

Lecture Notes in Electrical Engineering 1028

Ji Su Park

Laurence T. Yang

Yi Pan

Jong Hyuk Park *Editors*

Advances in Computer Science and Ubiquitous Computing

Proceedings of CUTE-CSA 2022

 Springer

Lecture Notes in Electrical Engineering

Volume 1028

Series Editors

Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Napoli, Italy

Marco Arteaga, Departamento de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico

Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, München, Germany

Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China

Shanben Chen, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore

Rüdiger Dillmann, Gebäude 07.21, University of Karlsruhe (TH) IAIM, Karlsruhe, Baden-Württemberg, Germany

Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China

Gianluigi Ferrari, Dipartimento di Ingegneria dell'Informazione, Sede Scientifica Università degli Studi di Parma, Parma, Italy

Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain

Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA

Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Janusz Kacprzyk, Intelligent Systems Laboratory, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Alaa Khamis, Department of Mechatronics Engineering, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt

Torsten Kroeger, Intrinsic Innovation, Mountain View, CA, USA

Yong Li, College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, China

Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA

Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore

Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany

Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA

Subhas Mukhopadhyay, School of Engineering, Macquarie University, NSW, Australia

Cun-Zheng Ning, Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA

Toyoaki Nishida, Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan

Luca Oneto, Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genova, Genova, Genova, Italy

Bijaya Ketan Panigrahi, Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India

Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi Roma Tre, Roma, Italy

Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

Gan Woon Seng, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Joachim Speidel, Institute of Telecommunications, University of Stuttgart, Stuttgart, Germany

Germano Veiga, FEUP Campus, INESC Porto, Porto, Portugal

Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Haidian District Beijing, China

Walter Zamboni, Department of Computer Engineering, Electrical Engineering and Applied Mathematics, DIEM—Università degli studi di Salerno, Fisciano, Salerno, Italy

Junjie James Zhang, Charlotte, NC, USA

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

Ji Su Park · Laurence T. Yang · Yi Pan ·
Jong Hyuk Park
Editors

Advances in Computer Science and Ubiquitous Computing

Proceedings of CUTE-CSA 2022

 Springer

Editors

Ji Su Park
Department of Computer Science
and Engineering
Jeonju University
Jeonju-si, Korea (Republic of)

Yi Pan
Department of Computer Science
Georgia State University
Atlanta, USA

Shenzhen Institute of Advanced
Technology
Shenzhen, China

Laurence T. Yang
Department of Computer Science
St. Francis Xavier University
Antigonish, NS, Canada

Jong Hyuk Park
Department of Computer Science
and Engineering
Seoul National University of Science
and Technology
Seoul, Korea (Republic of)

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-981-99-1251-3

ISBN 978-981-99-1252-0 (eBook)

<https://doi.org/10.1007/978-981-99-1252-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Message from the CSA 2022 General Chairs

International Conference on Computer Science and its Applications (CSA 2022) is the 14th event of the series of international scientific conference. This conference takes place in Vientiane, Laos, on December 19–21, 2022. CSA 2022 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications. CSA 2022 will provide an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of CSA. In addition, the conference will publish high-quality papers which are closely related to the various theories and practical applications in CSA. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA 2022 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA 2021(13th Edition: Jeju, Korea, 2021), CSA 2020 (12th Edition: Jeju, Korea, 2020), CSA 2019 (11th Edition: Macao, China, 2020), CSA 2018 (10th Edition: Kuala Lumpur, Malaysia), CSA 2017 (9th Edition: Taichung, Taiwan), CSA 2016 (8th Edition: Bangkok, Thailand, 2016), CSA 2015 (7th Edition: Cebu, December, 2015), CSA 2014 (6th Edition: Guam, December, 2014), CSA 2013 (5th Edition: Danang, December, 2013), CSA 2012 (4th Edition: Jeju, November, 2012), CSA 2011 (3rd Edition: Jeju, December, 2011), CSA 2009 (2nd Edition: Jeju, December, 2009), and CSA 2008 (1st Edition: Australia, October, 2008).

The papers included in the proceedings cover the following topics: mobile and ubiquitous computing, dependable, reliable and autonomic computing, security and trust management, multimedia systems and services, networking and communications, database and data mining, game and software engineering, grid and scalable computing, embedded system and software, artificial intelligence, distributed and parallel algorithms, web and Internet computing, and IT policy and business management.

Accepted and presented papers highlight new trends and challenges of computer science and its applications. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research. We would like to express our sincere thanks to

Steering Chairs: Jong Hyuk Park (SeoulTech, Korea), Young-Sik Jeong (Dongguk University, Korea), Namme Moon (Hoseo University, Korea), Yi Pan (Georgia State University, USA), Vincenzo Loia (University of Salerno, Italy), and Han-Chieh Chao (National Ilan University, Taiwan).

Our special thanks go to the program chairs: Ji Su Park (Jeonju University, Korea), Yan Li (Inha University, Korea), S. Vimal (National Engineering College, India), Alireza Souri (Islamic Azad University, Iran), Ping Liu (Chongqing University of Posts and Telecommunications, China) and all program committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

CSA 2022 General Chair

Jungho Kang
Baewha Women's University
Seoul, Korea

Kim-Kwang Raymond Choo
The University of Texas at San Antonio
San Antonio, USA

Piao Changhao
Chongqing University of Posts and Telecommunications
Chongqing, China

Message from the CSA 2022 Program Chairs

Welcome to the 14th International Conference on Computer Science and its Applications (CSA 2022) which will be held in Vientiane, Laos, on December 19–21, 2022. CSA 2022 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications.

CSA 2022 provides an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of computer science. In addition, the conference contains high-quality papers which are closely related to the various theories and practical applications in computer science. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA 2022 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA 2021(13th Edition: Jeju, Korea, 2021), CSA 2020 (12th Edition: Jeju, Korea, 2020), CSA 2019 (11th Edition: Macao, China, 2020), CSA 2018 (10th Edition: Kuala Lumpur, Malaysia), CSA 2017 (9th Edition: Taichung, Taiwan), CSA 2016 (8th Edition: Bangkok, Thailand, 2016), CSA 2015 (7th Edition: Cebu, December, 2015), CSA 2014 (6th Edition: Guam, December, 2014), CSA 2013 (5th Edition: Danang, December, 2013), CSA 2012 (4th Edition: Jeju, November, 2012), CSA 2011 (3rd Edition: Jeju, December, 2011), CSA 2009 (2nd Edition: Jeju, December, 2009), and CSA 2008 (1st Edition: Australia, October, 2008).

CSA 2022 contains high-quality research papers submitted by researchers from all over the world. Each submitted paper was peer-reviewed by reviewers who are experts in the subject area of the paper. Based on the review results, the program committee accepted papers.

For organizing an international conference, the support and help of many people are needed. First, we would like to thank all the authors for submitting their papers. We also appreciate the support from the program committee members and reviewers who carried out the most difficult work of carefully evaluating the submitted papers.

We would like to give our special thanks to Prof. Jong Hyuk Park, Prof. Young-Sik Jeong, Prof. Nammee Moon, Prof. Yi Pan, Prof. Vincenzo Loia, and Prof. Han-Chieh Chao the steering committee chairs of CSA for their strong encouragement

and guidance to organize the symposium. We would like to thank the CSA 2022 general chairs: Prof. Jungho Kang, Prof. Kim-Kwang Raymond Choo, and Prof. Piao Changhao. We would like to express special thanks to the committee members for their timely unlimited support.

CSA 2022 Program Chairs

Ji Su Park
Jeonju University
Jeonju-si, Korea

Yan Li
Inha University
Incheon, Korea

S. Vimal
National Engineering College
Kovilpatti, India

Alireza Souri
Islamic Azad University
Tehran, Iran

Ping Liu
Chongqing University of Posts and Telecommunications
Chongqing, China

CSA 2022 Organization

Honorary Chair

Doo-Soon Park, Soonchunhyang University, Korea

Steering Committee

James J. Park, SeoulTech, Korea
Young-Sik Jeong, Dongguk University, Korea
Nammee Moon, Hoseo University, Korea
Yi Pan, Georgia State University, USA
Vincenzo Loia, University of Salerno, Italy
Han-Chieh Chao, National Ilan University, Taiwan
Somphone Kanthavong, National University of Laos, Laos

General Chairs

Jungho Kang, Baewha Women's University, Korea
Kim-Kwang Raymond Choo, The University of Texas at San Antonio, USA
Piao Changhao, Chongqing University of Post and Telecom, China

Program Chairs

Jisu Park, Jeonju University, Korea
Vimontha Khievongphachanh, National University of Laos, Laos

Yan Li, Inha University, Korea
S. Vimal, National Engineering College, India
Alireza Souri, Islamic Azad University, Iran
Ping Liu, Chongqing University of Posts and Telecommunications, China

International Advisory Committee

Mo-Yuen Chow, North Carolina State University, USA
Ka Lok Man, Xi'an Jiaotong-Liverpool University, China
Shu-Ching Chen, Florida International University, USA
Mohammad S. Obaidat, Monmouth University, USA
Enrique Herrera-Viedma, University of Granada, Spain
Hang-Bae Chang, Chung-Ang University, Korea
Sherali Zeadally, University of Kentucky, USA
Jordi Mongay Batalla, National Institute of Telecommunications, Poland
Wanlei Zhou, Deakin University, Australia
Sethuraman Panchanathan, Arizona State University, USA
Yueh-Min Huang, National Cheng Kung University, Taiwan
Byoungsoo Koh, Korea Creative Content Agency (KOCCA), Korea
Junren Shi, Chongqing University of Posts and Telecommunications, China

Workshop Chairs

Michael Hwa Young Jeong, Kyung Hee University, Korea
Neil Y. Yen, The University of Aizu, Japan
Hyuk-Jun Kwon, Soonchunhyang University, Korea

Publicity Chairs

Arun Kumar Sangaiah, VIT University, India
Shailendra Rathore, Korea
Kwang-il Hwang, Incheon National University, Korea
Shailendra Rathore, University of Plymouth, UK
Fei Hao, Shaanxi Normal University, China
Min Choi, Chungbuk National University, Korea
Hyuk Joon Kwon, Soonchunhyang University, Korea
Jinho Park, Soongsil University, Korea
Pradip Kumar Sharma, University of Aberdeen, UK
Mingjie Liu, Changan Co., Ltd., China

Chunyun Fu, Chongqing University, China
Yongsheng Wang, Tsinghua University, China
Dexu Bu, Tsinghua University, China
Xiang Jiang, Chongqing Jiaotong University, China
Li Lu, Chongqing University, China
Chao Jiang, Chongqing Technology and Business University, China
Qian Zhang, Wuhan University of Technology, China
Bing Zhang, Xinxiang University, China

Industrial Cooperation Chairs

Yong Woo Lee, Ssangyong Information and Communications Corp., Korea
Sung Chul Yu, LG Hitachi Ltd., Korea
Sung Gil Kim, VAIV Company Inc., Korea
Bong Sang Seo, ALL4LAND Co., Ltd., Korea
Se Jong Kim, SJ Info and Communications Co., Ltd., Korea
Tae Yoon Kwon, Neighbor System Co., Ltd., Korea
Han Su Cheon, Selim TSG Co., Ltd., Korea
Seokhong Min, Mindata Corporation, Korea
Kailin Wan, Changan Co., Ltd., China

Local Arrangement Chairs

Duan Yaixin, Chongqing University of Post and Telecom, China
Byoungwook Kim, Dongguk University, Korea

Program Committee

Andrew Kusiak, The University of Iowa, USA
Chang Wu Yu, Chung Hua University, Taiwan
Chin-Fu Kuo, The National Kaohsiung University, Taiwan
Debajyoti Mukhopadhyay, Balaji Institute of Telecom and Management, India
Dion Hoe-Lian Goh, Nanyang Technological University, Singapore
Jehn-Ruey Jiang, National Central University, Taiwan
Paprzycki Marcin, Polish Academy of Sciences, Poland
Qian Yu, University of Regina, Canada
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Marco Listanti, DIET, Roma, Italy
Sun-Yuan Hsieh, National Cheng Kung University, Taiwan

Yue-Shan Chang, National Taipei University, Taipei
Ahmed EL Oualkadi, Abdelmalek Essaadi University, Morocco
Jung Hanmin, KISTI, Korea
Liu Chuan-Ming, National Taipei University of Technology, Taipei
Ali Shahrabi, Glasgow Caledonian University, UK
Bo-Chao Cheng, National Chung-Cheng University, Taiwan
Chang Yao-Chung, National Taitung University, Taiwan
Deok Gyu Lee, Seowon University, Korea
Min Choi, Chungbuk National University, Korea
Hang-Bae Chang, Chung-Ang University, Korea
Hong-Jun Jang, Jeonju University, Korea
Imad Saleh, University of Paris 8, France
Jinhyun Ahn, Jeju National University, Korea
Lai Kuan-Chu, National Taichung University, Taiwan
Lam-for Kwok, City University of Hong Kong, Hong Kong
Seung-Ho Lim, Hankook University of Foreign Studies, Korea
Ki Yong Lee, Sookmyung Women's University, Korea
Koojoo Kwon, Baewha Women's University, Korea
Yeong-Seok Seo, Yeungnam University, Korea
Yeong Wook Yang, Hanshin University, Korea
Yoo-Jae Won, Chungnam National University, Korea
Shi Junren, Chongqing University of Post and Telecom, China
Liu Ping, Chongqing University of Post and Telecom, China
Chen Junsheng, Chongqing University of Post and Telecom, China
Chia-Hung Yeh, National Sun Yat-sen University, Taiwan
Jin Gon Shon, Korea National Open University, Korea
Ka Lok Man, Xi'an Jiaotong-Liverpool University, China
Peng Ran, Chongqing University of Post and Telecom, China
Xu Zhang, Chongqing University of Post and Telecom, China

Message from the CUTE 2022 General Chairs

On behalf of the organizing committees, it is our pleasure to welcome you to the 16th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2022), which will be held in Vientiane, Laos, on December 19–21, 2022.

This conference provides an international forum for the presentation and showcase of recent advances on various aspects of ubiquitous computing. It will reflect the state-of-the-art of the computational methods, involving theory, algorithm, numerical simulation, error and uncertainty analysis, and/or novel application of new processing techniques in engineering, science, and other disciplines related to ubiquitous computing.

The papers included in the proceedings cover the following topics: ubiquitous communication and networking, ubiquitous software technology, ubiquitous systems and applications, and ubiquitous security, privacy, and trust. Accepted papers highlight new trends and challenges in the field of ubiquitous computing technologies. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to the steering committees: James J. Park (SeoulTech, Korea), Doo-Soon Park (Soonchunhyang University, Korea), Young-Sik Jeong (Dongguk University, Korea), Namme Moon (Hoseo University, Korea), Hsiao-Hsi Wang (Providence University, Taiwan), Laurence T. Yang (St. Francis Xavier University, Canada), Hai Jin (Huazhong University of Science and Technology, China), Chan-Hyun Youn (KAIST, Korea), Jianhua Ma (Hosei University, Japan), Mingyi Guo (Shanghai Jiao Tong University, China), and Weijia Jia (City University of Hong Kong, Hong Kong). We would also like to express our cordial thanks to the program chairs and program committee members for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

Finally, we would thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

CUTE 2022 General Chairs

Jungho Kang
Baewha Women's University
Seoul, Korea

Laurence T. Yang
St. Francis Xavier University
Antigonish, Canada

Message from the CUTE 2022 Program Chairs

Welcome to the 16th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2022), which will be held in Vientiane, Laos, on December 19–21, 2022.

The purpose of the CUTE 2022 conference is to promote discussion and interaction among academics, researchers, and professionals in the field of ubiquitous computing technologies. This year, the value, breadth, and depth of the CUTE 2022 conference continue to strengthen and grow in importance for both the academic and industrial communities. This strength is evidenced this year by having the highest number of submissions made to the conference.

For CUTE 2022, we received a lot of paper submissions from various countries. Out of these, after a rigorous peer-review process, we accepted only high-quality papers for CUTE 2022 proceeding, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation.

