of the face detection results, parameter settings, grid search parameter estimation, and gender recognition results, are presented and discussed.

A. Face Image Dataset

In the experiments, we used a benchmark face image dataset, called the color face recognition technology (ColorFERET) dataset. Firstly, we used ColorFERET for face detection purpose. Secondly, we divided the ColorFERET dataset into training and test sets using 2-fold (50:50) and 10-fold (90:10) cross-validation, respectively, for gender recognition.

The ColorFERET was introduced by J. Phillips and P. Rauss [11], [12] for a facial recognition system. This dataset consists of 14,126 face images from 1,199 subjects. The resolution of images in the dataset is 384x256 pixels. An example from the ColorFERET dataset is shown in Fig. 1.

B. Experimental Results

1) Face Detection Result:

The evaluation method of the face detection is given by:

$$Acc_{fd} = Ac_{fd} - Er_{fd} \quad (5)$$

when

$$Ac_{fd} = \frac{c*100}{N}, \qquad Er_{fd} = \frac{e*100}{N} \quad (6)$$

where

c is the number of face images, after using a face detection technique.

e is the number of the error face images

N is the total number of the face images in the dataset.

In this paper, we proposed to use the Haar-cascade classifier for the face detection process.

Firstly, we used the Faced face detection method to detect face from the ColorFERET dataset. The result showed that the Haar-cascade classifier significantly outperformed the Faced method. The Faced method was a time-consuming when compared to the Haar-cascade classifier method.

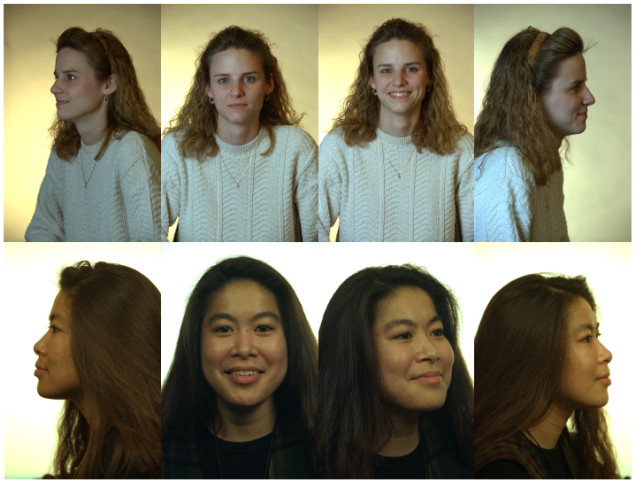
Secondly, we experimented with face detection using the Haar-cascade classifier. In the ColorFERET dataset, there are 13 different poses of each person, such as regular frontal image, profile left, half left, quarter left and also head turned to left and right between 15-75 degree as shown in Fig. 1. This method obtained an accuracy of 39.25% on ColorFERET dataset. The accuracy of the male and female faces was 36.87% and 41.63%, respectively.

Based on our experiments, the Haar-cascade classifier performed not very well when the face images turn to left or right, so the detection rate was quite low. On the other hand, this detection technique performed quite very well and fast when the regular frontal image were used. The results of the Haar-cascade face detection are shown in Table I.

TABLE I. Performance of the Face Detection using Haar-cascade Classifier on the ColorFERET Dataset

Gender	Number of male images	Number of face detected	Number of error detected
Male	7,139	2,854	222
Female	3,980	1,770	113

(a)



(b)

Fig. 1. Example face images of (a) male and (b) female from the ColorFERET dataset.

1) Parameter Settings:

We evaluated the performance of the HOG and SIFT descriptors using several parameters. The parameters of the HOG descriptor included orientations, pixels per cell, and cell per block [17]. The parameter of the SIFT descriptor comprised only patch size.

Note that the pixels per cell and cells per block of the HOG parameters and the patch size of the SIFT parameter are defined as a square. We use the SVM with the RBF kernel as a classifier and using the default C, γ parameters to find the best parameters of the HOG and SIFT descriptors. Also, 10-fold cross-validation over the training set was applied.

The best parameters of the HOG descriptor uses as 9 orientations, 8 pixels per cell, and 3 cells per block. The SIFT descriptor used patch size = 25 pixels. The accuracy results of the HOG and SIFT descriptors are shown in Table II and III.

(a)

(b)

Fig. 2. Sample face images of (a) male and (b) female after applying Haar-Cascade classifier from the ColorFERET dataset.

TABLE II. THE PERFORMANCE OF DIFFERENT HOG DESCRIPTOR PARAMETERS

HOG Descriptor Parameters			
Orientations	**Pixels per cell**	**Cells per block**	**Accuracy (%)**
4	8	1	94.6
8	16	1	92.2
8	16	2	92.8
9	8	1	94.8
9	**8**	**3**	**95.8**
9	16	1	93.3
24	16	1	92.0

2) Grid Search Parameter Estimation:

From the parameter settings section, the best feature descriptor parameter values were selected. Consequently, we have optimized the hyper-parameters of the SVM classifier with the RBF kernel. The grid search parameter method is suggested. We searched the hyper-parameter C and gamma between the number of 2^{-7} and 2^{7}. The best hyper-parameters found for our experiments are shown in Table IV.

TABLE III. THE PERFORMANCE OF THE SIFT DESCRIPTOR USING DIFFERENCE PATCH SIZES

SIFT Descriptor Parameters	
Patch sizes	**Accuracy (%)**
10	97.8
20	98.2
25	**98.4**
30	97.1
40	97.1
45	97.8
50	96.9

TABLE IV. THE BEST HYPER-PARAMETER VALUES FOR THE SVM CLASSIFIER WITH THE RBF KERNEL

Methods	C	γ
HOG	2^3	2^0
SIFT	2^3	2^{-5}

3) Gender Recognition Results:

The calculation of gender recognition accuracy is computed by multiply 100, with the total number of correct prediction and divided by the total number of face images in the dataset.

From the face detection result, we divided 4,624 face images into train and test sets with the ratio of 50:50 (2-cv) and 90:10 (10-cv). On this face images dataset, 2-fold cross-validation achieved accuracy of 96.50% when using the HOG descriptor. On the other hand, when performing the system with 10-fold cross-validation, the SIFT descriptor outperforms the HOG method with the accuracy of 99.20, which is the highest result based on our experiments. The accuracy results of gender recognition are shown in Table V.

IV. CONCLUSION

The main objective of this paper is to recognize gender (male and female) from facial images. First, the Haar-cascade Classifier was used to find the face from the whole image. Second, the face images were then assigned to the local gradient feature descriptors; the histogram of oriented gradients (HOG) and scale-invariant feature transform (SIFT) descriptors, to compute the feature vector.

TABLE V. THE ACCURACY (%) OF THE SVM CLASSIFIER OBTAINED WITH 2-FOLD AND 10-FOLD CROSS-VALIDATIONS

Methods	**Accuracy (%)**	
	2-cv	**10-cv**
HOG	96.50 ± 1.8	98.75 ± 2.5
SIFT	95.98 ± 0.4	$\mathbf{99.20 \pm 0.8}$

Finally, for gender recognition, finally, the invariant feature vector was classified using the support vector machine (SVM) with the radial basis function (RBF) kernel. From the experimental results, the SIFT descriptor outperformed the HOG descriptor when combined with SVM with RBF kernel. This method obtained very high recognition accuracy.

In future work, we plan to work on the deep convolutional neural network to detect the face (even from different poses) and compute the invariant feature vector. We also want to study the effect of the data augmentation to generate the new face images.

ACKNOWLEDGMENT

This research was supported by the Faculty of Informatics, Mahasarakham University, Thailand.

REFERENCES

[1] S. C. Nistor, A.-C. Marina, A. S. Darabant, and D. Borza, "Automatic gender recognition for 'in the wild' facial images using convolutional neural networks," in *Intelligent Computer Communication and Processing (ICCP), 13th IEEE International Conference on*, 2017, pp. 287–291.

[2] S. Yousefi and M. Zahedi, "Gender recognition based on SIFT features," *Int. J. Artif. Intell. Appl.*, vol. 2, no. 3, pp. 87–94, 2011.

[3] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, and S. Mahmood, "Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor," *J. Inf. Commun. Technol.*, vol. 14, no. 1, pp. 111–122, 2015.

[4] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," *2016 13th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2016*, no. November 2017, pp. 235–241, 2016.

[5] C. Zhang, H. Ding, Y. Shang, Z. Shao, and X. Fu, "Gender classification based on multiscale facial fusion feature," *Math. Probl. Eng.*, vol. 2018, pp. 1–6, 2018.