Finally, we would like to thank all of you for your participation in our conference and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

CUTE 2022 Program Chairs

Ji Su Park

Jeonju University

Jeonju-si, Korea

Vimontha Khieovongphachanh

National University of Laos

Vientiane, Laos

Tha Bounthanh

National University of Laos

Vientiane, Laos

Jin Wang

Yangzhou University

Yangzhou, China

Joon-Min Gil

Catholic University of Daegu

Gyeongsan, Korea

Neil Y. Yen

The University of Aizu

Aizuwakamatsu, Japan

CUTE 2022 Organization

Honorary Chairs

Jinmo Kang, President of KIPS, Korea
Somphone Kanthavong, National University of Laos, Laos

Steering Committee

James J. Park, SeoulTech, Korea
Doo-Soon Park, Soonchunhyang University, Korea
Young-Sik Jeong, Dongguk University, Korea
Nammee Moon, Hoseo University, Korea
Hsiao-Hsi Wang, Providence University, Taiwan
Laurence T. Yang, St. Francis Xavier University, Canada
Hai Jin, Huazhong University of Science and Technology, China
Chan-Hyun Youn, KAIST, Korea
Jianhua Ma, Hosei University, Japan
Mingyi Guo, Shanghai Jiao Tong University, China
Weijia Jia, City University of Hong Kong, Hong Kong

General Chairs

Jungho Kang, Baewha Women's University, Korea (Leading Chair)
 Laurence T. Yang, St. Francis Xavier University, Canada

Program Chairs

Ji Su Park, Jeonju University, Korea (Leading Chair)
 Vimontha Khievongphachanh, National University of Laos, Laos
 Tha Bounthanh, National University of Laos, Laos
 Jin Wang, Yangzhou University, China
 Joon-Min Gil, Catholic University of Daegu, Korea
 Neil Y. Yen, The University of Aizu, Japan

International Advisory Committee

Phonpadith Phoummavong, National University of Laos, Laos
 Yi Pan, Georgia State University, USA
 Byeong-Seok Shin, Inha, Korea
 C. S. Raghavendra, University of Southern California, USA
 HeonChang Yu, Korea University, Korea
 Im-Yeong Lee, Soonchunhyang University, Korea
 Han-Chieh Chao, National Ilan University, Taiwan
 Shu-Ching Chen, Florida International University, USA
 Keun Ho Ryu, Chungbuk National University, Korea
 Victor Leung, University of British Columbia, Canada
 Kwang-il Hwang, Incheon National University, Korea
 Sherali Zeadally, University of Kentucky, USA
 Jin Gon Shon, Korea University Open University, Korea
 Sherali Zeadally, University of Kentucky, USA
 Seok Cheon Park, Gachon University, Korea
 Yang Xiao, University of Alabama, USA

Publicity Chairs

Yu-Wei Chan, Providence University, Taiwan
 Byoungwook Kim, Dongshin University, Korea

Industrial Cooperation Chairs

Yong Woo Lee, Ssangyong Information and Communications Corp., Korea
 Sung Chul Yu, LG Hitachi Ltd., Korea
 Sung Gil Kim, VAIV Company Inc., Korea
 Bong Sang Seo, ALL4LAND Co., Ltd., Korea
 Se Jong Kim, SJ Info and Communications Co., Ltd., Korea
 Tae Yoon Kwon, Neighbor System Co., Ltd., Korea
 Han Su Cheon, Selim TSG Co., Ltd., Korea
 Seokhong Min, Mindata Corporation, Korea

Local Arrangement Chairs

Khanthanou Luangxaysana, National University of Laos, Laos
 Phonexay Vilakone, National University of Laos, Laos
 Jun-Ho Huh, Korea Maritime and Ocean University, Korea
 Jinho Park, Dongguk University, Korea

Program Committee

Andrew Kusiak, The University of Iowa, USA
 Chang Wu Yu, Chung Hua University, Taiwan
 Chin-Fu Kuo, The National Kaohsiung University, Taiwan
 Debajyoti Mukhopadhyay, Balaji Institute of Telecom & Management, India
 Dion Hoe-Lian Goh, Nanyang Technological University, Singapore
 Jehn-Ruey Jiang, National Central University, Taiwan
 Paprzycki Marcin, Polish Academy of Sciences, Poland
 Qian Yu, University of Regina, Canada
 Tzung-Pei Hong, National University of Kaohsiung, Taiwan
 Marco Listanti, DIET, Roma, Italy
 Sun-Yuan Hsieh, National Cheng Kung University, Taiwan
 Yue-Shan Chang, National Taipei University, Taipei
 Ahmed EL Oualkadi, Abdelmalek Essaadi University, Morocco
 Jung Hanmin, KISTI, Korea
 Liu Chuan-Ming, National Taipei University of Technology, Taipei
 Ali Shahrabi, Glasgow Caledonian University, UK
 Bo-Chao Cheng, National Chung-Cheng University, Taiwan
 Chang Yao-Chung, National Taitung University, Taiwan
 Deok Gyu Lee, Seowon University, Korea
 Min Choi, Chungbuk National University, Korea

Hang-Bae Chang, Chung-Ang University, Korea
Hong-Jun Jang, Jeonju University, Korea
Imad Saleh, University of Paris 8, France
Jinhyun Ahn, Jeju National University, Korea
Lai Kuan-Chu, National Taichung University, Taiwan
Lam-for Kwok, City University of Hong Kong, Hong Kong
Seung-Ho Lim, Hankook University of Foreign Studies, Korea
Ki Yong Lee, Sookmyung Women's University, Korea
Koojoo Kwon, Baewha Women's University, Korea
Yeong-Seok Seo, Yeungnam University, Korea
Yeong Wook Yang, Hanshin University, Korea
Yoo-Jae Won, Chungnam National University, Korea
Shi Junren, Chongqing University of Post and Telecom, China
Liu Ping, Chongqing University of Post and Telecom, China
Chen Junsheng, Chongqing University of Post and Telecom, China
Chia-Hung Yeh, National Sun Yat-sen University, Taiwan
Jin Gon Shon, Korea National Open University, Korea
Ka Lok Man, Xi'an Jiaotong-Liverpool University, China
Peng Ran, Chongqing University of Post and Telecom, China
Xu Zhang, Chongqing University of Post and Telecom, China

Contents

Random Forest in Federated Learning Setting	1
Thien-Phuc Doan, Bong Jun Choi, Kihun Hong, Jungsoo Park, and Souhwan Jung	
TL-SMOTE: Re-balancing Data in Federated Learning for Anomaly Detection	11
Linh Nguyen-Thuy, Long Nguyen-Vu, Jungsoo Park, Kihun Hong, and Souhwan Jung	
Gaussian Pseudospectral Optimization-Driven Neural Network Planning of Obstacle Avoidance Trajectory	19
Tianyi Chen, Chongchong Tan, Ping Liu, and Mingyuan Bian	
Research on Synchronous Control Method of Dual Motor Based on ADRC Speed Compensation	27
Fanqiao Zeng, Changhao Piao, Kuang Liang, and Jianmin Dang	
MDagg: A New Aggregation Method Using Mahalanobis Distance	37
Songi Gwak and Souhwan Jung	
Graphormer-NAS-X: A Novel Graphormer-Based Neural Architecture Search Method	45
Hu Pan, Yuan Hu, Ying Yang, and Xu Zhang	
Secondary Salient Feature-Based GNN for Few-Shot Classification	53
Chuang Liu, Wei Hou, Xinyue Wang, and Xu Zhang	
Mixup Cluster Contrastive Learning Over Harder Samples for Unsupervised Person Re-identification	61
Xin He, Hongwen Zhou, Die Dai, and Xu Zhang	
UWB and IMU Fusion Positioning Based on ESKF with TOF Filtering	69
Changhao Piao, Houshang Li, Fan Ren, Peng Yuan, Kailin Wan, and Mingjie Liu	

Observation Noise Covariance Matrix Initialization-Based Objective State Estimation for Kalman Filter Using SVR 79
Junhu Chen, Mingjie Liu, Fan Ren, Peng Yuan, Jianbin Chen, Yiwei Ma, Tai Liu, and Changhao Piao

Development of the Fire Analysis Framework for the Thermal Power Plant 89
Chai-Jong Song and Jae-Yun Park

Mortality Forecasting Using Data Augmentation 97
Kyu-suh Park, Ji Su Park, Joosoo Maeng, and Jin Gon Shon

Spatial–Temporal Reservation-Based Multi-vehicle Cooperative Speed Guidance 105
Changhao Piao, Jia He, and Yongsheng Wang

A Research on Quantum Repeater Platform Using Drone: Utilization Methods and Considerations 117
Tae Woo Kim, Seo Yeon Moon, Il Chul Shin, and Jin Sang Park

Implementation of Virtual Sensor for Semiconductor Process Verification Using Machine Learning 125
Jeong Il Shin, Ji Su Park, and Jin Gon Shon

An Analysis of the Longitudinal Changes in the Determining Factors for Adolescents’ Self-esteem with Random Forests 133
Hyejoo Lee and Euihyun Jung

Design of Multi-channel Ethernet System Based on FPGA 141
Yougu Ale, Yang Xu, and Chao Wu

Design of 2D to 3D Pose Estimation Using NeRF Image View Synthesis 149
Chan Park, Hyungju Kim, and Nammee Moon

Pet Multiple Behavior Recognition Through Sensor Data Synthesis 155
Hyungju Kim, Chan Park, Jaehyeon Park, and Nammee Moon

Research on AFS/DYC Coordinated Control Strategy for Four-Wheel Independently Driven Electric Vehicle Based on In-Wheel Motor 161
Changhao Piao, Hao Wang, Ping Liu, Junren Shi, and Jianmin Dang

A Study on the Construction of Crime Prevention System Using the Hadoop Model of Big Data- with Focus on Utilizing Big Data and CPTED Technology in Gimcheon-Si - 173
Bong-Soo Kim, Dong-wook Kang, Eun-Joo Lee, Dae-bum Lee, and Ji-Woong Ryu

Application of Multi-variable Double-Layer MPC in Energy-Saving Operation of Central Chiller 179
 Xiang Li and Kaibi Zhang

Study on AEB Control Strategy Considering Vehicle Load Variation ... 187
 Changhao Piao, Xu Han, Junren Shi, and Tai Liu

Enhancing System Utilization by Dynamic Reallocation of Computing Nodes 197
 Seungmin Lee, Hee Jin Jang, and Min Ah Kim

Efficient Detection of Skin Cancer Using Deep Learning Techniques and a Comparative Analysis Study 203
 Mehtab Hashim, Asad Masood Khattak, and Imran Taj

Autonomous Cooperative Driving Protocol Based on the Security Status of IVN 211
 Daekyo Shin and Soohyun Jang

Confidence Calibration of Hand Gesture Recognition Networks Based on Ensemble Learning 219
 Zongjing Cao, Yan Li, and Byeong-Seok Shin

Optimization Strategy for Residential Heating Based on Model Predictive Control 227
 Zhiwei Jia, Yue Gu, and Yanxue Li

A Study on the Development of a Crop Disease Diagnosis Mobile App Service Based on Deep Learning 235
 Dong In Lee, Ji Hwan Lee, Hoon Seo Jung, Seung Ho Jang, Seung Gi Ryu, Jiann An Yang, Ji Hwan Oh, Jae Won Kim, Se Jong Oh, and Ill Chul Doo

A Study on the Development of SDGs Implementation Analysis System at Universities Based on NLP 241
 Ha Jin Kwon, Hye Ju Yun, Hyun Bin Jin, Ji Eun Lee, Ju-Young Lee, Sang Hoon Oh, Kwang Wook Seo, Se Jong Oh, and Ill Chul Doo

Object Detection Improvements on Skin Burn Image Data via Data Augmentation and Semi-supervised Learning 247
 Dae Hwa Hong, Sin Yeong Kim, Min Kyu Park, Se Jong Oh, and Ill Chul Doo

Establishment of NLP-Based Greenwashing Pattern Detection Service 253
 Ji Sue Kim, Jun Bo Sim, Yoon Ju Kim, Min Kyu Park, Se Jong Oh, and Ill Chul Doo

Multiclass Classification of Online Comments Based on Toxicity Scale 261
 Jo eun Kang, Jisun Kim, Hyeon su Do, Sihyun Oh, Si hwa Lee, Min jee Jeung, Min seo Choi, Se Jong Oh, and Ill Chul Doo

Developing a Speech Recognition Service for Korean Speakers with Dysarthria 267
 Go Woon Choi, Min Hyuk Kim, Ha Young Park, Young Hae Choi, Se Jong Oh, and Ill Chul Doo

Korean Twitter Bot Detection via Deep Learning 273
 Hyun Seop Yoon, Se Yeon Yang, Dong Hyuk Yoo, Myung Hak Kim, Seo Yeon Im, Se Jong Oh, and Ill Chul Doo

Development of Literature Short-Form System Using Emotional Keyword 279
 Sang Won Roh, Sung Cheol Kim, Tae Ho Kim, Jeong Woo Bae, Ji In Han, Se Jong Oh, and Ill Chul Doo

English Pronunciation Correction Service for Hearing-Impaired People: BETTer, Focusing on the Personalized Speech Model 285
 Seon Hong Park, Hyun Jin Park, Yoon Jeong Kim, Young Woong Park, Jeong Min Lee, Ye Rin Jung, Seok Ryu Jo, Se Jong Oh, and Ill Chul Doo

Improving the Performance of Judicial Precedent Search by Fine-Tuning S-BERT 291
 Gilsik Park and Juntae Kim

Machine Learning-Enabled Distributed Framework for Attack Detection in Social Networks 299
 Sangthong Yotxay, Abir E. L. Azzaoui, and Jong Hyuk Park

Analysis of Noise Reduction Measures for 220 kV Substation Based on Acoustic Simulation Technology 305
 Haitao Sun, Yu Yang, and Peijie Liu

A Study on Metaverse Realistic Content Education Platform Using Deep Learning 313
 Hyunsook Lee and Sekyoung Youm

Strategies for Establishing u-health Care Center Through Research on u-health Service Awareness and Intention to Use 319
 Sekyoung Youm

Method of Predicting the Danger Zone of Tidal Greening Using Machine Learning 325
 Gayoung Kim

A Study on the Location Selection Model for a Remote Image Dispenser Using Machine Learning 331
Teachang Ryu, Changhoon Choi, and Ga-young Kim

Passive Indoor Human Tracking Using Commodity Wi-Fi 337
Lingchao Guo, Jun Ma, and Yue Xu

Research and Application on 5G-V2X-Based Enhanced Automated Valet Parking Key Technologies 347
Janghyun Baek, Yongcheol Ro, Junhyek Jang, Daekyo Shin, Sanghun Yoon, and Soohyun Jang

Sensitivity Analysis and Multi-optimization on Building Performance in Tube-Style Dwellings in the Rural Area of Northern Zhejiang Province 355
Li Wenxuan, Yin Zilu, Fan Bingxin, Wang Meiyan, and Long Jiang

Study on the Cooling Effect of Green Space in Shandong Peninsula Coastal Cities Based on Multi-source Remote Sensing Data 363
Yansu Qi, Xuefei Li, and Weijun Gao

Quantification Korean Skin Color Using Multi-color Space for Objective Skin Tone Classification 373
Seula Kye and Onseok Lee

Segmentation of UV Images Based on U-Net for Precisely Analyzing Cutaneous Pigmentation: A Pilot Study 381
Yejin Wee, Taeyeon Gil, and Onseok Lee

Korean-Vietnamese Neural Machine Translation at Sub-word Level 387
Van-Hai Vu, Quang-Phuoc Nguyen, and Cheol-Young Ock

Classification of the Factors Influencing Center of Pressure Using Machine Learning and Wavelet Analysis 395
Xueli Ning, Young Kim, Jong Gab Ho, Xin Guo, Rathna Damsmoun, A. Hyun Jung, and Se Dong Min

A Local Assessment Approach for Diagnosis of Thermal Imaging-Based Palmar Hyperhidrosis 403
Jiwon Lee, Juhyun Kim, and Onseok Lee

A Study on Internet Annotation Analysis System Using Opinion Mining 409
Yen-Tae Jeong and Byung-Hoon Jeon

Real-Time Pressure Feedback Sensor Designed for Manual Therapy—A Pilot Clinical Feasibility Study 415
Eun Hye Jo, Young Kim, Seungwan Jang, Hyun Young Lee, Seungmin Jeong, Ye Eun Kong, and Se Dong Min

A Blockchain-Based CCP Data Integrity Auditing Protocol for Smart HACCP 423
Siwan Noh, Muhammad Firdaus, Zhuohao Qian, and Kyung-Hyune Rhee

BPFL: Blockchain-Enabled Distributed Edge Cluster for Personalized Federated Learning 431
Muhammad Firdaus, Siwan Noh, Zhuohao Qian, and Kyung-Hyune Rhee

Healthcare Big Data Analysis Using Representative National Community Health Survey of 2021: Is Income Change Due to Covid-19 Pandemic Associated with Unmet Healthcare Needs in S. Korea 439
Yeong Jun Lee, Se Hyeon Myeong, and Sun Jung Kim

River Health Assessment Based on an Artificial Neural Network 449
Zonglin Pang, Yufeng Liu, Zitong Liu, and Chao Liu

Optimization of Low Impact Development Scenarios Based on Computer-Aided Design 457
Fanglin Zhu, Ruolan Mu, Han Li, and Xuefei Li

The Development of Expressive Vocabulary of Korean Learners as a Heritage Language 465
Sin Hye Nam, Chaerin Jang, and Sunyoung Kim

Applying Process Mining Techniques in Sewage Treatment Plant Management 471
Bin Shao, Guoqing Ni, Jingbo Zhao, and Sheng Miao

Intelligent Sewage Treatment Control System Based on Digital Twin ... 479
Hao Jiang, Zikun He, Shuqi Liu, Yang Hai, Chao Liu, and Sheng Miao

Visual Simulation and Demonstration of Building Quality Improvement in Existing Residential Areas Driven by Integrated Information System 487
Yue Fan, Li Zhang, Qiong Zhang, Guangxun Cui, Yueer He, and Yu Zhao

A Study on the Integrated Database of Sino-Korean Words in <Goryeo sa(高麗史)> 493
Joo-hee Lee, Qianhang Tang, and Ryang-jin Kim

A Case Study on Data Mining for Korean Painting Poetry 499
Haeyoung Park and Younghoon An

A Study on Digital Content Using Yun Dong-Ju’s Works 505
Minji Choi, Hwa-Young Jeong, Young-Hoon An, and Peerapat Luanmongkol

Digital Transformation of Simulation Process to Support Decision-Making on Green Remodeling 511
Woo-Seung Yun, Wontaek Ryu, Ye-Bon Jo, and Hyuncheol Seo

An Application of Deep Neural Network in Facade Planning of Coastal City Buildings 517
Xue Sun, Yue Wang, Ting Zhang, Yin Wang, Haoyue Fu, Xuechun Li, and Zhen Liu

A Study of Students Studying Abroad to Tang in the Late Silla’s Pass an Imperial Examination and Representation Using Database 525
Weiyu Chen, Wenxuan Zhao, Shaoshuo She, Young-Hoon An, and Hwa-Young Jeong

Study on the Spatial Evolution of Rural Housing’s Courtyard Based on the Perspective of Space Syntax 531
Xingkang Chai and Yu Chen

The Syntactic Changes of Traditional Rural House and Its Influence by Modernization 539
Jiacheng Xu and Yu Chen

A Quantitative Study on the Spatial Mechanism of Zhejiang Traditional Dwellings Based on Space Syntax 549
Keyou Xu and Yu Chen

Construction and Verification of the Evaluation Scale for Architectural Design Scheme Based on Knowledge Engineering 555
Yingjie Liu, Lan Chen, Chuanlong Wang, Dongmiao Zhao, and Chao Liu

Development of Indoor Air Quality Management Using Automatic Control of Home Appliances 563
Yuko Kuma and Fumiya Sakai

A Model for Detection of the Susceptible in Hospital Contact Network Based on Social Network Analysis 569
Yixuan Yang, Doo-Soon Park, Seok-Hoon Kim, Sony Peng, Sophort Siet, Sadriddinov Ilkhomjon Rovshan Ugli, and Phonexay Vilakone

A Modern Recommendation System Survey in the Big Data Era 577
Sony Peng, Doo-Soon Park, Dae-Young Kim, Yixuan Yang, Sophort Siet, Sadriddinov Ilkhomjon Rovshan Ugli, and HyeJung Lee

Privacy-Aware Intelligent Healthcare Services with Federated Learning Architecture and Reinforcement Learning Agent 583
Prohim Tam, Inseok Song, Seungwoo Kang, and Seokhoon Kim

ML/SDN-Based MEC Resource Management for QoS Assurances 591
Seyha Ros, Chanthol Eang, Prohim Tam, and Seokhoon Kim

Movie Recommendation System Using Community Detection Based on the Girvan–Newman Algorithm 599
 Sadriddinov Ilkhomjon Rovshan Ugli, Doo-Soon Park, Daeyoung Kim, Yixuan Yang, Sony Peng, and Sophort Siet

Efficient Container Management Scheme Based on Deep Learning Model 607
 Byeonghui Jeong, Jueun Jeon, Seungyeon Baek, and Young-Sik Jeong

Actual Resource Usage-Based Container Scheduler for High Resource Utilization 611
 Sihyun Park, Jueun Jeon, Byeonghui Jeong, Kyuwon Park, Seungyeon Baek, and Young-Sik Jeong

Solving Pre-erection Area Block Placement Problem Using Deep Reinforcement Learning 615
 Yoon Hyun Kim and Gijoo Yang

Design of Reliable Electronic Document Based on HTML and Blockchain 623
 Andrew Hwang and Ji Su Park

Multi-armed Bandit-Based Rate Control with Logarithmic Rates in CSMA/CA Wireless Networks 631
 Soohyun Cho

The Design and Implementation of Medicine Recognition System Using the TensorFlow Lite 639
 Ji un Park, Hee Jun Moon, and Dong Hyun Kim

Development of Ubiquitous Flip Learning Model Based on Backward Design 645
 Kil Hong Joo and Nam Hun Park

KoBART-Based Hybrid Literacy Improvement System for Each Level for Efficient Reading 653
 Soobin Cho and Seungmin Park

Extracting Low-Power Code Pattern Method Through Power Measurement of Software Code 661
 Bo Kyung Park and Su Nam Choi

A Study on Improved Pet Behavior Monitoring Through Data Imputation Using Multiple Data 669
 Jinah Kim, Hyungju Kim, Chan Park, Jeong-Hyeon Park, and Nammee Moon

Deep Learning SPIN Pattern Outlier Detection for Integrated Dynamic Rotary Machine 677
 Jieun Kang, Subi Kim, and Yongik Yoon

Optimized Vehicle Fire Detection Model Based on Deep Learning 685
 ByoungGun Park, Ji Su Park, and YounSoon Shin

Smarteria: Intelligent Cafeteria Using Multimodal Data 693
 Yuri Seo, Teh-Jen Sun, Seol Roh, In-hun Choi, Gi-dae Hong,
 Hong-Ju Jeong, Seung-Woo Jeong, Choong Seon Hong,
 Seong-Bae Park, and Eui-Nam Huh

**Post-training with Data Augmentation for Improving T5-Based
 Question Generator** 703
 Gyu-Min Park, Seong-Eun Hong, Choong Seon Hong,
 and Seong-Bae Park

**A Memory Replay-Based Continual Learning Utilizing Class
 Representative and Class Boundary Data** 711
 Jun-Beom Lee, Choong Seon Hong, and Seong-Bae Park

**Automatic Classification for Representative Spatio-temporal-Based
 Event Document Using Machine Learning** 717
 Byoungwook Kim, Yeongwook Yang, Ji Su Park, and Hong-Jun Jang

**Case Monitoring for Secondary Battery Transfer with Enhanced
 Fire Prevention and Stability** 723
 Kihak Lee, Heji Lee, Md. Jahirul Islam, Saetbyeol Kim,
 Wolyoung Kim, Tae-Young Byun, and Bonghwan Kim

**Effective Use of Knowledge Graphs in a Language Representation
 Model** 731
 Chanwook Min, Hyungsun Na, Yeji Song, Jinhyun Ahn,
 and Dong-Hyuk Im

**The Performance Analysis of Lightweight Neural Networks
 for Salient Object Detection** 739
 Thien-Thu Ngo, Eui-Nam Huh, and Choong-Seon Hong

**Pendulum: An I/O Efficient Algorithm for Nearest Surrounding
 Queries in 3D Space** 747
 Kyeongseok Hyun and Jaehwa Chung

**Proposal of Telehealth Service Reference Models as a Prerequisite
 for Cybersecurity Risk Analysis** 755
 In Hye Lee and Jungha Jin

**A Challenge of the Representative Spatiotemporal Document
 Classification and Spatiotemporal Information Extraction** 765
 Yeongwook Yang, Ji Su Park, Hong-Jun Jang, Byoungwook Kim,
 and Jin Gon Shon

**IoT Sensor Predictive Analysis System Using Apache Spark
in Edge Cloud Environment** 771
Gi-dae Hong, Young-Min An, Seol Roh, and Eui-Nam Huh

**MetaOps: Metaverse Operations for Large-Scale Metaverse
Services on Distributed Cloud** 779
Teh-Jen Sun, Seung-Woo Jeong, Hong-Ju Jeong, Choong Seon Hong,
Seong-Bae Park, and Eui-Nam Huh

**A Study on the Improvement of Internet Comment Accuracy
Using Deep Learning** 789
Heeji Park, Jimin Ha, and Jungho Kang

About the Editors

Ji Su Park received his B.S. and M.S. degrees in Computer Science from Korea National Open University, Korea, in 2003 and 2005, respectively, and Ph.D. degrees in Computer Science Education from Korea University, 2013. He is currently Professor in Department of Computer Science and Engineering from Jeonju University in Korea. His research interests are in mobile grid computing, mobile cloud computing, cloud computing, distributed system, computer education, and IoT. He is employed as Editor-in-Chief of *KIPS Transactions on Software and Data Engineering* and Managing Editor of *Human-centric Computing and Information Sciences* (HCIS), *The Journal of Information Processing Systems* (JIPS). In addition, he has been serving as Guest Editor for international journals by some publishers: Springer, MDPI, and KIPS. He has received “best paper” awards from the CSA2018, BIC2021, and CSA2021 conferences and “outstanding service” awards from CUTE2019 and BIC2020. He has also served as Chair, Program Committee Chair or Organizing Committee Chair at several international conferences including World IT Congress, MUE, FutureTech, CSA, CUTE, and BIC.

Dr. Laurence T. Yang is Academic Vice-President and Dean of School of Computer Science and Technology, Hainan University, China. He has published 250+ papers in the Cyber-Physical-Social Systems area on top IEEE/ACM Transactions including 8 and 41 papers as top 0.1% and top 1% highly cited ESI papers, respectively. He has been involved actively as Chair of four IEEE technical committees and Steering Chair for 10+ IEEE international conferences, invited to give around 50 keynote talks at various international conferences and symposia.

He is Member of Academia Europaea, the Academy of Europe (2021), and Fellow of Institution of Engineering and Technology (2020), Institute of Electrical and Electronics Engineers (2020), Engineering Institute of Canada (2019), Canadian Academy of Engineering (2017), respectively. He is Recipient of the John B. Stirling Medal (2021) from Engineering Institute of Canada, IEEE Sensor Council Technical Achievement Award (2020), IEEE Canada C. C. Gotlieb Computer Medal (2020), etc.

Dr. Yi Pan is currently Chair Professor and Dean of Faculty of Computer Science and Control Engineering at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China, and Regents' Professor Emeritus at Georgia State University, USA. He served as Chair of Computer Science Department at Georgia State University from 2005 to 2020. He has also served as Interim Associate Dean and Chair of Biology Department during 2013–2017. Dr. Pan joined Georgia State University in 2000, was promoted to Full Professor in 2004, named Distinguished University Professor in 2013, and designated Regents' Professor (the highest recognition given to Faculty Member by the University System of Georgia) in 2015.

Dr. Pan is Member of the Academy of the United Nations Sciences and Technology Organization, Foreign Member of Ukrainian Academy of Engineering Sciences, Member of European Academy of Sciences and Arts, Fellow of American Institute for Medical and Biological Engineering, Fellow of Institute of Engineering Technology, and Fellow of Royal Society for Public Health.

Jong Hyuk Park received Ph.D. degrees in Graduate School of Information Security from Korea University, Korea, and Graduate School of Human Sciences from Waseda University, Japan. From December 2002 to July 2007, Dr. Park had been Research Scientist of R&D Institute, Hanwha S&C Co., Ltd., Korea. From September 2007 to August 2009, he had been Professor at the Department of Computer Science and Engineering, Kyungnam University, Korea. He is now Professor at the Department of Computer Science and Engineering and Department of Interdisciplinary Bio, IT, and Materials, Seoul National University of Science and Technology (SeoulTech), Korea. Dr. Park has published about 200 research papers in international journals and conferences. He has been serving as Chair, Program Committee, or Organizing Committee Chair for many international conferences and workshops. He is Steering Chair of international conferences—MUE, FutureTech, CSA, CUTE, UCAWSN, and World IT Congress-Jeju. He is Editor-in-Chief of *Human-centric Computing and Information Sciences* (HCIS) by Springer, *The Journal of Information Processing Systems* (JIPS) by KIPS, and *Journal of Convergence* (JoC) by KIPS CSWRG. He is Associate Editor/Editor of 14 international journals including *JoS*, *JNCA*, *SCN*, *CJ*, and so on. In addition, he has been serving as Guest Editor for international journals by some publishers: Springer, Elsevier, John Wiley, Oxford Univ. press, Emerald, Inderscience, and MDPI. He got the best paper awards from ISA-08 and ITCS-11 conferences and the outstanding leadership awards from IEEE HPCC-09, ICA3PP-10, IEE ISPA-11, PDCAT-11, and IEEE AINA-15. Furthermore, he got the outstanding research awards from the SeoulTech, 2014. His research interests include IoT, human-centric ubiquitous computing, information security, digital forensics, vehicular cloud computing, multimedia computing, etc. He is Member of the IEEE, IEEE Computer Society, KIPS, and KMMS.

Random Forest in Federated Learning Setting



Thien-Phuc Doan , Bong Jun Choi , Kihun Hong , Jungsoo Park ,
and Souhwan Jung 

Abstract Federated learning has become popular nowadays since being started by Google in developing their Google Keyboard solution. Original federated learning proposed how to secure aggregating the gradient from the local training process of joining mobile devices. However, other machine mechanisms without gradient updating, such as random forest, are not supported. In this work, we propose **PriForest**, a new approach for building random forest in the federated learning setting. This framework covers all random forest processes, including building the forest, updating the forest, and using the forest for prediction. For providing privacy reserving, we apply a differential-privacy scheme while making bootstrapping sets, which modifies the value of existing records or appends new records before constructing trees. The experiment result shows that our approach is applicable with an affordable noise rate, which controls the sensitivity of the noise adding method.

Keywords Random forests · Differential privacy · Federated learning

T.-P. Doan · K. Hong · J. Park · S. Jung (✉)
School of Electronic Engineering, Soongsil University, Seoul 06978, Korea
e-mail: souhwanj@ssu.ac.kr

T.-P. Doan
e-mail: phucdt@soongil.ac.kr

K. Hong
e-mail: khong@ssu.ac.kr

J. Park
e-mail: jspark@ssu.ac.kr

B. J. Choi
School of Computer Science and Engineering, Soongsil University, Seoul 06978, Korea
e-mail: davidchoi@soongil.ac.kr

1 Introduction

In 2016, Google proposed a new collaborative learning system called federated learning (FL) [1]. Its core concept is to train a machine learning model using sub-updates from dispersed datasets, which speeds up model training and prevents direct privacy leakages. FL is gaining popularity among academics thanks to its attractive features, and it is used in many applications, including voice recognition and e-healthcare. Furthermore, since the discovery of the data reconstruction attack against sub-updates, FL has continued to explore ways to improve its privacy-preserving technique.