[6] Z. Stawska, "Support Vector Machine in Gender Recognition," *Inf. Syst. Manag.*, vol. 6, no. 4, pp. 318–329, 2017.

[7] J. van de Wolfshaar, M. F. Karaaba, and M. A. Wiering, "Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition," in *IEEE Symposium Series on Computational Intelligence*, 2015, pp. 188–195.

[8] P. P. Saragni, B. S. P. Mishra, and S. Dehuri, "Pyramid histogram of oriented gradients based human ear identification," *Int. J. Control Theory Appl.*, vol. 10, no. 15, pp. 125–133, 2017.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on*, 2001, vol. 1, pp. 511–518.

[10] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[11] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-

recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.

[12] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on*, 2005, pp. 886–893.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[15] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[16] O. Surinta, M. Karaaba, L. Schomaker, and M. Wiering, "Recognition of handwritten characters using local gradient feature descriptors," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 405–414, 2015.

[17] X. Li and Z. Lin, "Face Recognition Based on HOG and Fast PCA Algorithm," in *Intelligent Data Analysis and Applications, the Fourth Euro-China Conference on*, 2017, vol. 682, pp. 10–21.

[18] H. Bristow and S. Lucey, "Why do linear SVMs trained on HOG features perform so well?," 2014.

[19] V. Codreanu *et al.*, "Evaluating automatically parallelized versions of the support vector machine," *Concurr. Comput. Pract. Exp.*, vol. 28, no. 7, pp. 2274–2294, 2014.

[20] O. Surinta, L. R. B. Schomaker, and M. A. Wiering, "A Comparison of Feature Extraction and Pixel-Based Methods for Recognizing Handwritten {B}angla Digits," in *Document Analysis and Recognition (ICDAR), The 12th International Conference on*, 2013, pp. 165–169.

Develop the Framework Conception for Hybrid Indoor Navigation for Monitoring inside Building using Quadcopter

S. Khruahong
Department of Computer Science and Information Technology, Faculty of Science,
Naresuan University, Thailand
sanyak@nu.ac.th

O. Surinta
Multi-agent Intelligent Simulation Laboratory (MISL)
Faculty of Informatics,
Mahasarakham University, Thailand
olarik.s@msu.ac.th

***Abstract*— Building security is crucial, but guards and CCTV may be inadequate for monitoring all areas. A quadcopter (drone) with manual and autonomous control was used in a trial mission in this project. Generally, all drones can stream live video and take photos. They can also be adapted to assist better decision-making in emergencies that occur inside a building. In this paper, we show how to improve a quadcopter's ability to fly indoors, detect obstacles and react appropriately. This paper represents a new conceptual framework of hybrid indoor navigation ontology that analyzes a regular indoor route, including detection and avoidance of obstacles for the auto-pilot. An experiment with the system demonstrates improvements that occur in building surveillance and maintaining real-time situational awareness. The immediate objective is to show that the drone can serve as a reliable tool in security operations in a building environment.**

Keywords—semi-autonomous quadcopter; indoor navigation; object detection; image processing; ontology

I. INTRODUCTION

Buildings, such as schools, universities building, office buildings, or shopping malls, etc. are guarded by staffs who monitor both inside and outside the buildings. They are concerned about preventing all dangerous situations. Some buildings need to be high security and may require much investment in guards and technologies. Such buildings will have Closed-Circuit Television (CCTV) and an operations room for monitoring and controlling the situation. However, the CCTV may not cover all area of the buildings, or there may be blind spots in the CCTV coverage. Technology to check the blind spots is needed to increase building security.

A quadcopter or drone [1] is a popular technology for taking photos and video, usually used for outdoor missions. The quadcopters can fly under user control or be autonomous. Although normally used outdoors they can be adapted for indoor missions. However, an indoor environment presents difficulties, especially where the building has many floors, and each floor has many objects, both static and moving. The quadcopter should be able to fly to a given destination anywhere in a building while avoiding obstacles (people, furniture) in its path. A hybrid or semi-autonomous approach for controlling the quadcopter may be appropriate.

In this paper, we discuss the indoor use of a quadcopter for patrolling a building. A quadcopter's route inside a building may be different from a person's; it can fly above head-height and can fly to different floors of the building. However, to increase the speed of quadcopter to arrive in the situation area, it needs an indoor navigation route which we have developed using an ontology method which can provide a more accurate, robust flight path. Furthermore, it needs to detect the properties of operating stability. Our approach is beneficial for the building guards as well. If they see a suspicious situation on CCTV in the building but have not enough information, they can send the quadcopter to that location to view the situation in order to decide on a future course of action. They do not need to learn how to use and control the quadcopter; they can use our application.

We propose a hybrid approach leading to an improved real-time situational awareness, as shown in Figure 1. Firstly, we have developed an analysis of the best route for the quadcopter in the building with indoor navigation ontology providing the optimum flight path. As GPS does not work inside the building, a Bluetooth Low Energy device (BLE) [2] is adapted for calculating the current position of the quadcopter. Secondly, while the quadcopter is flying, it detects some objects which may affect its flight mission. Our approach includes obstacle detection by using image processing for identifying the objects and avoiding them. While the quadcopter is in flight, it can communicate and receive real-time flight information from the control room via Wi-Fi in the building. We believe that our approach may improve security analytics and threat intelligence, enhancing the security of the building.

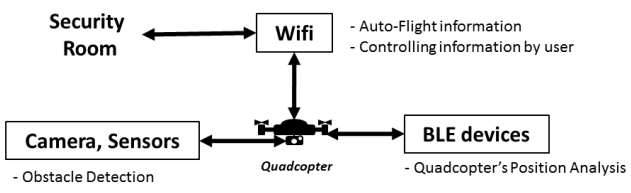


Fig. 1. Overview of the hybrid framework of the quadcopter

The section of this article is as follows. In Section II, we review the literature on quadcopters, ontology and obstacle

detection. In Section III, we show hybrid indoor navigation for an indoor quadcopter. In Sections IV, we detail our conceptual approach. Section V presents our experiment. Part VI discusses our results and proposes future research directions.

II. Relate Work

In this section, we discuss some relevant related work on indoor navigation involving indoor quadcopter, indoor ontology, and obstacle detection.

A quadcopter or quadrotor [1] is a helicopter with four rotors. They are used for surveillance [3], construction inspections, or mission in the farms [4-6] etc. A quadcopter was developed to inspect the vertical infrastructure of building [7] but is limited to inspection of the exterior of the building. An AR-Drone was adapted for robotic research [8], discussing technologies on position stabilization, autonomous navigation, etc. However, it was not developed for an indoor situation. I. Sa and P. Corke detail quadcopter for use indoors and outdoors [9], but the user needs to learn and attain the skills to use it and also needs to know the structure of the building in which the quadcopter will be flown. SmartCopter is a technique for controlling a quadcopter without GPS; it can automatically fly both outdoors and indoors by using vision-based tracking [10], but vision-based tracking may not be sufficient for autonomous flight. A Camera Measurement Algorithm was used for estimating distances in a building [11]. However, this approach may be too slow for processing for indoor navigation where the requirements of the mission need a fast response.

Ontology is developed within many research fields [12]. J. Scholz and S. Schabus propose an Indoor Navigation Ontology which is used for movement of production in an indoor environment [13]. They designed the elements of the ontology for representing the indoor space. However, their article focused only on the indoor production for autonomous navigation in the indoor space. Web ontology can be combined in an indoor navigation system called OntoNav [14, 15], in which both the navigation paths and the guidelines are presented to users to develop an Indoor Navigation Ontology (INO) [16]. Nevertheless, in this article can find the route for the specific user profile for the recommendation the best route to them, which need to collect the user profile, if it applies to our research, may need to maintain the quadcopter attributes for being suitable.

Obstacle Detection is applied to many techniques for navigation such as using image processing for the autonomous micro aerial vehicles [17]. This article detail that it did not focus on indoor navigation. Obstacle Detection and the 3D indoor model are developed for indoor navigation by using the laser scanner [18], this method may difficult to collect the building planning information for creating a route from the building structure. Similarity, the 3D model is designed for navigation for autonomous vehicle [19], but his approach focuses on the outside navigation. Computer vision was used for indoor autonomous drone racing, but this article needs more information with using Deep Learning technique [20]. Imaging geometric relationship [21] is used for obstacle detection to navigate inside the building by using four cameras based on the bird's eye view images; we will apply this approach to our research.