As a type of tree model, random forest (RF) [2] is very popular and widely used in the risk control scenarios. It provides an effective solution to the performance shortage of the decision tree model by integrating the multiple trees. Because of its efficiency and lower complexity than other deep learning methods, RF is commonly employed in the industry. Figure 1 illustrates a peer-to-peer protocol implemented in a Federated Learning setup to tackle the issue of money laundering detection faced by banks from different countries. In this protocol, banks hold a variety of user data associated with transactions, but they cannot share data due to privacy concerns. This presents a challenge for distributed training, despite the banks having a common set of features.

Decentralized learning models have been proposed to solve this problem. Secure multi-party computation (secure MPC) [3] is proposed. MPC is a cryptographic technique that distributes a calculation across numerous parties without allowing one participant to access the data of the others. However, the cost of implementation, particularly for the peer-to-peer approach, is enormous. On the other hand, random forest creates bunch of decision trees, which can be considered as a data encoder. Feng and his colleagues [4] have successfully developed an auto-encoder using tree ensemble learning, which demonstrates that a training dataset may be recovered from a small number of decision trees.

Random forest is excellent at decreasing expenses, but it has some weaknesses to security attacks, such as data extension attack. As a result, in this paper, we aim

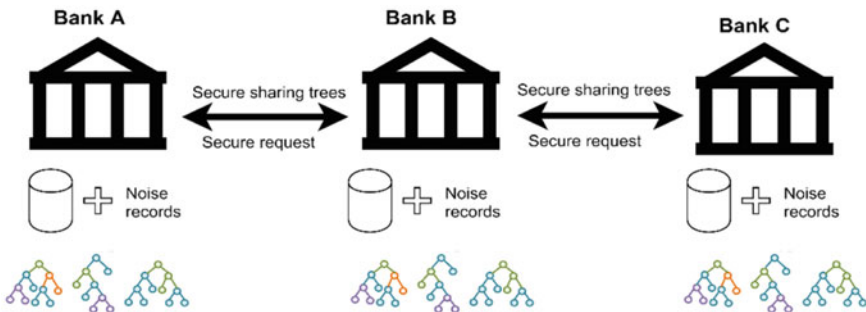


Fig. 1 Peer-to-peer random forest in federated learning setting

to use a differential-privacy technique [5] to ensure that the model remains private when we use random forest in federated learning setting.

2 Related Works

Recent studies [1, 6–10] investigated algorithms that can be applied to decision tree classification in MPC.

Federated Forest [11] is a privacy-preserving machine learning model built on CART trees and bagging. The algorithm’s detailed steps are as follows. To begin, the active party selects a subset of features and samples randomly from the complete dataset and discreetly notifies each passive party of the selected features and sample IDs. Secondly, each passive party determines the local ideal split feature and impurity value and sends this encrypted data to the active party. The active party can determine the optimum feature for splitting by comparing the purity improvement values of all alternative features. Finally, the active party informs the passive party with the best splitting capability. The samples will be divided into left–right by the passive party.

SecureGBM is a secure multi-party gradient boosting based on semi-homomorphic encryption [12]. At first, SecureGBM must align the data of the active and passive parties, and researchers typically use private sets intersection techniques [13]. Second, SecureGBM trains a decision tree to fit the negative direction of the gradient as a gradient boosting method. Like the federated forest, passive parties must calculate local feature separation and relay that information to the active party. As the node is introduced to the decision tree, the active party can have the best split feature. Finally, repeat the procedures above.

RevFRF [14] presented a revocable federated RF architecture and established the revocable FL idea. They provided federated RF without revealing private data in construction and prediction processes. To do this, researchers applied homomorphic encryption (HE)-based secure protocols to achieve privacy-preserving. Besides, they conducted thorough research to demonstrate that RevFRF could handle all the security issues associated with a revocable FL framework.

PHPR [10] proposed a privacy-preserving and high-accurate outsourced disease predictor on random forests. The data first is encrypted at the data owner side with secure multi-party computing schemes. Then, the central server constructs the tree securely using an additive homomorphic cryptosystem. The inference phase is also securely operated with encrypted data sent from client to server.

3 System Architecture

3.1 Motivating Example

Banks in different places want to link their data to solve money laundering problems. However, transaction data is confidential because it contains sensitive information about customers. If data is shared without encryption, it exposes the risk of being stolen by malicious users. In contrast, banks will not be able to use encrypted data to train the AI model. Even using a homomorphic scheme, stakeholders consume a lot of computational resources. It is necessary to have a lightweight AI algorithm-based framework based on a FL setting and ensure privacy preservation for sensitive data at the lowest cost.

3.2 Architecture Design

We propose **PriForest**, a random forest constructing and sharing framework for peer-to-peer HFL. By adopting differential privacy, **PriForest** ensures the privacy and stability of the serving model. We also describe how to protect the serving model from integrity attacks by applying cryptography tools to establish secure channels among peers. The **PriForest** constructing and sharing protocol is shown in Fig. 2.

Suppose that we are given a set of clients (e.g., banks) $C = \{C_1, C_2, \dots, C_N\}$. Each client contains a dataset D_n that has the same features set $X \in R^d$ to other clients, i.e., we set up a horizontal federated learning problem. In this work, we assume that the client’s local data is non-i.i.d., and the dataset contains sensitive information. All the clients want to share their training model to enhance the accuracy of solving specific tasks while ensuring the data’s privacy. The local dataset $D_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$ contains sensitive records (e.g., transferring money records).

Firstly, all cryptography and tree constructing configurations will be established in the *Config Initializing* phase. This phase ensures a secure communication channel by establishing elliptic-curve Diffie–Hellman (ECDH) hyperparameters [15]. In addition, the random forests model configuration has to be shared among parties.

The *bootstrapping sets generation* is the key method in this work. To provide privacy-preserving, we leverage the differential-privacy technique to draw bootstrapping sets from the client’s local dataset. Algorithm 1 describes making a bootstrap set for constructing one decision tree. First, a standard bootstrapping set D^* is generated

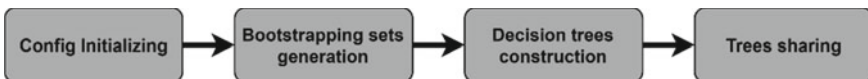


Fig. 2 Tree construction steps of **PriForest**

in the same way in a typical random forest algorithm. D^* contains N^* records are drawn from D . Note that from here, D and D^* belong to a single client that represents the local dataset. The N^* is controlled by *example subsampling* rate r_{exp} . The number of attributes used to make bootstrapping set $d < d^*$, where d is the number of attributes of D .

Algorithm 1 PriForest bootstrapping set generation

Input: Local dataset D has N records and d features. The attribute subsample rate r_{att} , the example subsample rate

r_{exp} , the noise rate r_{noise}

Output: The bootstrapping dataset for constructing a tree.

$D^* = \emptyset$

for i in range($0, N \cdot r_{\text{exp}}$) **do**

D^* .append($D[\text{rand}()]$)

end for

for i in range($0, N \cdot r_{\text{exp}} \cdot r_{\text{noise}}$) **do**

$\text{random_att} \leftarrow \text{rand}(0, d.\text{ratt})$

$\text{random_rec} \leftarrow \text{rand}(0, d.\text{rexp})$

$\text{ndrand} \leftarrow \text{rand}(0, d.\text{rexp})$

if (noise samples replacement method), then

$D^*[\text{random_rec}, \text{random_att}] = D^*[\text{ndrand}, \text{random_att}]$

else if (noise appending method) then

$\text{tmp_rec} \leftarrow D^*[\text{ndrand}, \text{random_att}]$

$D^*.\text{append}(\text{tmp_rec})$

end if

end for

We adapt the traditional differential-privacy method described in [16] when the size of D^* and D'^* is equal. In that method, D'^* is constructed from D^* by modifying the one-row so-called *noise samples replacement* method. The sensitivity of this method now is equal to 1. Note that we only modify one single value of the chosen row by replacing this element with the existing value of the corresponding feature in D . We then ensure the correct type of the modified element as well as keep the reasonable range of the value.

Besides the above traditional method, we also adapt the noise appending method. In this case, the size of D'^* could be larger than D^* . We first randomly choose one existing record in D^* , then duplicate this row but modify one single value of the new row using the same modifying way of the *samples replacement* method.

To control the sensitivity of the above methods, we set the noise rate r_{noise} for both methods. With the *noise samples replacement method*, there are $N \cdot r_{\text{exp}} \cdot r_{\text{noise}}$ modified records. With the *noise appending method*, there are $N \cdot r_{\text{exp}} \cdot r_{\text{noise}}$ new records.

Decision tree construction is conducted locally in the client system. The following hyperparameters configure the random forests model in this work:

- **num_trees**: Number of trees to be generated.
- **depth_limit**: The limitation of depth level while constructing trees. The rest will be cut off using tree pruning protocol.
- **example_subsample_rate** (r_{exp}): The rate to control the number of samples in bootstrapping sets.
- **attr_subsample_rate** (r_{att}): The rate to control the number of attributes in bootstrapping sets.
- **bootstrap_noise_rate** (r_{noise}): The rate of noise added to bootstrapping sets.

Finally, tree sharing is conducted when all clients have finished constructing their decision trees. This method can be called tree aggregation. Each client broadcasts its trees to the others. The tree object is compressed and signed by the responsible client to ensure the confidentiality and integrity of the aggregation. We will not discuss the details of compression and digital signature. We implement the state-of-the-art solution as default.

4 Evaluation

4.1 Experimental Setup

We use the money laundering dataset available at Kaggle.¹ The dataset contains 2340 records about money transferring or cash-in types. We randomly split the dataset into a training set and a test set, with the ratio of the test set being 0.2. The training set is randomly split into multiple parts for multiple clients. We assume no client drops out during the training and trees sharing phases. We set the number of trees per client during tree construction as 10, and the limitation of the tree depth level is 10. The example subsample and attribute subsample rates are 0.4 and 1, respectively. In our experiment, we change the bootstrapping set noise rate to evaluate the effect of adding noise on the model's accuracy. The result is shown in Fig. 3.

4.2 Result

Both types of bootstrapping set generation reduce the accuracy of the model. The accuracy slightly changes when $r_{\text{noise}} < 1$ while significantly reducing when $r_{\text{noise}} > 1$ if we use the replacement method (a). There seems to be little change in the other

¹ <https://www.kaggle.com/datasets/maryam1212/money-laundering-data>.

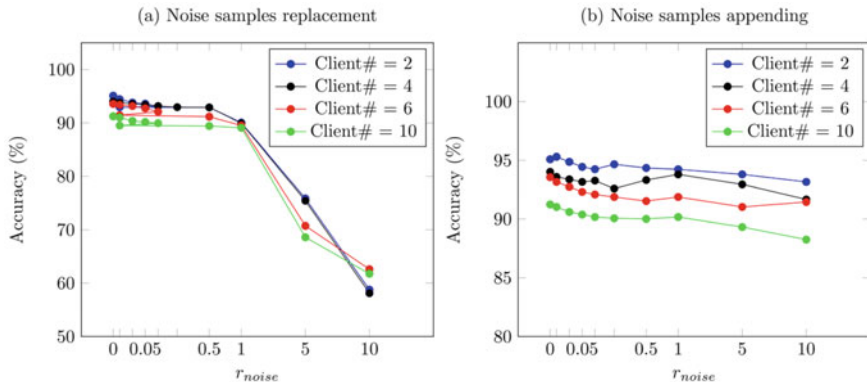


Fig. 3 Impact of bootstrapping set generation on the accuracy of the random forest in the federated learning setting

type (b), which appends the noise into the normal bootstrapping set. This can be explained as follows:

- (a) **Noise samples replacement (original differential-privacy method):** When $r_{\text{noise}} < 1$, only one attribute of at most $N \cdot r_{\text{exp}}$ changes the value. The change cannot guarantee the change of impurity, which leads to the different split of the decision tree. Therefore, the noise contribution in bootstrapping set generation is not much. When $r_{\text{noise}} > 1$, the changed record has more changing periods, which means there could be more than one attribute value that is changed. In that case, the tree construction is more different, which leads to the accuracy deduction.
- (b) **Noise samples appending:** The noise samples are now appended to the normal bootstrapping set. In that case, the original samples remain. However, the distribution of the dataset is slightly changed leading to a slight decrease in the model's accuracy.

5 Conclusion and Discussion

This work proposed **PriForest**, a differential-privacy approach to solving the problem of using random forests in a horizontal federated learning setting. By adding noise to the bootstrapping set in two ways, noise samples replacement and noise appending, this solution ensures sensitive data privacy. The result of this work experiment shows that with the suitable noise rate r_{noise} , the privacy is preserved while the random forest model remains an acceptable performance. This work also describes using cryptographic tools to protect the serving model from integrity and confidentiality attacks. However, this work needs more security analysis, and we design to let this task for future study.

Acknowledgements This research was partly supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2022-2020-0-01602) supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) and Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00952, Development of 5G Edge Security Technology for Ensuring 5G+ Service Stability and Availability).

References

1. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K (2017) Practical secure aggregation for privacy-preserving machine learning. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security (CCS'17). Association for Computing Machinery, New York, NY, pp 1175–1191. <https://doi.org/10.1145/3133956.3133982>
2. Ho TK (1995) Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, Aug 1995, vol 1. IEEE, pp 278–282
3. Canetti R, Feige U, Goldreich O, Naor M (1996) Adaptively secure multi-party computation. In: Proceedings of the twenty-eighth annual ACM symposium on theory of computing, July 1996, pp 639–648
4. Feng J, Zhou ZH (2018) Autoencoder by forest. In: Thirty-second AAAI conference on artificial intelligence, Apr 2018
5. Dwork C, McSherry F, Nissim K, Smith A (2006) Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference, Mar 2006. Springer, Berlin, Heidelberg, pp 265–284
6. Backes M, Berrang P, Bieg M, Eils R, Herrmann C, Humbert M, Lehmann I (2017) Identifying personal DNA methylation profiles by genotype inference. In: 2017 IEEE symposium on security and privacy (SP), May 2017. IEEE, pp 957–976
7. Bost R, Popa RA, Tu S, Goldwasser S (2014) Machine learning classification over encrypted data. Cryptology ePrint archive
8. Hamada K, Hasegawa S, Misawa K, Chida K, Ogishima S, Nagasaki M (2017) Privacy-preserving Fisher's exact test for genome-wide association study. In: International workshop on genome privacy and security (GenoPri), pp 99–102
9. Wu DJ, Feng T, Naehrig M, Lauter K (2015) Privately evaluating decision trees and random forests. Cryptology ePrint archive
10. Ma Z, Ma J, Miao Y, Liu X (2019) Privacy-preserving and high-accurate outsourced disease predictor on random forest. *Inf Sci* 496:225–241
11. Liu Y, Liu Y, Liu Z, Liang Y, Meng C, Zhang J, Zheng Y (2020) Federated forest. *IEEE Trans Big Data*
12. Feng Z, Xiong H, Song C, Yang S, Zhao B, Wang L et al (2019) SecureGBM: secure multi-party gradient boosting. In: 2019 IEEE international conference on big data (big data), Dec 2019. IEEE, pp 1312–1321
13. Pinkas B, Schneider T, Zohner M (2014) Faster private set intersection based on $\{OT\}$ extension. In: 23rd USENIX security symposium (USENIX security 14), pp 797–812
14. Liu Y, Ma Z, Yang Y, Liu X, Ma J, Ren K (2021) RevFRF: enabling cross-domain random forest training with revocable federated learning. *IEEE Trans Depend Secure Comput*

15. Bernstein DJ (2006) Curve25519: new Diffie–Hellman speed records. In: International workshop on public key cryptography, Apr 2006. Springer, Berlin, Heidelberg, pp 207–228
16. Dwork C, Rothblum GN, Vadhan S (2010) Boosting and differential privacy. In: 2010 IEEE 51st annual symposium on foundations of computer science, Oct 2010. IEEE, pp 51–60

TL-SMOTE: Re-balancing Data in Federated Learning for Anomaly Detection



Linh Nguyen-Thuy , Long Nguyen-Vu , Jungsoo Park , Kihun Hong , and Souhwan Jung 

Abstract Due to the significant growth of computing devices, attacks on the network have received increasing consideration. In traditional anomaly detection methods, an agent must collect all data to train the model, potentially leading to data leakage. Thus, the security challenges have encouraged the use of federated learning to address this issue by anomaly detection while improving the efficiency and privacy of the training models. However, in an extensive network, the imbalance of data in the training set for each client and the large volume of data point distribution between classes are significant challenges for training models. Thus, it is necessary to re-balance the training dataset before anomaly tasks. In this paper, we propose a re-balancing scheme for mitigating the impact of imbalanced training data. This work combines k-nearest neighbors with the Tomek link synthetic minority over-sampling method. While Tomek Link eliminates a pair of samples from two distinct classes (one majority class and one minority class) that are closest, k-nearest neighbors enrich the minority class by providing artificial examples in the minority class. Tomek link makes advantage of minority class data that k-SMOTE oversampled to obtain more accurate class clusters. Our experiments demonstrate the importance of acknowledging class imbalance.