After a literature review, we found that the semi-autonomous quadcopter will be used for our research, because it can be automatic fly by it after the user selects the destination inside the building and can control by the building guards, especially, controlling the camera on the quadcopter when arriving the target location. Our conceptual framework consist of two contributions are analysis the indoor route and the obstacle detection.

III. Hybrid Indoor Navigation for Indoor Quadcopter

This study, we describe the hybrid indoor navigation model monitoring indoor security by using a semi-autonomous quadcopter. While the quadcopter is flying on the mission, it will communicate with the security room via the internet by Wi-Fi network in the building. It can send to images, VDO and position information back to guards. The quadcopter's position can lead to present to the current position on the building map. The quadcopter begins and finishes the mission at the security room.

In section includes Indoor Quadcopter Position, Indoor Navigation Ontology, and Obstacles Detection.

A. Indoor Quadcopter's Position

Global Positioning System (GPS) for calculating the position is not working inside the building. We use Bluetooth Low Energy devices (BLE) for analysis the quadcopter's position [22]; these devices are called "iBeacon." These BLEs will be installed in the building, and one each will be set on the quadcopter, they are used for distance measurement and help to us for calculating the quadcopter's position in the tower as shown Figure 2.

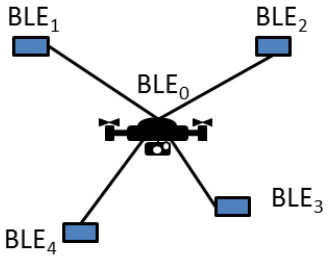


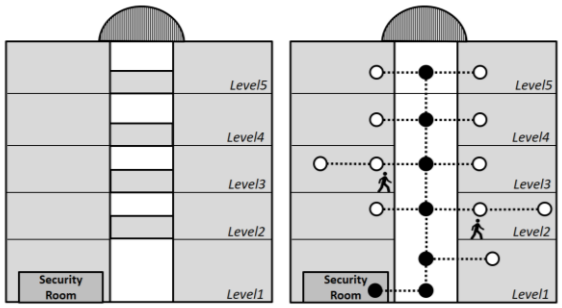
Fig. 2. BLE for analysis the quadcopter's position

B. Indoor Navigation Ontology

The building has some different floors plan and some levels. Normally, almost towers have the corridor for people walking inside the building, and some buildings have the lift and escalator. In this research, we focus on the flying route of the quadcopter. Therefore, the route will use any airspace for flying direction in the building. We develop the coordinate on the airspace for linking to another, vertical view, as shown in Figure 3 and map view on the floor as illustrated in figure 4.

The quadcopter's direction is not similarity with human way. Therefore quadcopter path will be designed in the position where it can fly and can link o different floor. The

indoor navigation ontology is applied to creating the path of flying. The ontological foundations are used to a model of navigation on indoor space. This technique can increase the speed of quadcopter for flying the destination. Guard will set the target before quadcopter going. The quadcopter will fly follow on the coordinate and follow up until reaching the target location, which all coordinate will be set value as flight information for travelling to the next coordinate.



(a) (b)

Fig. 3. Example of (a) the instruction of indoor building where has any airspace (b) indoor route of semi-autonomous quadcopter inside the building

This approach is flexible for maintenance because all routes are set on the ontology as a dynamic database. If the building was renovated, the user could change data on the ontology for updating the fly direction. The example of primary attributes on the ontology is designed as shown in Table I.

TABLE I. TERMINOLOGIES FOR SOME ATTRIBUTES IN INDOOR ONTOLOGY FOR INDOOR QUADCOPTER

Name	Description
ID_dro	ID is a unique label for a coordinate for quadcopter (droBLD1.lev1.cr01)
x, y, z	(x, y, z) in Euclidean air space inside buildig (x=1500,y=560,z=195)
Defualt_direction	The default position of quadcopter when arriving this coordinate, the quadcopter will be set the direction about inspecting point as same as compass degree (352)
Building	Building Name (Bld1)
Level	Level of building (level3, level5)
Status	Status of a coordinate on the map (On, Off)

C. Obstacle Detection

Commonly, most quadcopters have the sensor for protecting when flying nearly some objects, but not enough for it. In the building, some properties are caused by flight problems, especially humans or cabinets. These obstacles are over control because they can move to anywhere in the building. Although the flying route is set for auto-pilot, these objects may affect the quadcopter and lead to decreasing speed and have accident flight. Therefore auto obstacle detection on the real-time is critical for the semi-autonomous quadcopter.

All level of the building, the flight route will be designed on the map of the construction plan, as shown in Figure 4. The quadcopter can fly on the flight line with the typical situation. However, it needs to get the object detection for instantly stabilized flight. The obstacle detection recognizes the objects for getting the size and dimension of them by using image processing. This research focuses on the detection of object colour.

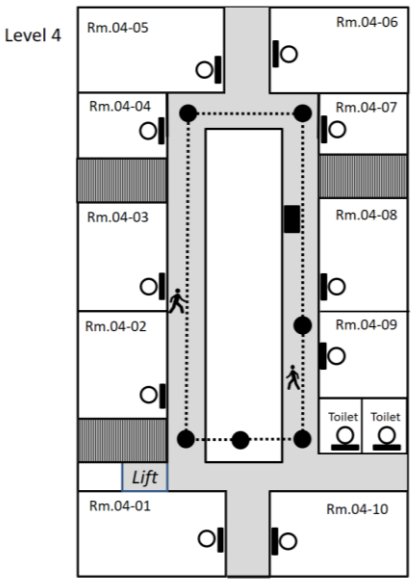


Fig. 4. For an example of direction and coordinate in the building

IV. THE CONCEPTION OF AN ALGORITHM FOR INDOOR QUADCOPTER

Our approach is using the quadcopter to fly inside the building for the indoor mission. We describe the step of an algorithm for monitoring indoor security. The building has different information for a guide to indoor travel; design should follow the real indoor space which focuses on the quadcopter flight. Figure 5 is an example of a building.

Fig. 5. Example, inside the building

The flight mission will analyze the route by the algorithm and then send information to the quadcopter. The flight

information will set as the array; it consists of the coordinate information. Each coordinate has an attribute for supporting the fight until to the destination; this approach can help the speed of the flight, as shown in Figure 6. It is a drone route in the building, and it starts from the security room on level one to room number *Rm.04-07*. It will follow node of route. However, the calculation of quadcopter's position is crucial for stable flying. Our algorithm uses four BLE devices for calculating the current quadcopter's position [22, 23]. This approach looks like finding the position of the satellite, which can lead to development to show the current position of the quadcopter on the application.

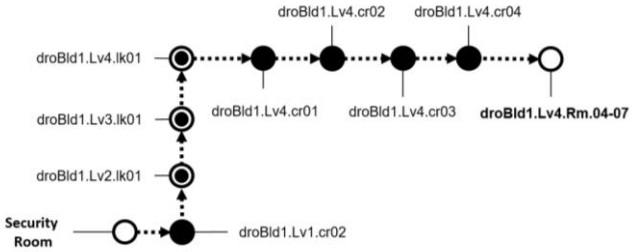


Fig. 6. The direction of quadcopter's flight

While the quadcopter arrives at each coordinate, it will get the value of attributes; this information can provide analysis for flight controlling data, for still flying to next coordinate in the direction mission. This information can help to control the flying for the *front, back, left, right, up, down, clockwise and counter-clockwise*. This controlling data communicates between the guard room and quadcopter via the internet with Wi-Fi technology.

Finally, our approach includes obstacle detection, while the quadcopter is flying on the route; we improve flight stability by using image processing for object detection. It analyses the dimension of objects for avoiding them. After that, it will fly back in the same mission route until the destination area. Also, finished task, it will use the same path for the flight back to the security room. This research develops the simple web application for controlling and monitoring with our algorithm.

V. Experiment and Discussion

We design the experiment for validation, analysis the route and obstacle detection with the simple web application. This research, we use the Quadcopter or AR-Drone 2.0 in our research, as shown in Figure 7. It flies high precision control and automatic stabilization features which be suitable for development to the indoor building. We separate to be two testings, consists of auto flight on the route and obstacle detection.

The "Estimote Location" devices [22] are Bluetooth Low Energy devices which can use to support to measure the distance inside the building. We use the stick on the wall in the building where they can send the signal to BLE on the quadcopter.