Keywords Federated learning · Imbalanced data · Anomaly detection

L. Nguyen-Thuy · L. Nguyen-Vu · J. Park · K. Hong · S. Jung (✉)
School of Electronic Engineering, Soongsil University, Seoul, Korea
e-mail: souhwanj@ssu.ac.kr

L. Nguyen-Thuy
e-mail: linhntt@soongsil.ac.kr

L. Nguyen-Vu
e-mail: longnv@soongsil.ac.kr

J. Park
e-mail: jspark@ssu.ac.kr

K. Hong
e-mail: khong@ssu.ac.kr

1 Introduction

In recent years, the demand for information transmission between unauthorized devices expanded quickly, while released devices typically lack security concerns [1], which is easily attackable, making the network insecure. Anomaly detection is a crucial research topic in several domains and applications in network security, in which machine learning (ML) approaches have become indispensable to solve these kinds of issues. Typically, the data generated by devices must be transferred to a cloud system or a central server to train models using traditional centralized ML methodologies. When there is a considerable increase in the amount of data accessible, this activity may often be ineffective and wasteful.

The emergence of distributed ML systems, which let each device do training locally and only submit updates to the central server to maintain privacy, can overcome these constraints of standard ML. Federated learning (FL) is particularly promising for training decentralized machine learning models [2]. In such instances, the local display of the learning process and the fact that the data is never removed from the locations of the learning task participants assist in safeguarding the datasets' privacy. Despite the advantages, FL also faces one major challenge: data from various sources may contain their preferences, and the resulting diversity may make the convergence of the global model challenging and slow [3]. Moreover, the phenomenon of class imbalance happens with high frequency in real scenarios [4]. When the imbalance occurs, the number of samples in majority classes is massive compared to that in minority classes, directly impacting accuracy reduction in minority classes. Besides, in the case of anomaly detection, the type of anomalies observed by each device are distinct. However, the quantity of abnormalities often makes up a relatively tiny portion of the entire training set. Therefore, it poses a difficult challenge when those minority classes play an essential role beyond their proportion in data.

In this paper, we overcome the problem of anomaly detection where data imbalances happen in the supervised learning-based scenario. We propose the TL-SMOTE algorithm to re-balance the data using FL. TL-SMOTE adaptively improves the synthetic minority oversampling technique (SMOTE) by mixing this state-of-the-art re-balancing technique with another down-sampling method called Tomek link (T-link).

The remainder of the paper is organized as follows. Section 2 summarizes related work. Section 3 describes the proposal, and Sect. 4 is the experimental results. Finally, Sect. 5 presents the conclusions and future work.

2 Related Work

2.1 Federated Learning

Many studies have been done on organizing many devices for distributed learning, using both centralized and decentralized [5] approaches. In recent years, a growing number of local client devices (such as mobile phones) has participated in model learning. The local device training data becomes more personal and sensitive in these conditions. Federated learning has emerged as a viable approach for avoiding privacy leaks. In general, neither local participants nor federated learning's central server can see the individual gradient in plain form during training using these methods. Inferring information from specific clients is still exceedingly tricky, despite different inference attacks. As a result, we focus on doing so in this paper to remedy the class imbalance.

2.2 Re-balancing Data

In real-world categorization jobs, unbalanced data is a prevalent problem. When trained with unbalanced datasets, general machine learning models struggle, especially when distinguishing samples from the minority class. Re-balancing procedures are commonly used to overcome this constraint and increase model performance. Instead of focusing solely on total accuracy, several studies have developed additional metrics [6] to quantify model performance with class imbalance. There are three previous efforts to address class imbalance: data-level, algorithm-level, and hybrid methods.

Because they violate privacy rules, data-level approaches cannot be used in federated learning. Due to the mismatch between local and global data distributions, cost-sensitive approaches require specific knowledge of minority classes, treating them as positive classes, and calculating false positives and negatives to build a new loss form.

3 Methodology

This section describes TL-SMOTE, the proposed method for training ML-based anomaly detection systems with FL and imbalanced datasets.

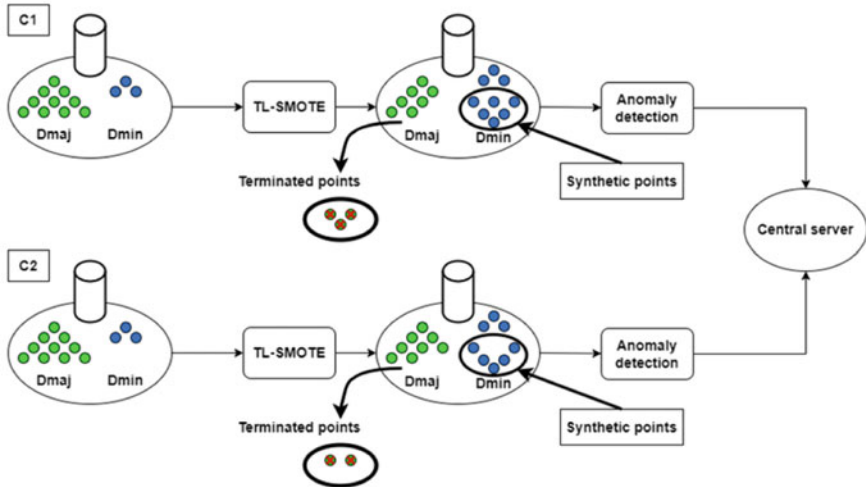


Fig. 1 System overview of TL-SMOTE: each client has a local dataset and a local anomaly detection model. Samples from both the majority (green dots) and minority (blue dots) classes

3.1 Computational Procedure

Figure 1 describes the operation of our system. TL-SMOTE will be deployed on the local devices. The local device can be a gateway that learns to detect any abnormal behaviors and trains a regional model based on local, on-device data to send the local parameters to the server. Overall, the TL-SMOTE system consists of three phases: (1) re-balancing, (2) optimization of the anomaly detector, and (3) reporting back to the central server.

3.2 TL-SMOTE

This method combines the k-SMOTE ability to generate synthetic data for minority classes and the Tomek links method to remove the data that are identified as Tomek links. The process of TL-SMOTE to re-balance the dataset is shown in Fig. 2.

Firstly, the k-SMOTE based on Smote algorithm selects the i samples in the minority class. It takes all the k -nearest neighbors instead of picking one at random,

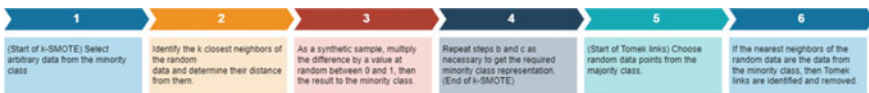


Fig. 2 Re-balancing process

denoted as $N = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k\}$, to produce the final synthetic data point using linear interpolation. The new equation can be re-written as

$$x_{\text{syn}} = x_i + \frac{1}{|N|} \sum (\hat{x}_n - x_i) \times \delta \quad (1)$$

where \hat{x}_n and x_i are both from minority classes, \hat{x}_n is one of the nearest neighbors from N and δ denotes a random number in $[0, 1]$. The synthetic data point x_{syn} will be added to the dataset for training.

After finishing, minority class examples that are oversampled can balance class distributions. However, class clusters are frequently not clearly defined since some majority class instances may be encroaching on the territory of minority classes. Because some situations can lead to an overfitting challenge, so Tomek links method is applied to the oversampled training set as a data cleaning technique to get better-defined class clusters.

The Tomek links method employs the rule to collect the pair of observations. Let $d(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j , where x_i and x_j denote the samples that belong to the minority and majority class, respectively. If there is no other sample x_k that satisfies the below condition

$$d(x_i, x_k) < d(x_i, x_j), \text{ or } d(x_j, x_k) < d(x_i, x_j) \quad (2)$$

then the pair of (x_i, x_j) is a Tomek link [7]. When two samples connect via a Tomek connection, either one is noisy, or both are borderline. In this case, instances from both classes are eliminated rather than only the examples from the majority class.

4 Experiments and Results

Our experiments correspond to FL's common framework [8], and we mainly implement the algorithms in PyTorch. We used the FEMNIST dataset [9], which is a well-known FL dataset with significant feature heterogeneity, to perform several experiments. We use ResNet18 a convolution neural network, with a local training batch size of 32, a learning rate $\lambda = 0.001$, and the SGD optimizer for the dataset.

4.1 Experiment Results

The detection accuracy and Ratio Lost [3] during the training process are shown in Figs. 3 and 4. If the accuracy is closer to 100%, our estimation is more accurate. After 100 iterations, the TL-SMOTE reaches a high detection accuracy of 99.8% on the training set and 93.90% on the testing set when using the TL-SMOTE method to balance the data. Most results are over 98% (from the fifth to the last round). Opposite,

finding a set of weights and biases that, on average, have minimal loss across all examples is the aim of training a model. Per our expectation, the TL-SMOTE model achieves a slight loss of 0.0017 after 100 epochs.

Moreover, from Figs. 3 and 4, compared with the Smote model, the TL-SMOTE model achieved a slight increase in accuracy while the loss value also tended to decrease slightly. Besides, Table 1 also shows the detailed comparison between Smote model and the extended TL-SMOTE model. Our model outperforms the traditional Smote algorithm in binary classification experiments.

Regarding security analysis, the Area Under the Curve (AUC) is a specific performance measure for anomaly detection based on the receiver operating characteristic curves (ROC curves) [1]. AUC values range from 0.5 to 1, with 1 denoting excellent detection and 0.5 indicating random detection. Table 1 displays an excellent AUC score of 0.9972, which means we can learn to detect anomalies not far from perfect (1.0).

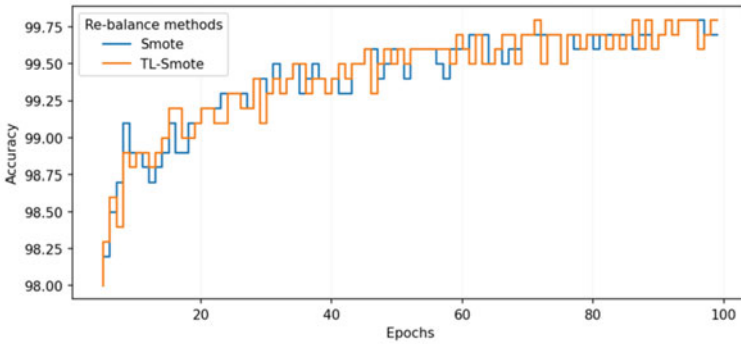


Fig. 3 Accuracy of the training model

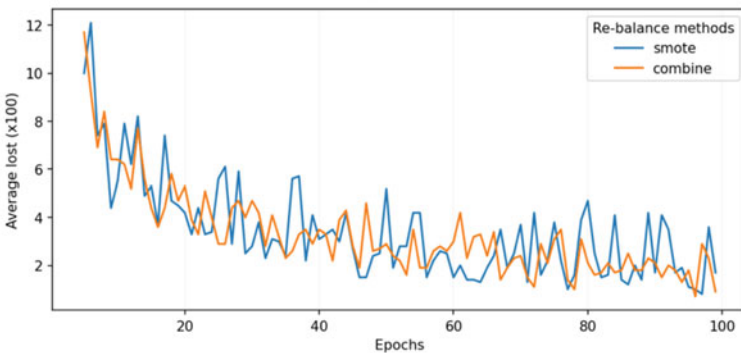
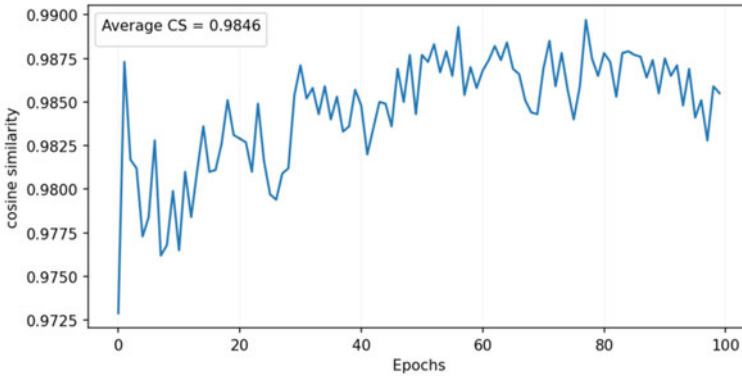


Fig. 4 Average ratio lost

Table 1 Comparison between TL-SMOTE and Smote model

Values	TL-SMOTE model	Smote model
Training accuracy (after 100 epochs) (%)	99.8	99.7
Testing accuracy (%)	93.90	93.67
Average lost (after 100 epochs)	0.009	0.017
AUC score	0.997200	0.997051

**Fig. 5** Similarity between our estimation of the global data composition and the ground truth

For each training round's evaluation, the server computes a data composition vector v_j . By using a cosine similarity (CS) score, we can compare it to the ground truth, which is represented by the value V .

$$CS_j = (v_j \cdot V) / (\|v_j\| \|V\|) \quad (3)$$

The more precise our assessment is, the nearer it is to 1. Figure 5 shows the result of the CS score, and we can see that our estimation of the data composition is exceptionally close to the ground truth. The similarity score is greater than 0.98 most of the time and is over 0.98 on average. These outcomes indicate the potency of our strategy.

5 Conclusion

In this paper, we have introduced TL-SMOTE, which addresses the class imbalance issue in federated learning for anomaly detection. Using multiple closest neighbors to compute linear interpolation for the generation of complicated synthetic points, we have enhanced SMOTE and made the points more unexpected and available for

sharing among several clients taking part in FL. By incorporating up-sampling and down-sampling, we succeed in handling imbalanced data issues.

Acknowledgements This research was partly supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2022-2020-0-01602) supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) and Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00952, Development of 5G Edge Security Technology for Ensuring 5G+ Service Stability and Availability).

References

1. Albanese D, Filosi M, Visintainer R, Riccadonna S, Jurman G, Furlanello C (2013) minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics* 29(3):407–408
2. Wang H, Muñoz-González L, Eklund D, Raza S (2021) Non-IID data re-balancing at IoT edge with peer-to-peer federated learning for anomaly detection. In: *Proceedings of the WiSec'21*
3. Wang L, Xu S, Wang X, Zhu Q (2021) Addressing class imbalance in federated learning. In: *AAAI*
4. Rao RB, Krishnan S, Niculescu RS (2006) Data mining for improved cardiac care. *ACM SIGKDD Explor Newsl* 8(1):3–10
5. Sergeev A, Del Balso M (2018) Horovod: fast and easy distributed deep learning in TensorFlow. *arXiv preprint [arXiv:1802.05799](https://arxiv.org/abs/1802.05799)*
6. Wang S, Liu W, Wu J, Cao L, Meng Q, Kennedy PJ (2016) Training deep neural networks on imbalanced data sets. In: *2016 international joint conference on neural networks (IJCNN)*. IEEE, pp 4368–4374
7. Pereira RM, Costa YMG, Silla CN Jr (2020) MLTL: a multi-label approach for the Tomek Link undersampling algorithm. *Neurocomputing* 383:95–105
8. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D (2016) Federated learning: strategies for improving communication efficiency. *CoRR [arXiv:1610.05492](https://arxiv.org/abs/1610.05492)*
9. Caldas S, Wu P, Li T, Konečný J, McMahan HB, Smith V, Talwalkar A (2018) Leaf: a benchmark for federated settings. *arXiv preprint [arXiv:1812.01097](https://arxiv.org/abs/1812.01097)*

Gaussian Pseudospectral Optimization-Driven Neural Network Planning of Obstacle Avoidance Trajectory



Tianyi Chen, Chongchong Tan, Ping Liu, and Mingyuan Bian

Abstract This work proposed an improved Gaussian pseudospectral optimization data-driven neural network trajectory planning algorithm to shorten the planning time and improve the speed of obstacle avoidance trajectory. Firstly, the obstacle avoidance trajectory planning optimal control problem is established by deriving the kinematics model of unmanned vehicle and obstacle model. Gaussian pseudospectral optimization is then derived in detail to tackle the infinite time optimal control problem (OCP) and obtain the obstacle avoidance optimization data. By using the neural network training data set, an intelligent planning method for on-line calculation of obstacle avoidance trajectory is designed. Simulation results show that the proposed method can efficiently decrease trajectory planning time.