Fig. 7. Indoor quadcopter, model is AR-Drone 2.0 (*Image reference: https://www.parrot.com*)

The indoor route is designed with indoor navigation ontology. In the programming of the algorithm, the route will be set in the array. Our experiment, we set three coordinates for every route with five routes for testing. For example, the first path includes *droBld1.LV1.cr02*, *droBld1.Lv2.lk01*, *droBld1.Lv3.lk01*, *droBld1.Lv4.lk01*, *droBld1.Lv4.cr01*, *droBld1.Lv4.cr02*, *droBld1.Lv4.cr03*, *droBld1.Lv4.cr04* and *droBld1.Lv4.Rm.04-07*. The algorithm will use coordinate information for flight navigation which the values will be sent to quadcopter for flight control. Python was developed to be the simple application for controlling the quadcopter with auto-flight. Our validation uses the quadcopter to fly to a destination for one way direction. After that, we determine from the distance of closely flying the coordinates, where is set up on the route. The result is shown in Table II. The quadcopter can fly to pass all three coordinates on five routes.

TABLE II. The result of the flight on the route

Routes No.	The distance of Quadcopter with Coordinate(meters)		
	Coordinate No.1	*Coordinate No.2*	*Coordinate No.3*
1	1.5 meters	1.3 meters	1.5 meters
2	0.8 meters	1.5 meters	1.2 meters
3	1.5 meters	1 meters	1.5 meters
4	2 meters	1.5 meters	2 meters
5	1.5 meters	2 meters	1.5 meters

Indoor flight has narrow space for flying, with the result in Table II; the precision of position on the route is not 100%. The result show flight of quadcopter missing from coordinate around 0.8-2 meters. Therefore, the autonomous flight of quadcopter needs to use some approach to be stable fly.

On the other hand, although AR-Drone 2.0 will include sensors for objects detection, it may not be enough for a quick flight in the environment. This research, we present that increase a part of obstacle detection, the object is detected by the camera on the quadcopter with the colour analysis by using the image processing technique as shown in Figure 8. We set the distance between camera and object for photos collection around 0.5, 1, 1.5, 2, 2.5 meters respectively. After that, they have recognized the colour in green, red and blue five times. The result shows in Table 2.

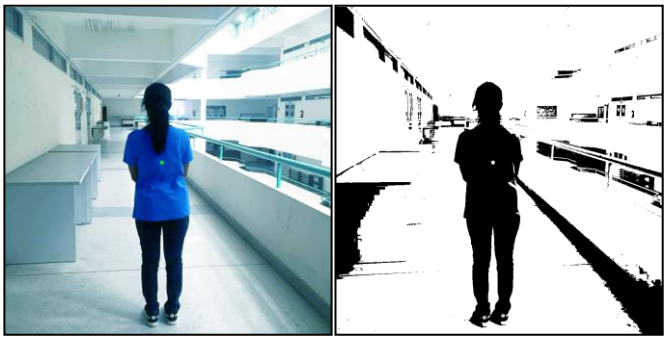


Fig. 8. Detecting the colour with an image processing technique

TABLE III. THE RESULT OF COLOUR DETECTION

Colour	Percentage of Color Detection in Different Distances				
	0.5 meters	**1 meters**	**1.5 meters**	**2 meters**	**2.5 meters**
Green	100%	100%	96.66%	96.66%	86.66%
Red	80%	40%	10%	0%	0%
Blue	96.66%	93.33%	50%	33.33%	13.33%

Table III presents all colour can analysis the colour is extremely precision in the distance 0.5-1 meters. Green got to high accuracy detection, more than 80%. Red cannot detect the colour in the distance more than two meters.

Also, these experiment results, we think that they can be adapted to use in the Hybrid Indoor Navigation for inspection inside the building by using the quadcopter.

VI. CONCLUSION AND FUTURE WORK

In this research, we developed the framework conception for hybrid indoor navigation of the quadcopter for supporting the building security. Our method focuses on selecting the best route and obstacle detection. Firstly, Multi-level Indoor Navigation Ontology is applied to our framework, which can design to support the quadcopter indoor route. All coordinates on the map used to be the information for navigation. After the design, we evaluated some routes which the result is right, can lead to developing to autonomous flight. Secondly, object detection is designed for our research. However, we just validated the colour detection with the camera on the quadcopter. The result can detect the colour object inside the building, and it can extend to being object detection.

In future research, we will work on the auto-flight of quadcopter for improving the efficient model. Our design may create to being air squadron or using many quadcopters on the mission, which can get video or photos information to supporting for deciding on building security. Nevertheless, the object detection should add the other techniques for helping to auto-pilot of the quadcopter as well.

REFERENCES

[1] T. Luukkonen, "Modelling and control of quadcopter," *Independent research project in applied mathematics, Espoo*, 2011.

[2] B. Yu, L. Xu, and Y. Li, "Bluetooth Low Energy (BLE) based mobile electrocardiogram monitoring system," in *Information and Automation (ICIA), 2012 International Conference on*, 2012, pp. 763-767: IEEE.

[3] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y.-D. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 29-35, 2018.

[4] T. Pobkrut, T. Eamsa-Ard, and T. Kerdcharoen, "Sensor drone for aerial odor mapping for agriculture and security services," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2016, pp. 1-5: IEEE.

[5] U. R. Mogili and B. Deepak, "Review on application of drone systems in precision agriculture," *Procedia computer science*, vol. 133, pp. 502-509, 2018.

[6] P. Patel, "Agriculture drones are finally cleared for takeoff [News]," *IEEE Spectrum*, vol. 53, no. 11, pp. 13-14, 2016.

[7] I. Sa and P. Corke, "Vertical infrastructure inspection using a quadcopter and shared autonomy control," in *Field and Service Robotics*, 2014, pp. 219-232: Springer.

[8] T. Krajník, V. Vonásek, D. Fišer, and J. Faigl, "AR-drone as a platform for robotic research and education," in *International Conference on Research and Education in Robotics*, 2011, pp. 172-186: Springer.

[9] I. Sa and P. Corke, "System identification, estimation and control for a cost effective open-source quadcopter," in *Robotics and automation (icra), 2012 ieee international conference on*, 2012, pp. 2202-2209: IEEE.

[10] D. R. M. Liming Luke Chen, P. Dr Matthias Steinbauer, A. Mossel, M. Leichtfried, C. Kaltenriner, and H. Kaufmann, "SmartCopter: Enabling autonomous flight in indoor environments with a smartphone as on-board processing unit," *International Journal of Pervasive Computing and Communications*, vol. 10, no. 1, pp. 92-114, 2014.

[11] Y. S. Vintervold, "Camera-Based Integrated Indoor Positioning," Institutt for teknisk kybernetikk, 2013.

[12] N. Guarino, "Formal ontology and information systems," in *Proceedings of FOIS*, 1998, vol. 98, pp. 81-97.

[13] J. Scholz and S. Schabus, "An indoor navigation ontology for production assets in a production environment," in *International conference on geographic information science*, 2014, pp. 204-220: Springer.

[14] C. Anagnostopoulos, V. Tsetsos, and P. Kikiras, "OntoNav: A semantic indoor navigation system," in *1st Workshop on Semantics in Mobile Environments (SME05), Ayia*, 2005: Citeseer.

[15] P. Kikiras, V. Tsetsos, and S. Hadjiefthymiades, "Ontology-based user modeling for pedestrian navigation systems," in *ECAI 2006 Workshop on Ubiquitous User Modeling (UbiqUM), Riva del Garda, Italy*, 2006.

[16] L. Yang and M. Worboys, "A navigation ontology for outdoor-indoor space:(work-in-progress)," in *Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness*, 2011, pp. 31-34: ACM.

[17] M. Nieuwenhuisen, D. Droeschel, M. Beul, and S. Behnke, "Obstacle detection and navigation planning for autonomous micro aerial vehicles," in *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*, 2014, pp. 1040-1047: IEEE.

[18] L. Díaz Vilariño, P. Boguslawski, K. Khoshelham, H. Lorenzo, and L. Mahdjoubi, "Indoor navigation from point clouds: 3D modelling and obstacle detection," 2016: International Society for Photogrammetry and Remote Sensing.

[19] A. Broggi, S. Cattani, M. Patander, M. Sabbatelli, and P. Zani, "A full-3D voxel-based dynamic obstacle detection for urban scenario using stereo vision," in *Intelligent Transportation Systems-(ITSC),*

2013 16th International IEEE Conference on, 2013, pp. 71-76: IEEE.

[20] S. Jung, S. Hwang, H. Shin, and D. H. Shim, "Perception, guidance, and navigation for indoor autonomous drone racing using deep learning," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2539-2544, 2018.

[21] M. Jia, Y. Sun, and J. Wang, "Obstacle detection in stereo bird's eye view images," in *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, 2014, pp. 254-257.

[22] S. Khruahong, X. Kong, K. Sandrasegaran, and L. Liu, "Multi-Level Indoor Navigation Ontology for High Assurance Location-Based Services," in *The 18th IEEE International Symposium on High Assurance Systems Engineering*, 2017, Singapore.

[23] S. Khruahong, X. Kong, K. Sandrasegaran, and L. Liu, "Develop An Indoor Space Ontology For Finding Lost Properties for Location-Based Service of Smart City," in *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, 2018, pp. 54-59: IEEE.

DDOS Attack Detection & Prevention in SDN using OpenFlow Statistics

Nisha Ahuja[1], Gaurav Singal[2]
Department of CSE, *Bennett University*, Greater Noida, India
na6742@bennett.edu.in[1], gauravsingal789@gmail.com[2]

***Abstract*—Software defined Network is a network defined by software, which is one of the important feature which makes the legacy old networks to be flexible for dynamic configuration and so can cater to today's dynamic application requirement. It is a programmable network but it is prone to different type of attacks due to its centralized architecture. The author proposed a method to detect and prevent DDOS attack in the paper. Mininet [5] which is a popular emulator for Software defined Network is used. We followed the approach in which collection of the traffic statistics from the various switches is done. After collection we calculated the packet rate and bandwidth which shoots up to high values when attack take place. The abrupt increase detects the attack which is then prevented by changing the forwarding logic of the host nodes to drop the packets instead of forwarding. After this, no more packets will be forwarded and then we also delete the forwarding rule in the flow table. Hence, we are finding out the change in packet rate and bandwidth to detect the attack and to prevent the attack we modify the forwarding logic of the switch flow table to drop the packets coming from malicious host instead of forwarding it.**

***Index Terms*—SDN, Mininet, Network attack, Traffic simulation, DDOS**

I. Introduction

Software-Defined Network [1] is the network is the network which can be programmed unlike the traditional network. It is also known as next-generation network. Software-Defined Network is a breakthrough in networking as it breaks the year old barriers of rigidity in network configurations. With the advent of SDN the network can be programmed at run time as per the user requirement. Open Networking Foundation(ONF) [1] defines SDN as the network in which the control plane is segregated from the data plane where the control plane is the brain of the network. With SDN there is a global vision of the network i.e. the status of all the switches existing in the topology is known so the rules for traffic routing can be decided intelligently. SDN is invented as a brainchild of a Ph.D Student Martin Casado during his thesis writing in 2005 at Stanford University. In nutshell we can say that Software-defined network is a network which is defined and managed by software. Also, when combined with the concept of Network function virtualization (NFV) [1] it brings out virtualization of entire networking hardware.

The traditional network cannot be programmed by the software. The devices in the network like switch, router are hard to configure as the two planes in the device i.e. Data Plane and Control Plane are tightly coupled. It is basically due to the distributed nature of traditional network and thus inflexible [1]. But Software-Defined Network is based on the concept of splitting the network into data plane and Control Plane. Control Plane [2] is the one responsible for making all the decisions in the network. We can see the control plane in SDN like the brain in human body and Data Plane is known for its functionality of just forwarding the packets.

This architectural change has brought a lot of flexibility to the network. In nutshell, traditional network routing and forwarding decisions are made by the switches based on the destination IP address which is unlike Software Defined Network. In SDN the routing and forwarding decisions [3] are decoupled and here traffic forwarding is based on flow-based scheme. A flow is defined as a rule based on the match which is then used to forward the traffic.

Various companies have also started looking for the option of investing in SDN. One of the big industry giants like Google is using SDN in their data centers. Firms like NEC to Hewlett-Packard have setup a SDN network. SDN consists of folllowing layers namely a) Forwarding layer consist of dumb switches b) Control layer consist of Controller c) Application layer consisting of application programs which the user can make to interact with the controller. These three layers communicate using Northbound and Southbound APIs [1]. A northbound interface is defined as the set of API's [1]between the application plane and the control plane whereas southbound interface is the set of API's between the controller and the device plane.

Our motivation to work on DDOS attack is because the work which has been done till now is mostly focused on the idea of finding the randomness [5] in the traffic, based on Shannon's theory. But the proposed work here involve calculation at switches which is less time-consuming. DDOS attack can result in the following scenarios:

- Exhausting the memory of switch and controller.
- Exhausting the control channel bandwidth.
- Exhausting the computational power at data and control plane.

Our contribution in this paper is to detect and prevent the DDOS attack by using Open Flow statistics which are extracted from the switches. Continuous monitoring of the

switch is done and whenever the value of the packet rate is found to be more than '100' which is a already set threshold value, the attack is detected. The prevention method is also used in which the controller modify the flow entry from forward packets to drop the packets coming from the identified malicious host.

The paper is organized as outlined below, Section II detailed on the literature survey whereas Section III illustrates the method used for detecion and pevention of attack. and algorithm strategy. Section IV provides the Experimental Setup and Results and finally, Section V concludes the paper.

II. Related Work

Here we discuss the various state-of-the-art in the field of security [4] of the network specially in the context of DDOS attack. Security techniques [6] which have been discussed does not necessarily pertains to SDN, we have diuscussed the core techniques only, it is not necessary that they are SDN specific.

Hong et.al. [7] discuss the concept of mitigation of slow DDOS attack on the server by following a technique employed at the server. The algorithm for detection and mitigation is running on Controller. A Tcp connection has to be already established between SHDA (Secure hash Decryptable Analysis) and web server. Now, whenever attacker tries to connect to the web server it will send large amount of connection request $half - openconnections$ and when the count reaches a threshold, the attack is detected. When detection process running in the server detects the attack, it forwards the half-open connection packets to the mitigation algorithm running on the controller which will continue receiving of the half-open connections. After waiting for the stipulated time in which half-open connection of slow client will complete but the half-open connection of attacker will never complete. After waiting for a threshold limit of time the traffic is classified as attacked traffic or benign traffic. Once the controller add the IP's suspected into the blocked list, the attack is mitigated.

Kalkan et.al. [9] describes the concept of using entropy together with the attributes of TCP layer. For example the randomness in the destination IP address will not solely give accurate information of the attacker but some other parameters like Protocol type, TCP flags, destination port, source port, destination IP, Source IP and packet size are also important to consider for detecting the attack. It creates pairing of traffic features mentioned above and calculates the joint entropy (for example: Destination IP with TTL i.e. time to live) when in non-attacked position and compare it with the same pair after attack. If the difference in entropy values for the normal and attacked traffic exceeds the threshold value then the attack is detected and mitigation module start executing in which detected pairs during the previous phase are taken in ratio with the same during normal traffic flow and a ratio is calculated. If the value of the ratio comes to be greater than the set limit value then it starts dropping packets for that pair unless the bandwidth decreases to acceptable limits.

Da Silva et.al. [10] detailed on the architecture for detecting, classifying and preventing the attack in SDN. The author used the concept of Shannon's theory for the purpose of attack detection in two phases: In the first phase, illegal attack traffic is found by using entropy-based approach and in second phase the dump flows analysis and its classification is done using various machine learning classifiers which uses the past information collected to classify the abnormal traffic.

NisharaniMeti et.al. [11] proposed the use of various Machine-Learning algorithms to detect the attack in Software-Defined Network (SDN). Various Machine learning algorithms are tested for their performance in classification task. The dataset used is a TCP traffic set between certain locations obtained from the experimental results. As the controller is the brain of the network, it is the only one with the help of which the detection and prevention take place.Controller keeps an access contro list (ACL) with the help of which it segregates the TCP traffic set into normal or malicious traffic and then the labelled traffic set is used by the machine learning classifiers and it is found that SVM algorithm gives better results as compared to other machine learning algorihms.

Rasool et.al. [12] discuss the use of deep learning based algorithm for traffic classification into two classes i.e malicious and legitimate. The author has done Link Flooding attack in which the controller is disconnected from the data plane by sending slow legitimate traffic on the control channel. This type of attack was initially handled by the use of traffic filter but it was still a challenge. So the paper analyses the statistics during attack and then use Deep Learning algorithms to classify the traffic as malicious or legitimate. The author used mininet for real time traffic generation. It uses Artificial Neural Network for model training and then use the statistics collected as a test set for predicting the traffic class.

III. Distributed Denial of Service Attack and Countermeasures

In this section first we describe the Distributed Denial of Service(DDOS) attack which leads to data plane and control plane saturation. Later on, we present our proposed countermeasure to detect and mitigate the attack.