Keywords Gauss pseudospectral optimization · Data driven · Obstacle avoidance · Neural network

1 Introduction

Unmanned driving has become very popular in recent years, the intelligent driving technology has important research implications and great application value in reducing the incidence of road traffic accidents [1]. The local path planning [2], also known as obstacle avoidance path planning, considers the geometric relationship between local vehicles and obstacles to find a path to avoid collision with obstacles. It is the key content of the planning part, which is related to the safety and comfort of unmanned vehicles. Therefore, how to correctly calculate the optimal driving trajectory of the unmanned vehicle to complete the driving tasks is the urgent

T. Chen · C. Tan · P. Liu (✉)
College of Automation, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China
e-mail: liuping_cqupt@cqupt.edu.cn

M. Bian
School of Vehicle and Mobility, Tsinghua University, Beijing 100091, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_3

matter. At present, the relevant research institutions and universities have carried out extensive research on unmanned local trajectory planning [3, 4] and carried out a lot of work. As one of the most effective numerical optimization methods, the pseudospectral method [5] has been applied to many optimal control problems of unmanned vehicles [6, 7]. Due to faster speed, longer obstacle avoidance distance and complex vehicle conditions, the actual obstacle avoidance planning requires that the obstacle avoidance planning time be as short as possible to ensure effective obstacle avoidance within a safe distance. Therefore, how to improve the efficiency of Gauss pseudospectral method in solving obstacle avoidance problem has become the primary task. In order to shorten the planning time and improve the safety and speed of obstacle avoidance planning, an improved Gaussian pseudospectral optimization data-driven neural network trajectory planning algorithm is proposed in this paper. Firstly, the Gaussian pseudospectral optimization method is used to solve the typical trajectory planning problem, and the optimized obstacle avoidance trajectory is stored in the database through off-line calculation. Then the neural network is then used to train the obstacle avoidance trajectory in the database, and the median filter is used to obtain a similar trajectory. In order to verify the effectiveness of this method, numerical experiments are carried out on the basis of the trajectory planning model of unmanned vehicle, and the performance of this method is compared with traditional Gaussian pseudospectral method.

2 Unmanned Vehicle Trajectory Planning Model

The vehicle kinematic model [8] is a set of time-varying mathematical models that describe the position and orientation of vehicle. On this basis, the vehicle is placed on the earth and a reference point is randomly selected as the origin of the coordinate system to establish the XOY coordinate system. It is assumed that the whole obstacle avoidance time is $[t_0, t_f]$. Accordingly, the center point of the rear axle is defined as $(X_r(t), Y_r(t))$, and the center point of front axle is presented by $(X_f(t), Y_f(t))$; $\varphi(t)$ is the heading angle; $\delta(t)$ represents the deflection angle of the front wheel, the center speed of the rear axle is $V_r(t)$, and the center speed of the front axle is $V_f(t)$; l is a fixed value and is the wheelbase. Its kinematic model can be described as:

$$\begin{bmatrix} \dot{X}_r(t) \\ \dot{Y}_r(t) \\ \dot{\varphi}(t) \end{bmatrix} = \begin{bmatrix} \cos \varphi(t) \\ \sin \varphi(t) \\ \tan \delta(t)/l \end{bmatrix} V_r(t) \quad (1)$$

Then P-norm is employed in this work to construct a geometric model. Meanwhile, assume that the geometric center of the obstacle is (x_t, y_t) , Y_{\min} represents the lower boundary of the road and Y_{\max} is the upper boundary of the road. Accordingly, to avoid the obstacle (x_t, y_t) , the center point of the rear axle $(X_r(t), Y_r(t))$ should satisfy the following equation with using the P-norm:

$$\left(\left[\left| \frac{X_r(t) - x_r}{r_x} \right| \right]^p + \left[\left| \frac{Y_r(t) - y_r}{r_y} \right| \right]^p \right)^{1/p} \geq 0, \quad Y_{\min} \leq Y_r(t) \leq Y_{\max} \quad (2)$$

where p represents the order of graph and is chosen as 4 in this work. Finally, the goal of control is to find a feasible state curve that minimizes performance indicators and is presented by an optimal control problem (OCP) stated below:

$$\begin{aligned} \min J &= \phi(X_{t_f}, Y_{t_f}, \varphi_{t_f}, t_f) + \int_{t_0}^{t_f} g(X_r(t), Y_r(t), \varphi(t), \delta(t), V_r(t), t) dt \\ \text{s.t. Eqs. (1)–(2)} \\ S(X_{t_0}, Y_{t_0}, \varphi_{t_0}, X_{t_f}, Y_{t_f}, \varphi_{t_f}) &= 0 \\ Y_{\min} \leq Y_r(t) \leq Y_{\max}, \quad -\pi/2 \leq \delta(t) \leq \pi/2 \\ V_{\min} \leq V_r(t) \leq V_{\max}, \quad t_0 \leq t \leq t_f \end{aligned} \quad (3)$$

where V_{\min} and V_{\max} are the minimal and maximal speeds of the vehicle, respectively. Meanwhile, $S(X_{t_0}, Y_{t_0}, \varphi_{t_0}, X_{t_f}, Y_{t_f}, \varphi_{t_f})$ represents the initial and terminal constraints.

3 Optimization-Driven Neural Network Planning

3.1 Gauss Pseudospectral Optimization

Let $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]^T = [X_r(t), Y_r(t), \varphi(t)]^T$, $\mathbf{u}(t)$ denotes the control vector. Therefore, OCP is formulated into the trajectory optimization problem (TOP) stated below:

$$\begin{aligned} \min J &= \phi(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t) dt \\ \text{s.t. } \dot{\mathbf{x}}(t) &= f(\mathbf{x}(t), \mathbf{u}(t), t), \quad t \in [t_0, t_f] \\ G_0(\mathbf{x}(t_0), \mathbf{x}_0) &= 0, \quad T(\mathbf{x}(t_f), \mathbf{x}_T) = 0 \\ C(\mathbf{x}(t), \mathbf{u}(t), t_0, t_f) &\leq 0, \quad t_0 \leq t \leq t_f \\ \mathbf{x}_L \leq \mathbf{x}(t) \leq \mathbf{x}_U, \quad \mathbf{u}_L \leq \mathbf{u}(t) \leq \mathbf{u}_U \end{aligned} \quad (4)$$

Since t_f is a time free variable, let $T = t_f$. In this work, let $t = T\tau$, $\tau \in [0, 1]$. Then, the new state and control variables are defined as $\tilde{\mathbf{x}}(\tau) = \mathbf{x}(T\tau) = \mathbf{x}(t)$, $\tilde{\mathbf{u}}(\tau) = \mathbf{u}(T\tau) = \mathbf{u}(t)$. Therefore, it can be obtained that:

$$\begin{aligned}
\min \quad & J = \phi(\tilde{\mathbf{x}}(1), 1) + \int_0^1 T \times g(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), \tau) d\tau \\
\text{s.t.} \quad & \dot{\tilde{\mathbf{x}}}(\tau) = T \times f(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), T\tau), \quad \tau \in [0, 1] \\
& G_0(\tilde{\mathbf{x}}(0), \mathbf{x}_0) = 0, \quad T(\tilde{\mathbf{x}}(1), \mathbf{x}_T) = 0 \\
& C(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), 0, 1) \leq 0, \quad \mathbf{x}_L \leq \tilde{\mathbf{x}}(\tau) \leq \mathbf{x}_U, \quad \mathbf{u}_L \leq \tilde{\mathbf{u}}(\tau) \leq \mathbf{u}_U \quad (5)
\end{aligned}$$

Based on Legendre Gauss (LG) collocation theory [9], the approximation of state vector is obtained on the LG collocation points:

$$\begin{aligned}
\mathbf{x}(\tau) \approx \mathbf{X}(\tau) &= \sum_{i=1}^{N+1} L_i(\tau) \mathbf{X}(\tau_i) = \sum_{i=1}^{N+1} L_i(\tau) \mathbf{X}_i \\
L_i(\tau) &= \prod_{j=1, j \neq i}^{N+1} \frac{\tau - \tau_j}{\tau_i - \tau_j} = \frac{b(\tau)}{(\tau - \tau_i) \dot{b}(\tau_i)} \quad (6)
\end{aligned}$$

where $b(\tau) = \prod_{i=1}^{N+1} (\tau - \tau_i)$. The state and control vector are approximately expressed by $\mathbf{U}(\tau) \approx \mathbf{U}(\tau) = \sum_{i=1}^{N+1} L_i(\tau) \mathbf{U}_{\tau_i}$ and $\dot{\mathbf{x}}(\tau) \approx \dot{\mathbf{X}}(\tau) = \sum_{i=1}^{N+1} \dot{L}_i(\tau) \mathbf{X}_i$. Finally, problem (5) then is transformed into the following nonlinear programming (NLP) problem:

$$\begin{aligned}
\text{Min}_{\mathbf{X}, \mathbf{U}, T} \quad & J = \phi(\tilde{\mathbf{x}}(1), 1) + T \times \sum_{k=1}^N w_k g(\mathbf{X}(s_k), \mathbf{U}(s_k), s_k, -1, 1) \\
\text{s.t.} \quad & \sum_{i=0}^N D_{k,i} \mathbf{X}_i - \frac{T}{2} f(\mathbf{X}_k, \mathbf{U}_k, s_k; -1, 1) = 0 \\
& G_0(\mathbf{X}_0, -1) = 0, \quad T(\mathbf{X}_T, 1, T) = 0 \\
& C(\mathbf{X}_k, \mathbf{U}_k, T, s_k; -1, 1) \leq 0 \\
& \mathbf{x}_L \leq \mathbf{X}_k \leq \mathbf{x}_U, \quad \mathbf{u}_L \leq \mathbf{U}_k \leq \mathbf{u}_U, \quad k = 1, 2, \dots, N \quad (7)
\end{aligned}$$

Usually, gradient-based NLP methods can be used to solve problem (7) with high precision [10].

3.2 Neural Network Trajectory Planning

With using the Gauss pseudospectral optimization method to solve the TOPs, plenty of optimal trajectories can be obtained. Therefore, the optimization data set then can be established. In this work, the neural network tool is employed to train the data and output a similar trajectory according to the actual situation of the vehicle.

The structure of the proposed planning method is shown in Fig. 1 and the detailed implementation steps are stated below. In this work, the function fitting results and error histogram obtained are shown in Fig. 2.

- **Step 1:** Use Gaussian pseudospectral optimization method to solve TOPs;
- **Step 2:** Establish the obstacle avoidance optimization data base;
- **Step 3:** Employ the neural network tool to construct the planning model and use the data set to train the model;
- **Step 4:** Optimize the parameters of the model;
- **Step 5:** Output the obstacle avoidance trajectory.

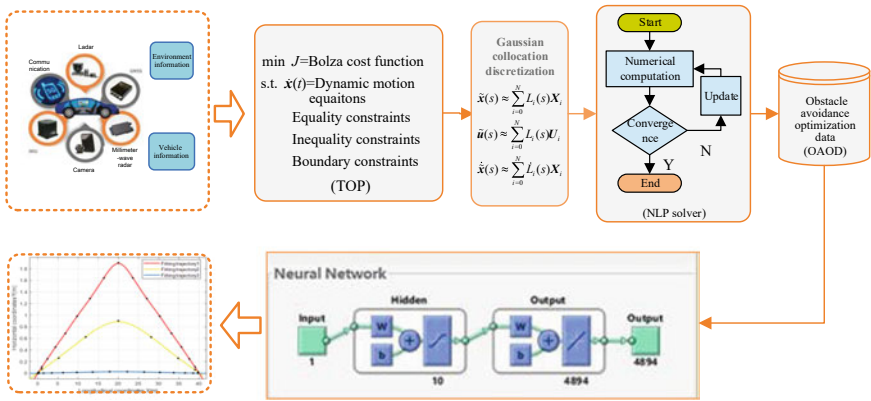


Fig. 1 Optimization-driven neural network obstacle avoidance planning

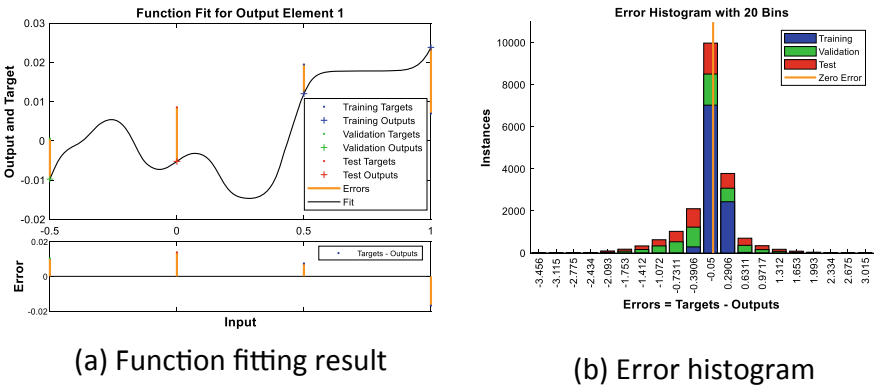


Fig. 2 Neural network planning results

4 Numerical Tests

Simulation experiments are carried out on an unmanned vehicle model and the traditional Gaussian pseudospectral method (GPM) is used for comparison. All simulation tests are carried out by MATLAB software with Intel core i5/2.5 GHz CPU processor and DDR3L/1600 MHz 8 GB memory personal computer. Correspondingly, the parameter values of vehicle and constraint parameters are given in Table 1.

In the simulation test, it is assumed that the vehicle is moving with a constant speed of 10 km/h and the obstacle is a static object. Furthermore, the performance index is avoiding the obstacle with minimal time. To verify the performance of proposed neural network-median filtering planning method (NN-MF), different obstacle positions are chosen for testing. Specifically, the center of the obstacle is set as 0.3, 0.5 and 0.9 m away from the horizontal line. The test results are given in Table 2. It can be found that both the two methods can calculate the obstacle avoidance trajectories, however, the computation time of neural network planning is much lower than that of GPM, revealing the on-line efficiency of proposed strategy. Meanwhile, the average error between the two trajectories is very small, showing the effectiveness of proposed method for obstacle avoidance. Accordingly, the trajectory curves of 3 situation tests are shown in Fig. 3.