A. Distributed Denial of Service Attack

In Distributed Denial of Service Attack (DDOS) shown in figure 1, adversary aim is to saturate the target host with so many packets that it lets suffer the legitimate user in using the network services. Adversary performs the attack in such a way that it leads to both data and control plane saturation [8]. Adversary performs the following attacks:

- Flooding attack: Adversary performs the flooding attack to compromise the bandwidth and switch memory and thus controller delays access to benign users. The flooding attack is done by the use of the ping command from the source. For example if we assume h1 as a target host in our topology and h2, h3, h4 as the malicious hosts trying to attack H1. Syntax for one of the attack can be given as: h2 ping –f h1 will send flood of packets from

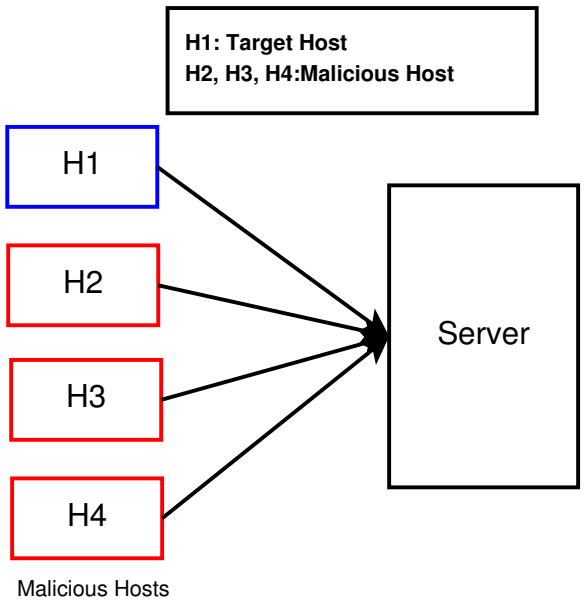


Fig. 1. DDOS Attack

h2 to h1. Similarly, we can perform attack from h3 and h4.

- TCP-SYN Attack: Adversary performs the TCP Syn attack by directing the Syn Packet to the target but the destination host is yet to send the ACK to the source in the mean time adversary send other Syn Packet for connection establishment, so it become difficult for the target to handle the traffic.This is because of the reason as controller already allocates all its resources for the previous Syn requests and so controller reaches its saturation to handle the packets. To detect such an attack we have proposed a threshold-based approach which is explained in the next section.

B. Countermeasures

We proposed a countermeasure known as Continuous watch. It detects the attack by the malicious node by continuously analyzing the traffic statistics on switch. Detection and prevention of the attack [14] is done in two phases. In the first phase, attack is detected by analyzing the rate of change in the packet rate and bandwidth. In the second phase, the attack is prevented by changing the forwarding logic of the target host in the switch flow table.

1) Primary phase of detection: In this phase a thread is run continuously to monitor the switch. At regular intervals of time it collects the flow statistics which include various parameters which are mentioned in I & II

Packet count is used in the paper to calculate the packet rate. The packet rate value is used to check for the attack, whenever the packet rate is greater than a predefined threshold value of 100, the attack is detected and appropriate action is taken. Algorithm 1 shows the pseudo-code for attack detection.

TABLE I
FLOW STATISTICS

S.No	*Flow statistics*	*Meaning*
1	byte_count	Count of bytes
2	duration_n_secs	flow duration in seconds
3	priority	Priority of the flow
4	hard time out	set time when flow is removed
5	idle timeout	set idle time when flow is removed
6	len	lenghth of the message
7	match	can be IP,Mac address,Port No
8	packetcount	it is the total count of packets

TABLE II
PORT STATISTICS

S.No	Flow statistics	Meaning
1	txbytes	Bytes sent on a port in network
2	rxbytes	Bytes recieved on a port in network
3	txpackets	Packets sent on a port in network
4	rxpackets	Packets received on a port in network
5	txerrors	Errors during sending the packets
6	rxerrors	Errors during receiving of the packets
7	port id	Port number
8	datapath	switch id
9	duration	time during transmission
10	in_port	entry port
11	out_port	exit port

Algorithm 1 Algorithm for detection of attack in SDN

Input: Traffic statistics at switches
Output: Attack detected successfully

Initialisation :packetratelimit=100
For each flow collect the packet count and tx_ & rx_ byte statistics
for $i = packet1$ to $packet\_n$ **do**
  packetrate=packet_count/duration
  Bandwidth=((tx_bytes+rx_bytes)*8)/1024
  if (packetrate(i) $\geq packetratelimit$) **then**
    Attack detected
  end if
end for
return Flow with $SourceIP$,$destinationIP$

Secondly, we also compute the bandwidth for every 30 seconds. Bandwidth is computed for all the ports by specifying the OFPP_ANY [15] to the OFPFlowStatsRequest [5], [15] to measure the bandwidth of all the ports. We can specify the specific port number if we want to measure it from a particular port. Bandwidth is also found to be increasing high when the attack take place, but once the prevention technique is used then bandwidth will reduced. Flowchart in figure 2 shows the process of attack detection ad prevention.

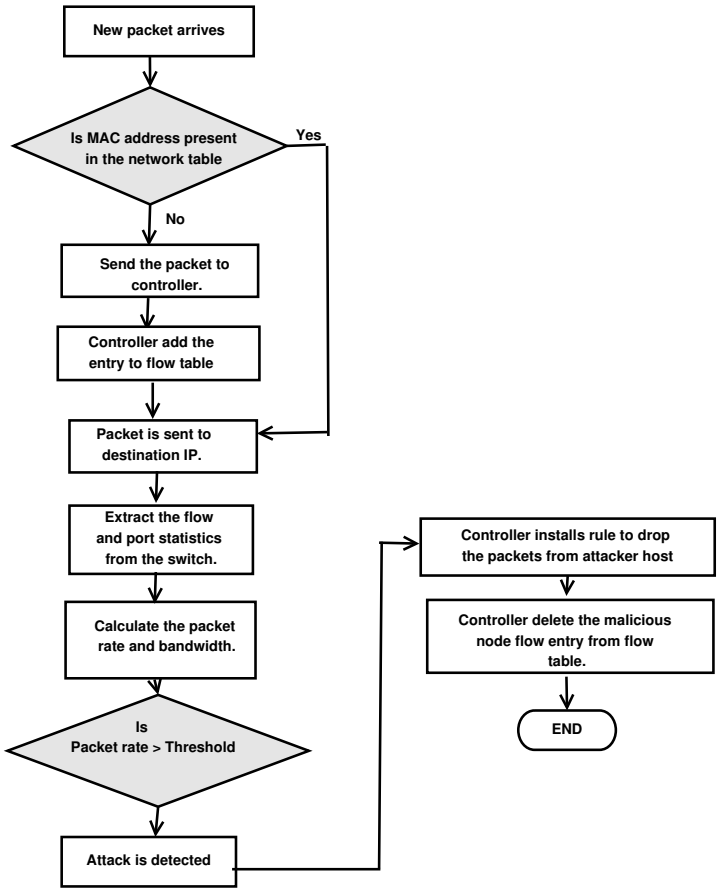


Fig. 2. Attack detection and prevention process

Algorithm 2 Algorithm for prevention of attack in SDN

Input: Flow for the Packet Rate,Bandwidth
Output: Deleted BlockedList

1: For the flow which has the identified packet rate and bandwidth
2: **for** $i = flow1$ to $flow\_n$ **do**
3: append the corresponding source IP of that flow into blocked list which is a list of IPs who have exceeded the treshold packet rate.
4: **if** (sourceIP(i) == $BlockedList$(i) **then**
5: Modify the action of that flow to drop the packets from source IP identified.
6: **end if**
7: **end for**
8: **return** Deleted $BlockedList$

2) Secondary phase & Countermeasures: In this phase, Attack prevention take place. The attack which is detected during the primary phase is prevented by changing the forwarding logic of the switches by modifying the flow rule in the flow table for discarding the further packets by using OFPFlowMod message [5] instead of forwarding. Algorithm 2 shows, the pseudo-code for the prevention of the attack.

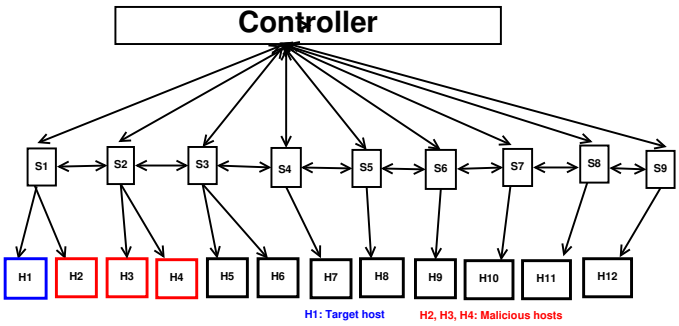


Fig. 3. Simulated topology

IV. Experimental Setup & Results

In this section, we show the setup conditions during experiments, hardware, and software used, their configurations and how we calculate the parameters associated with the experiment. Finally, the impact of attack with time is also shown.