Table 1 Parameters of the trajectory planning mathematical model

Symbol	Value	Symbol	Value	Symbol	Value
V_{\min}	10 [km/h]	Y_{\min}	- 2 [m]	r_y	0.9 [m]
V_{\max}	10 [km/h]	Y_{\max}	6 [m]	r_x	2.4 [m]
φ_{t_0}	0 [rad]	(x_{t_0}, y_{t_0})	(20, 0) [m]	l	2 [m]
φ_{t_f}	Free	(X_{t_f}, Y_{t_f})	(0, 0) [m]	(X_{t_f}, Y_{t_f})	(40, 0) [m]

Table 2 The test result comparison of two methods

Method	Obstacle distance [m]	Obstacle avoidance time [s]	Computation time [s]	Average error [m]
GPM	0.9	14.493	2.98	0.0372
NN-MF		14.544	0.196	
GPM	0.5	14.481	6.56	0.0632
NN-MF		14.524	0.209	
GPM	0.3	14.455	4.27	0.0558
NN-MF		14.645	0.139	

Bold means the result is better between the two methods

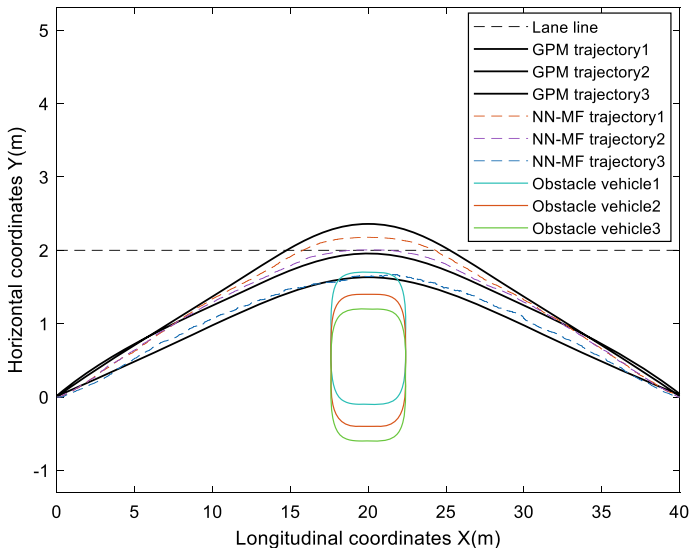


Fig. 3 Trajectory planning curves of 3 situation tests

5 Summary

In this paper, an improved trajectory planning optimization method is proposed. The main idea of this paper is to construct the trajectory data by using Gaussian pseudospectral optimization method, so as to drive the neural network to obtain the on-line obstacle avoidance trajectory. Based on the kinematic model and obstacle model, the optimal control problem of obstacle avoidance trajectory planning is established. Then Gaussian collocation discretization is used to deal with the optimal control problem, and the discretized nonlinear programming is obtained. On this basis, the Gaussian pseudospectral optimization method is used to establish the obstacle avoidance optimization database and solve the typical obstacle avoidance optimization control problem off-line. Combining off-line optimization with on-line neural network planning, a data-driven neural network planning algorithm for obstacle avoidance trajectory of unmanned vehicle is designed to realize on-line obstacle avoidance trajectory planning under different vehicle conditions. Experimental results show that compared with GPM trajectory planning, this method can reduce the calculation time by more than 90% and obtain the similar trajectory.

Acknowledgements This work was supported by the National key R&D program (No. 2022YFE0101000).

References

1. Wang H, Wang B, Liu B et al (2016) Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle. *Robot Auton Syst* 88:71–78
2. Shi P, Liu Z, Liu G (2022) Local path planning of unmanned vehicles based on improved RRT algorithm. In: Presented at the 2022 4th Asia Pacific information technology conference
3. Oleksak K, Wu Y, Abella M et al (2021) Trajectory optimization of unmanned aerial vehicles for wireless communication with ground terminals. In: AIAA Scitech 2021 forum, Shanghai
4. Shibata K, Shibata N, Nonaka K et al (2018) Model predictive obstacle avoidance control for vehicles with automatic velocity suppression using artificial potential field. *IFAC-PapersOnLine* 51(20):313–318
5. Li Y, Chen W, Yang L (2021) Multistage linear Gauss pseudospectral method for piecewise continuous nonlinear optimal control problems. *IEEE Trans Aerosp Electron Syst* 57(4):2298–2310
6. Wei K (2010) Rate of convergence for the Legendre pseudospectral optimal control of feedback linearizable systems. *J Control Theory Appl* 8(4):391–405
7. Xu Z, Shi P, Zhao Y (2015) Optimal reconfiguration control of electromagnetic spacecraft formation using Gauss pseudospectral method. In: 2015 2nd international conference on information science and control engineering, Virtual Event
8. Kang CM, Lee SH, Chung CC (2018) Multirate lane-keeping system with kinematic vehicle model. *IEEE Trans Veh Technol* 67(10):9211–9222
9. Long X, Jiong Y, Xinggao L et al (2017) An effective simultaneous approach with variable time nodes for dynamic optimization problems. *Eng Optim* 49(10):1761–1776
10. Loxton R, Teo KL, Rehbock V (2011) Robust suboptimal control of nonlinear systems. *Appl Math Comput* 217(14):6566–6576

Research on Synchronous Control Method of Dual Motor Based on ADRC Speed Compensation



Fanqiao Zeng, Changhao Piao, Kuang Liang, and Jianmin Dang

Abstract To solve the problem of inconsistent speed response time of dual motors under load disturbance, a synchronous control method of dual motors based on the combination of virtual spindle and fuzzy-LADRC (VS-FLADRC) is proposed in this paper. In this method, the speed error between two motors is corrected and compensated by the fuzzy linear active disturbance rejection speed compensator, which is input into the virtual spindle together with the target speed to achieve accurate control, so as to solve the problems of the double motor synchronous control system in practical application, such as hysteresis and poor load characteristics. The experimental results show that VS-FLADRC can significantly improve the speed response ability and synchronization control precision of the synchronous control system under load disturbance, also has strong anti-interference ability and robustness, which effectively realizes the precision collaborative control of dual motors.

Keywords Dual-motor synchronous control · Virtual spindle · LADRC

1 Introduction

With the continuous improvement of industrial automation level, dual-motor synchronous cooperative control is increasingly used in industrial production, the response speed, synchronization accuracy, and anti-interference ability of two-motor cooperative control system directly affect the efficiency and quality of industrial production [1]. There are two main problems in the current dual-motor synchronous control: one is the tracking accuracy of the motor to the target speed, the second is the synchronization error between the two motors [2]. The choice of compensation method for speed synchronization error of dual-motor synchronous control system

F. Zeng (✉) · C. Piao · K. Liang
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: S200332011@stu.cqupt.edu.cn

J. Dang
Chongqing Changan Automobile Co. Ltd., Chongqing 400023, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_4

will directly affect the stability and synchronization of the control system. Therefore, it is of great significance to improve the control performance of dual-motor cooperative control from the perspective of control structure and control strategy.

At present, the dual-motor synchronous control structure includes parallel control, master-slave control, cross-coupling control [3], virtual spindle control [4], etc. In recent years, domestic and foreign scholars have also applied control theories such as neural network control [5], sliding mode control [6], and active disturbance rejection control [7] in order to improve the motor speed regulation performance. In Ref. [8], neural network is used to realize adaptive synchronous control of dual motors, which has strong learning ability, however, the control structure is similar to parallel control, and the error is large under load disturbance. In Ref. [9], a model predicted velocity controller was designed to improve the anti-disturbance performance of the system, but it was difficult to set many parameters. In Ref. [10], the sliding mode control structure is improved, the speed of the system does not oscillate during operation, but the algorithm is relatively complex and difficult to be widely applied.

Considering the shortcomings of the above control methods, this paper proposes a dual-motor synchronous control method with virtual spindle and fuzzy-LADRC, also verifies the feasibility of the control method through experiments. This method retains the advantages of virtual spindle control, gives full play to the maximum performance of differential effects, has strong robustness, improves the performance of the dual-motor synchronous control system in practical applications, such as speed response capability, anti-disturbance capability, synchronous control accuracy, and so on, which is conducive to the wide application of the dual-motor synchronous control system in the automation field.

2 Construction of Dual-Motor Synchronous Control System

The virtual spindle control is to feed back the driving torque of the slave shaft to the forward channel of the spindle controller to realize the coupling feedback of parameters between the spindle and the slave shaft, so as to overcome the system unsynchronization caused by the asynchronous transient speed when the speed changes [11]. The structure of the dual-motor cooperative control system is shown in Fig. 1.

According to the mechanical principle, the angular acceleration output by the spindle in Fig. 1 and its torque meet the following relation:

$$T - (T_1 + T_2) = J_m \dot{\omega}_m. \quad (1)$$

In this case, the driving torque of the virtual axis is calculated as follows:

$$T = b_m(\omega_r - \omega_m + u) + K_m \int (\omega_r - \omega_m + u) dt, \quad (2)$$

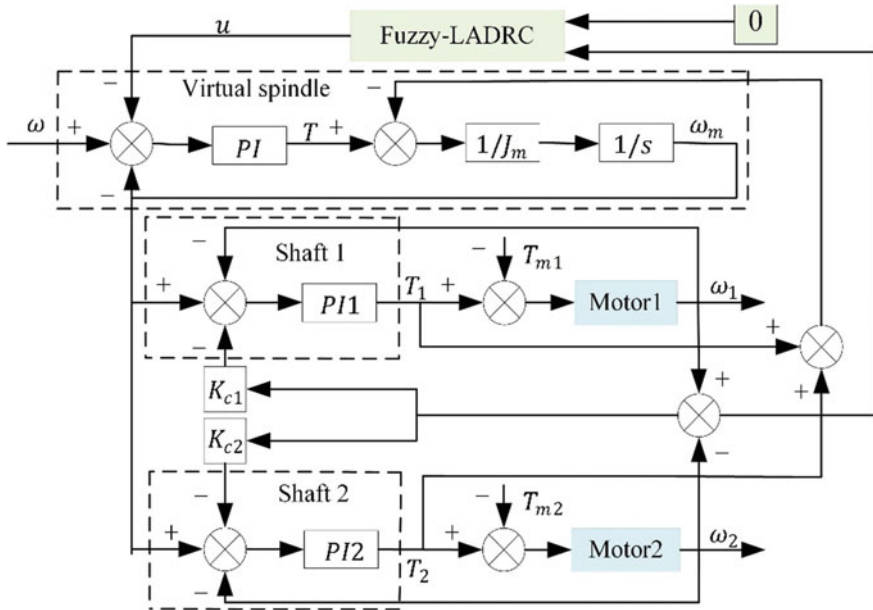


Fig. 1 Structure of dual-motor cooperative control system

where T is the driving moment of the imaginary axis, T_1 and T_2 are the torques fed back to the imaginary axis by each motion axis, J_m and $\dot{\omega}_m$ are imaginary axes of inertia and angular acceleration, respectively, b_m and K_m are the attenuation coefficient and elastic coefficient of the virtual axis, respectively, ω_r is the reference angular velocity, u is the output of fuzzy linear active disturbance rejection controller.

The formula for calculating the driving torque of each movement axis is

$$T_i = b_{si}(\omega_m - \omega_i - K_{ci} \Delta\omega) + K_{si} \int (\omega_m - \omega_i - K_{ci} \Delta\omega) dt, \quad (3)$$

where K_c is the velocity compensation coefficient of the motion axis, and $K_{c1} > 0 > K_{c2}$, $\Delta\omega$ is the speed difference between two axes of motion, b_s and K_s are the attenuation coefficient and elastic coefficient of the motion axis, respectively.

3 Design of Speed Compensator

The LADRC does not depend on the tracking object model, the tracking effect is obvious, the response speed is fast, and it has a good engineering application prospect [12]. Although the differential tracker can improve the dynamic performance of the control system, it will also increase the high-frequency gain of the system, resulting

in the decrease of the anti-interference ability of the system. In this section, the differential tracker is improved and the fuzzy-LADRC is designed.

3.1 LADRC

The LADRC consists of a linear extended state observer (LESO) and a linear state error feedback control law (LSEF) [13]. LADRC has the characteristics of strong robustness and easy implementation in practical operation, which also makes active disturbance rejection technology overcome the problem of parameter tuning. LESO's equation of state is only related to observer bandwidth λ_c , and LSEF can also be parameterized to make its control law only related to controller bandwidth λ_0 . The control quantities of LADRC are

$$u = -(\lambda_c^2 z_1 + 2\lambda_c z_2 - z_3) + \lambda_c^2 r, \quad (4)$$

where r is the target speed error value, z_1 , z_2 , and z_3 are the estimated quantities of LESO.

3.2 Design of Fuzzy-LADRC

In order to reduce the influence of noise interference on the stability of the control system, the fuzzy controller is used to control the LADRC's differential time constant. The structure of fuzzy controller is shown in Fig. 2.

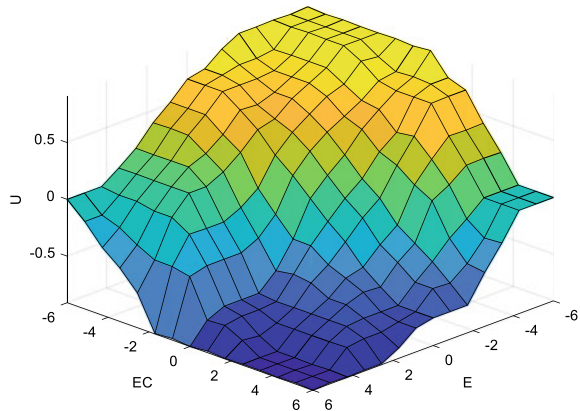
Integrating the computational volume and control accuracy, the fuzzy sets of both input and output are set to 7. The speed error E between two motors and the rate of change of speed error EC are chosen as the input of the fuzzy controller, and the theoretical domains of E and EC are noted as $\{-6, 6\}$, and the theoretical domain of the fuzzy controller output U is $\{-1, 1\}$. Considering the control accuracy and response speed, the membership function of the edge with large error is selected as the gentle Z-type, which can improve the control precision, and the membership function of the middle part is triangle-type, which can improve the response speed. The fuzzy control rule surface is shown in Fig. 3.

The fuzzy output derived by fuzzy control rules is solved by gravity center method to obtain the exact output. At this time, the output of fuzzy linear active disturbance



Fig. 2 Structure of the fuzzy controller

Fig. 3 Rule surface of the fuzzy control



rejection controller is

$$u = -(\lambda_c^2 z_1 + K'_d z_2 - z_3) + \lambda_c^2 r. \quad (5)$$

4 Experimental Verification and Analysis

4.1 Experimental Verification

The experimental platform is an electric long slide rail suitable for car seats, including a pair of slide rails, a pair of slide seats, a pair of BLDCM, and a seat base, as shown in Fig. 4. The motor is fixed on the slide seat and connected to the slide rail through a flexible shaft. The left and right slide seats are used for fixing the seat base, and the travel of the whole slide rail is 420 mm. The BLDCM has a rated voltage of 12 V, a polar log of 7, and a maximum allowable speed of 9900 rpm. In practice, the parameters of each motor cannot be completely consistent, for the BLDCM with Hall sensor whose backemf waveform is close to sine wave, the switching frequency is selected as 50 kHz, and the vector control method of adaptive lead angle is adopted. This control method only needs the minimum stator current to generate the required torque, so as to reduce the copper loss and improve the efficiency. High-speed CANFD is adopted for communication between the two-motor controllers, and the sampling cycle is 5 ms. After each sampling, a synchronous operation will be carried out. The two motors will be distinguished by master and slave. VS-FLADRC and other common synchronous control methods were used to carry out synchronous control experiments of two motors, respectively, and the synchronous control effect of the cooperative control system without load disturbance and single motor under load disturbance was analyzed.

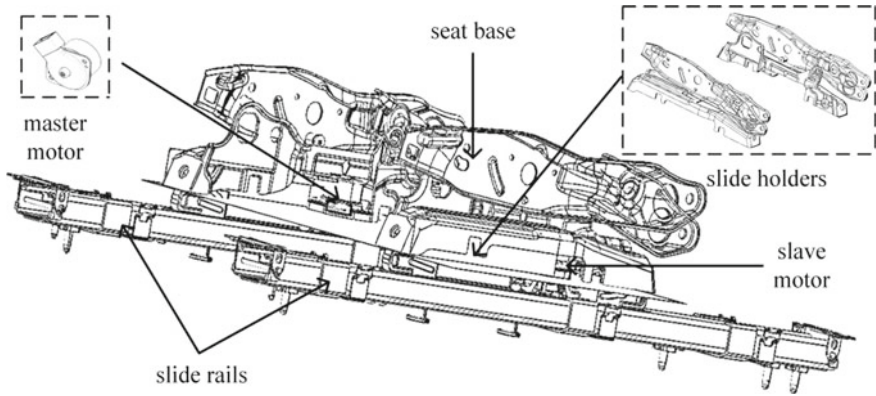


Fig. 4 Schematic diagram of the dual-motor synchronous experimental platform

The seat runs normally with no load until t is equal to 5 s, and a load disturbance of 500 N is applied to the side of the master motor at t is equal to 5 s. The experimental results are shown in Figs. 5, 6 and Table 1.