A. Experimental setup

We used Mininet [5] as an emulator to analyze the network traffic, the attacks and their countermeasures along with Ryu as a network controller. The controller is written entirely in python and we have deployed our application logic as an application on Ryu.

Various simulation parameters are shown in Table III. Topology considered under investigation consists of 9 switches and 12 hosts is shown in figure 3. In the topology, H1 is the target host and Host 2, Host 3 and 4 are the malicious hosts. Host 1 is connected on port 1 of switch S1, Host 2 is connected on port 2 of switch S1. Host 3 and 4 are connected on port 1 & 2 of switch s2 and switch S2 is connected on port 3 of switch S1.

TABLE III
Simulation Environment

S.No	Parameters	Value
1	Host OS	Windows10
2	Guest OS	Ubuntu16.04
3	VirtualBox	5.1.26
4	Emulator	Mininet
5	Controller	Ryu
6	# Controller	1
7	# Switches	9
8	# Hosts	12
9	Protocol Used	OpenFlow
10	Graphical package	MiniEdit
11	Traffic Generation tool	Iperf, Curl, Ping
12	Controller Port Number	6653
13	Bandwidth	100 kbps
14	Simulation Time	300 seconds
15	Packet rate threshold used	100 packet per second
16	Statistics collection interval	30 seconds
17	Bandwidth plot interval	30 seconds

B. Performance Metrics

We have used two parameters for performance measurement to evaluating the effects of attacks and countermeasures.

- Packet Rate: It is defined as the number of packets in the network per second. From the statistics collected total packet count can be collected from a switch but packet rate has to be calculated by subtracting the previous packet count from the current packet count and dividing by the total duration which will give the number of packets transmitted per second. Packet Rate is calculated as under :

$$packetrate = packetcount/duration \quad (1)$$

- Bandwidth: It is the total number of bytes transferred and received on a port. We are collecting the port statistics as mentioned in table II by calling OFPPortStatsRequest [15] which return the Port related statistics. Out of which we are using txbyte & rxbytes(transmitted bytes & received bytes) to calculate the bandwidth. By calculating the bandwidth we found that bandwidth suddenly shoots up in case of attack. Thus confirming that attack is done. We are calculating the bandwidth after every 30 seconds. Bandwidth is calculated as under:

$$bandwidth = (txbytes + rxbytes) \quad (2)$$

C. Result Analysis

To analyze the results we have plotted the graph of the two performance metrics without prevention technique and with prevention technique which will help us understand the scenario of attack in better way.

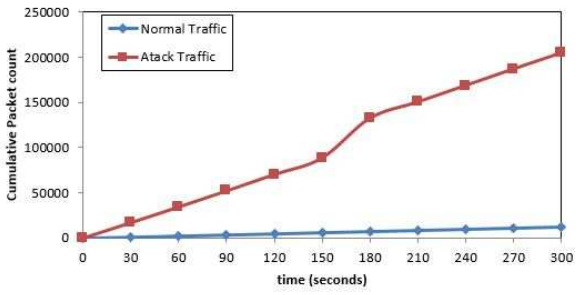


Fig. 4. Packet count vs time

Figure 4 shows the effect of attack on packet count. Figure 4 shows the effect of attack on number of packets. As can be seen from the figure that when there is no attack the number of packets over the network are in the range of (1238-12380) packets when measured after every 30 seconds. The packet count range increases in such a way because packet sending is done at uniform rate of 100 kbps by all the three hosts in the packet range from 1238 packets measured after every 30 seconds which equals to 12380 packets as the experiment is run for 300 seconds. Similarly, when a malicious node attack the target host (H1) the packet count shows a abrupt increase and reaches in the range of (18000-200000) packets. It is a strong indication of the attack. The increase in packet count is due to the malicious nodes(H2, H3, H4) attacked the target host (H1) at 100 packets per second by three malicious nodes respectively. Host 2 will send on port 2, so there will be 3000 packets every 30 seconds and on port 3 there will 6000 packets every 30 second. So three malicious nodes attack the target host and packet count increases in the interval of 9000 packets every 30 seconds and a total of 18000 as twice count for request and response.

For preventing the attack we have calculated the packet rate which is the per second count of the packets in the network. To check the attack, packet rate threshold value of '100' is used and if the packet rate at a particular switch crosses the threshold rate then the attack is detected and the control goes to the prevention module.

The prevention module alters the forwarding logic of the host from further forwarding the packets by changing the action part of the flow table to drop packets instead of forward and in this way the malicious source IP will not be able to send the packets again.

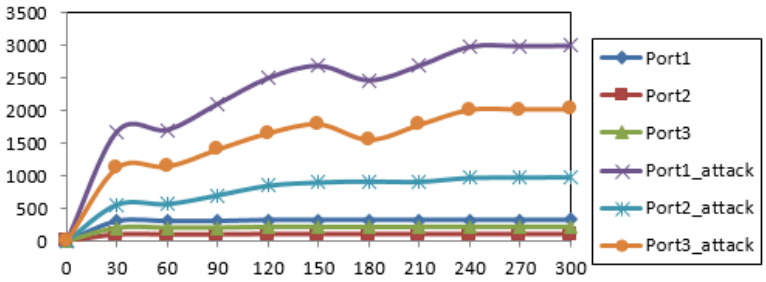


Fig. 5. Bandwidth during normal vs attack traffic.

Figure 5 depicts the bandwidth graph plotted between normal and attacked scenario. From the graph, we can infer that bandwidth of the network during normal traffic ranges in between (100-300 kbps) which shoots up during the attack in the range of (1600-3000 kbps) which is a indication of attack. Port1 in yellow is the port which is connected to host1 which is the target host and it is the sum of bandwidth at host 2, host 3 and host 4 who have attacked the target host at the rate of 100 packet per second.

Figure 6 depicts the bandwidth of the network when countermeasure is applied which stays in the range of (400-600 kbps). It is because of the packet rate threshold which has been set to 100 packets per second, due to which when the attacker (H2, H3, H4) continue sending packets, they are dropped and port1 drop the packets and its bandwidth falls to zero.There is packet sending still done by Host 2,3,4 but Host 1 drop the packets can be seen from the figure 5 that port2, 3,4 have not fall to zero.

Figure 7 depicts the graph between packet rate before and after the countermeasure is applied. When the

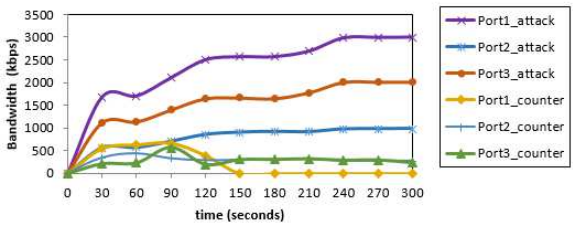


Fig. 6. Traffic Bandwidth in attacked vs after countermeasure applied

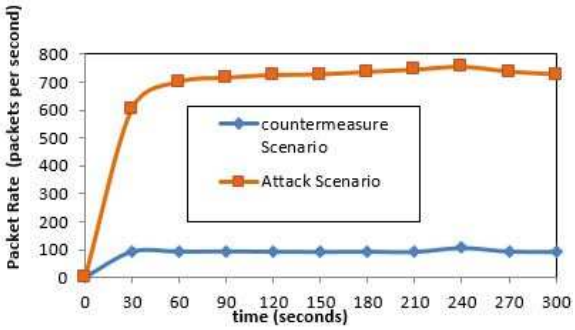


Fig. 7. Packet Rate during attack vs after prevention applied.

malicious hosts(H2, H3, H4) attack the target host (H1). The attack rate shoots up in the range of (600-744) but after applying the countermeasure whenever the packet rate goes above 100 the packets are dropped which is shown in the figure 7 that when packet rate becomes 106 at 240 seconds the switch started dropping packets and so packet rate comes down to 92-93.

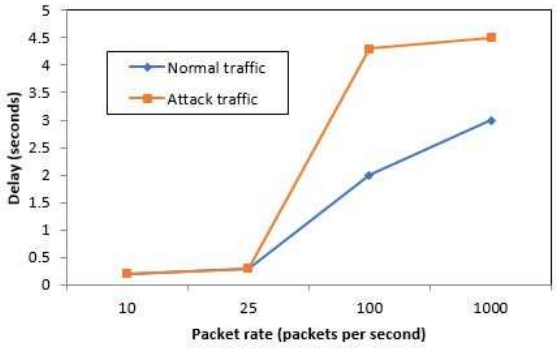


Fig. 8. Delay caused to legitimate user in normal traffic vs attack traffic.

Figure 8 shows the graph between delay caused to the legitimate user in sending the data when the attack take place. Graph depicts that when the attack rate increases from 100 pkt/sec, further increase is not increasing the delay. The delay decreases at higher attack rates because the controller is already busy in processing of previous packets. So delay reaches saturation.

V. Conclusion

In Software-defined networks, security [13] is a major concern and also data breaches is increasing with time. In this paper, we have implemented a method for detecting and preventing DDOS attack in SDN by collecting the flow statistics and port statistics at regular interval from the switches. We have calculated the packet rate and bandwidth for performance evaluation. If the packet rate is above the defined threshold then the source is not allowed to send further packets by changing the forwarding logic in the flow table that maintained by switch.

In future, we will be implementing machine learning and deep learning approach for detection and prevention of such attacks using traffic logs. In SDN, multiple new attacks are also coming these days, so we are planing to implement a single stop reliable approach for all the attacks.

References

[1] M. Dabbagh, B. Hamdaoui, M. Guizani, A. Rayes, Software-defined networking security: pros and cons, IEEE Communications Magazine 53 (6) (2015) 73–79. doi:10.1109/MCOM.2015.7120048.

[2] D. Kreutz, F. M. Ramos, P. Verissimo, Towards secure and dependable software-defined networks, in: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, HotSDN '13, ACM, New York, NY, USA, 2013, pp. 55–60. doi:10.1145/2491185.2491199.

[3] S. Shin, G. Gu, Attacking software-defined networks: A first feasibility study, in: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, HotSDN '13, ACM, New York, NY, USA, 2013, pp. 165–166. doi:10.1145/2491185.2491220.

[4] M. Brooks, B. Yang, A man-in-the-middle attack against OpenDayLight SDN controller, in: Proceedings of the 4th Annual ACM Conference on Research in Information Technology, RIIT '15, ACM, New York, NY, USA, 2015, pp. 45–49. doi:10.1145/2808062.2808073.

[5] www.mininet.org.

[6] Poisoning Network Visibility in Software-Defined Networks: New Attacks and Countermeasures. doi:10.14722/ndss.2015.23283.

[7] Hong, Kiwon, Youngjun Kim, Hyungoo Choi, and Jinwoo Park. "SDN-assisted slow HTTP DDoS attack defense method." IEEE Communications Letters 22, no. 4 (2017) 688-691.

[8] C. Yoon, S. Lee, H. Kang, T. Park, S. Shin, V. Yegneswaran, P. Porras,G. Gu, Flow wars: Systemizing the attack surface and defenses in software-defined networks, IEEE/ACM Transactions on Networking 25 (6) (2017) 3514–3530. doi:10.1109/TNET.2017.2748159.

[9] Kalkan, Kübra, Levent Altay, Gürkan Gür, and Fatih Alagöz. "JESS: Joint Entropy-Based DDoS Defense Scheme in SDN." IEEE Journal on Selected Areas in Communications 36, no. 10 (2018): 2358-2372.

[10] da Silva, Anderson Santos, Juliano Araujo Wickboldt, Lisandro Zambenedetti Granville, and Alberto Schaeffer-Filho. "ATLANTIC: A framework for anomaly traffic detection, classification, and mitigation in SDN." In NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium, pp. 27-35. IEEE, 2016.

[11] Meti, Nisharani, D. G. Narayan, and V. P. Baligar. "Detection of distributed denial of service attacks using machine learning algorithms in software defined networks." In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1366-1371. IEEE, 2017.

[12] Rasool, Raihan Ur, Usman Ashraf, Khandakar Ahmed, Hua Wang, Wajid Rafique, and Zahid Anwar. "Cyberpulse: A Machine Learning Based Link Flooding Attack Mitigation System for Software Defined Networks." IEEE Access 7 (2019): 34885-34899.

[13] K. K. Karmakar, V. Varadharajan, U. Tupakula, M. Hitchens, Policy based security architecture for software defined networks, in: Proceedings of the 31st Annual ACM Symposium on Applied Computing, SAC '16, ACM, New York, NY, USA, 2016, pp. 658–663. doi:10.1145/2851613.2851728.

[14] A. Shukhman, P. Polezhaev, Y. Ushakov, L. Legashev, V. Tarasov, N. Bakhareva, Development of network security tools for enterprise software-defined networks, in: Proceedings of the 8th International Conference on Security of Information and Networks, SIN '15, ACM, New York, NY, USA, 2015, pp. 224–228. doi:10.1145/2799979.2800009.

[15] E. Haleplidis, K. Pentikousis, S. Denazis, J. H. Salim, D. Meyer, O. Koufopavlou, Software-defined networking (SDN): Layers and architecture terminology, RFC 7426, RFC Editor, http://www.rfc-editor.org/rfc/rfc7426.txt (January 2015).

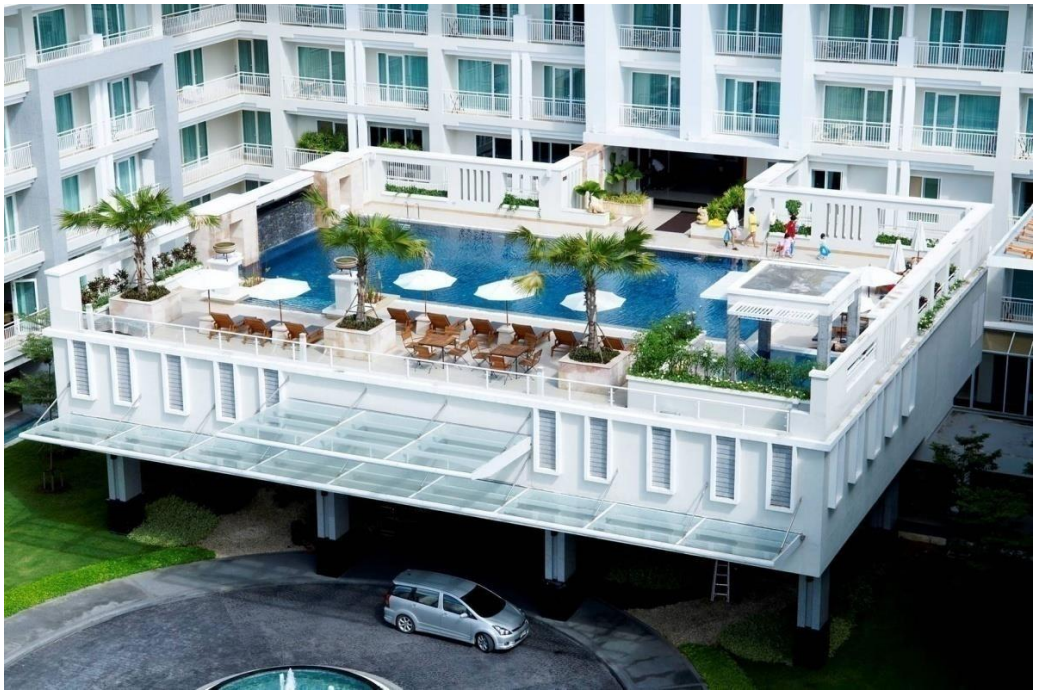
Conference Venue

Kantary Hills HotelWebsite

44, 44/1-4 Nimmanhaemin Road, Soi 12, Suthep, Muang, Chiang Mai 50200, Thailand.
Tel: 66 (0)53 222 111, 66 (0)53 400 877 Fax: 66 (0)53 223 244

Accommodation

Tastefully and intelligently appointed studios, one-bedroom and two-bedroom apartments have been designed to be the last word in contemporary Asian style. All our rooms and suites blend with the natural beauty of the region, providing you with everything you need for short and long-term visits.



iSAI-NLP 2019 Organizers

iSAI-NLP 2019was organized by

AiAT
ARTIFICIAL INTELLIGENCE ASSOCIATION OF THAILAND

MUBAN CHOM BUENG RAJABHAT UNIVERSITY

SIIT

NECTEC
a member of NSTDA

IEEE
THAILAND SECTION

THAMMASAT UNIVERSITY

Damaynti Attanam Panditā
CHIANG MAI UNIVERSITY SINCE 1964


PHETCHABURI RAJABHAT UNIVERSITY


กฟผ
การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย

EGAT


KMUTNB
Management Information Systems