As can be seen from Fig. 5, except for the master-slave control, the proposed method and the other two commonly used synchronous control methods have the same actual speed of the two motors in the process of motion, indicating that the synchronization performance is better when there is no load disturbance.

As can be seen from Fig. 6, in parallel control, only the motor speed on the side where the load changes has drastic changes, while the other motors have no

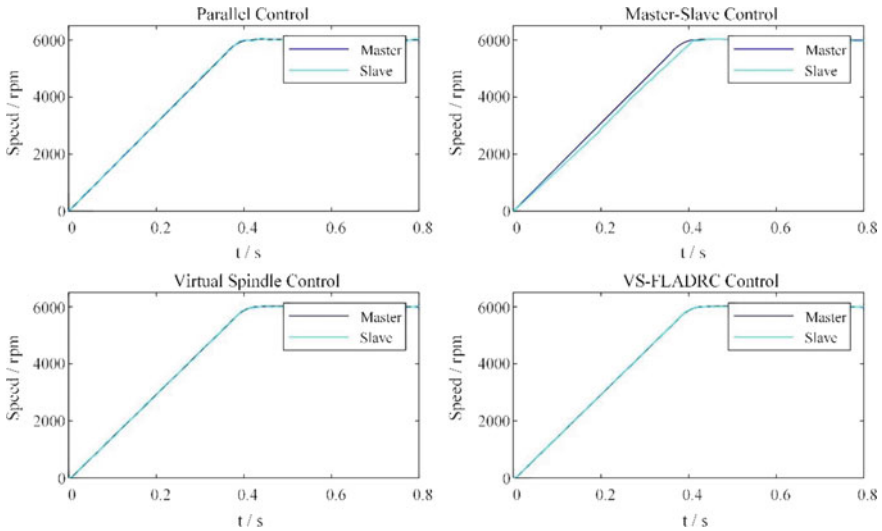


Fig. 5 Speed comparison of four control methods in motion without load disturbance

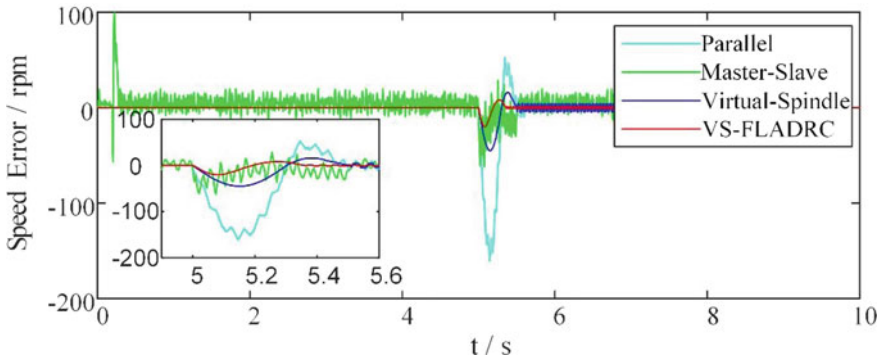


Fig. 6 Comparison of synchronization errors of four control methods when the master motor is subjected to load perturbation

Table 1 Synchronization performance when the master motor is disturbed by a 500 N load

Control method	Maximum speed error (rpm)	Speed response time (s)
Parallel	161	0.62
Master-slave	61	0.66
Virtual-shaft	44	0.49
VS-FLADRC	20	0.35

reaction, and the synchronization performance is poor. Master-slave control when the master motor is disturbed by the load, the slave motor will adjust according to the actual speed of the master motor, but the synchronization error will produce large fluctuations during the whole operation. In the traditional virtual spindle control, the motor on the other side can also make corresponding adjustment when the load is disturbed on either side. VS-FLADRC is significantly better than other commonly used synchronous control methods in terms of synchronous control effect when the load on either side of the motor changes abruptly during the seat motion, due to the inclusion of a fuzzy-LADRC speed compensator in the control, which effectively suppresses the speed error between motors, improves the speed response time of the system under load disturbance, the process of response adjustment is smoother and has better tracking performance. As can be seen from Table 1, VS-FLADRC has a response time less than or equal to 0.35 s under 500 N load disturbance, which are improved by 38.2% compared with traditional virtual spindle control. The speed synchronization error less than or equal to 20 rpm is reduced by 54.5% compared with traditional virtual spindle control, and the synchronization control performance is significantly improved.

4.2 Analysis of Experimental Results

Analysis of the experimental verification results shows that parallel control and master-slave control have poor synchronization performance in practical applications. Traditional virtual spindle control can ensure a certain degree of system synchronization when subjected to load disturbance, but the regulation time is long and the anti-interference capability is insufficient. VS-FLADRC on synchronous control performance is superior to other common synchronous control method, fast response time, high precision synchronous control, advanced control, speed response ability, and the coordinated control system significantly enhance load characteristic, stronger anti-interference ability, and response to adjust process more smooth and has good dynamic performance and steady-state performance. VS-FLADRC incorporates a fuzzy-LADRC on the basis of virtual spindle control, and when the cooperative control system is disturbed by the load, it can make a reasonable planning of the differential action according to the abrupt speed synchronization error and make an overcorrection to the system, so that the control system can respond quickly when the speed synchronization error is large. At the same time, the differential action gradually decreases when the control system becomes stable, reducing the influence of high-frequency noise interference on the stability of the system and achieving the maximum performance of the differential action in the control system.

5 Conclusions

Aiming at the problem of inconsistent speed response time under load disturbance, this paper integrates fuzzy-LADRC as the speed compensator of the control system on the basis of virtual spindle control, which improves the speed response ability and load characteristics of the dual-motor synchronous control system in the actual application process. The experimental results show that, compared with the traditional virtual spindle control, the response time of VS-FLADRC under 500 N load disturbance is improved by 38.2%, and the speed synchronization error is reduced by 54.5%, and the synchronization control accuracy is significantly improved. It can also effectively suppress the output fluctuation under noise interference and has strong anti-interference ability and robustness, which can ensure the stable operation of the control system.

References

1. Hung C-W, Lee RCL, Huang B-K, Yu S-T (2019) Multi-motor synchronous control with CANopen. *J Robot* 37–40
2. Lim C, Levi E, Jones M, Rahim NA, Hew W (2014) A comparative study of synchronous current control schemes based on FCS-MPC and PI-PWM for a two-motor three-phase drive. *IEEE Trans Ind Electron* 3867–3878
3. Koren Y (1980) Cross-coupled biaxial computer control for manufacturing systems. *J Dyn Syst Meas Control* 265–272
4. Anderson RG, Meyer AJ, Valenzuela MA (2001) Web machine coordinated motion control via electronic line-shafting. *IEEE Trans Ind Appl* 247–254
5. Liu G-H, Liu P-Y, Shen Y (2008) Neural network generalized inverse decoupling control of two-motor variable frequency speed-regulating system. In: *Proceedings of the CSEE*, pp 98–102
6. Abjadi NR, Soltani J, Askari J (2009) Nonlinear sliding-mode control of a multi-motor web-winding system without tension sensor. *IET Control Theory Appl* 419–427
7. Liu X-Q, Hu J-Q, Zhou L (2010) Active disturbance rejection control of three-motor synchronous control system. In: *Proceedings of the CSEE*, pp 80–85
8. Huang Z, Wan H, Jin F, Gu M (2020) Research on adaptive synchronous control of dual motors based on neurons. In: *2020 IEEE international conference on advances in electrical engineering and computer applications*, pp 17–20
9. Francis M, Hoach TN, Han HC, Jin WJ (2017) Finite set model predictive control of interior PM synchronous motor drives with an external disturbance rejection technique. *IEEE Trans Mechatron* 762–773
10. Yeam T-I, Lee D-C (2021) Design of sliding-mode speed controller with active damping control for single-inverter dual-PMSM drive systems. *IEEE Trans Power Electron* 5794–5801
11. Yehao X, Feng P (2020) Application of virtual axis synchronization control strategy in tension control. *Manuf Autom* 58–62
12. Dan W, Tong Z, Ken C (2013) Research and industrial applications of active disturbance rejection control to fast tool servos. *Control Theory Appl* 1534–1542
13. Lili D (2013) Application of active disturbance rejection control to micro-electro-mechanism system transducers. *Control Theory Appl* 1543–1552

MDagg: A New Aggregation Method Using Mahalanobis Distance



Songi Gwak  and Souhwan Jung 

Abstract Federated Learning has been praised for several years for collaborating machine learning paradigm while providing the data privacy of users who participate in the FL network. Even though there are many advantages of adopting FL architecture on their system, this is still suffered from security concerns because of various components. Mainly, a poisoning attack can be applied in the FL network to compromise the gradient update of the global model. In this case, a parameter server might lose the robustness of the global model. Several gradient aggregation rules have been proposed to prevent Byzantine from attacking the global model by adding malicious gradients. However, existing aggregation solutions such as Krum, Multi-krum, and Geometric Median cannot offer a high detection rate or effectively filter outliers. To handle untargeted poisoning attacks by Byzantine on the FL network, we propose a new aggregation rule MDagg which uses Mahalanobis Distance to consider the distance between gradients and covariance among gradients simultaneously. Whenever the server receives gradients from clients, the server computes Mahalanobis Distance between all gradients by calculating the covariance matrix and multiplying the difference between gradients' vectors. Our evaluation results show that MDagg provides the robustness of the FL network, wherein Byzantine compromises some of the clients.

Keywords Distributed computing · Federated Learning · Untargeted poisoning attack · Byzantine attack · Byzantine resilience · Gradient aggregation rule

S. Gwak

School of Information and Communication Engineering, Soongsil University, Seoul, South Korea
e-mail: song2@soongsil.ac.kr

S. Jung (✉)

School of Electronic Engineering, Soongsil University, Seoul, South Korea
e-mail: souhwanj@ssu.ac.kr

1 Introduction

Federated Learning (FL) is a machine learning setting in which many workers collaboratively train a global model from a server. This paradigm has come out for security purposes of the risk of sharing private data with the central server. If the server uses FL network, this does not need to keep a bunch of data from many participants. The server then aggregates and averages gradients from distributed clients. The server and clients do not share raw data directly. It seems safer than a previous machine learning framework. However, some researchers have tried to attack the FL network. Among these attacks, there is an attack called the poisoning attack. Especially in the case of the untargeted attack, Byzantine does not have a specific goal of the attack. Byzantine workers try to manipulate a gradient that they must send to the server in a learning process to compromise a global model.

Some Byzantine resilience algorithms, such as Krum, Multi-krum, and Geometric Median, emerged to prevent these poisoning attacks. However, they have solid assumptions and limitations in their research. Because of that, [1, 2] suggest powerful poisoning attacks which compromise the global model even though a Byzantine resilience algorithm is set up in the network. We study existing Gradients Aggregation Rules (GAR) to point out the problem of current solutions. Then, we design and implement a new aggregation method that solves problems of existing algorithms using Mahalanobis Distance (MD), widely used for anomaly detection. Finally, we experiment with our algorithm and others on the FL network wherein Byzantine tries to compromise the accuracy of the global model. By comparing the experiment results, we assert that our algorithm can provide the robustness of the FL network and reduce the computation overhead compared to Multi-krum, which is the most accurate algorithm among the existing algorithms we handle.

2 Backgrounds

2.1 Federated Learning

Federated Learning is a machine learning approach that enables distributed devices to train a global model collaboratively without security concerns such as data privacy [3–5]. In this setting, participants joining the FL network download the current model and train the model locally. Then, each participant sends the updated model to the server. Now, the server aggregates gradients from participants and updates the server’s gradient using the aggregated result. In this scenario, there may be a Byzantine worker (or more) who wants to compromise the global model. A general FL setting with Byzantine is shown in Fig. 1.

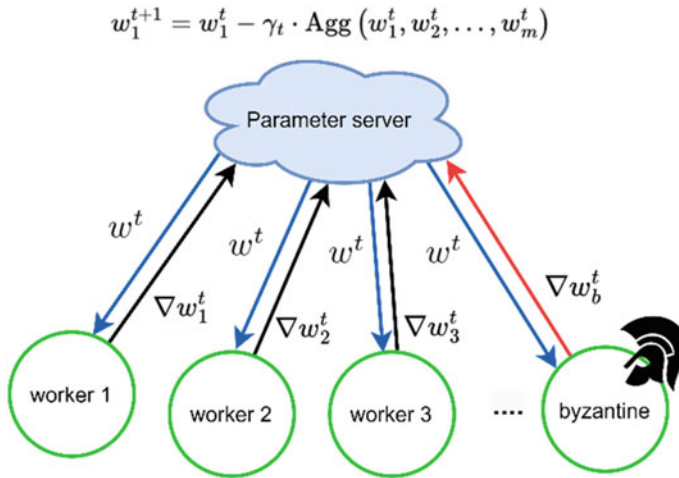


Fig. 1 Federated Learning setting with Byzantine worker

2.2 Poisoning Attack in FL

Even though this paradigm comes out for security purposes, there are still some security problems. Specifically, some attacks come out against the FL, categorized as Robustness and Privacy [6, 7]. For the system's robustness, there are two types of attacks: (1) untargeted poisoning attack and (2) targeted poisoning attack. An attacker tries to compromise the global model by sending a manipulated gradient in an untargeted attack. However, the attacker does not have a target. The goal is to make the global model lose its integrity. This attack is well-known as the Byzantine attack. In a targeted attack, an attacker has a target label and tries to induce the global model to return the target label.

2.3 Gradient Aggregation Rule

FedAvg. FedAvg is the most used aggregation algorithm [3]. This is called FedAvg, based on SGD optimization problem. A single server selects a group of clients to train the FL process to aggregate gradients from distributed devices. The server deploys a global model to the clients to make them compute their own local model and send it back to the server, then the server leverages averaging calculation to compute the weights' sum of all the received local model updates.

Krum and Multi-krum. Krum is an aggregation rule which satisfies Byzantine resilience [8]. In this method, for each local model w_i , Krum selects the gradient from $m-f-2$ gradients, which are the closest to w_i in the Euclidian norm space. This

algorithm is from an intuition that the Byzantine workers' gradients are far from the normal clients' to damage the global model. To leverage the Krum effectively, Multi-krum selects gradients using Krum, makes a selection set, and calculates the average of the gradients in a selection set. This outperforms Krum in the point of model accuracy.

Geometric Median. Geometric Median is an aggregation rule that uses the Euclidean norm for secure aggregation [9]. This can prevent the FL process from aggregating the mean of all gradients by allowing the server to choose a point with a Geometric Median.

2.4 Mahalanobis Distance

The Mahalanobis Distance is a way to measure the distance between a point P and a distribution D . This differs from the Euclidean distance, which is used in Krum method because the MD reflects the correlations among variables. The formulation of the MD is

$$\sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \quad (1)$$

In this formulation, the \vec{x} is a vector that we want to calculate the distance with a distribution and the μ is the mean of the distribution. The S is the matrix of covariance among variables. This distance is widely used in anomaly detection since this considers the distribution and the correlation between variables. As a result, this concludes, "How much is this point far from the distribution?"

3 Proposed Algorithm

3.1 Problem Definition

Existing aggregation rules are insufficient to prevent Byzantine from attacking the Federated Learning system. Krum and Multi-krum are similarity-based aggregation rules [10], which assume that malicious gradients must be far from benign. As a result, Krum selects the gradient from its input set. Geometric Median also selects the gradients, which are geometrically median, adopting the mean of gradients based on Euclidean norm space. All these aggregation rules are mentioned in Sect. 2.3. Filter some gradients doubted to be malicious by caring values of gradients. In this paper, we simply remove some malicious gradients based on intuition. The way to filter outliers will be a future work. Otherwise, they adopt the gradient, which is not likely to be malicious. According to the researchers [1, 2], if Byzantine carefully crafts the

gradient, Euclidean distance can be similar to the benign gradients even though the direction of gradients is opposed. The idea of this paper is to care about the values of gradients and the correlations among variables simultaneously.

Algorithm 1 Aggregation Method with Mahalanobis Distance (MDagg)

```

for  $t = 0, 1, 2, \dots, T-1$  do
  Parameter server: send  $w^t$  to all the working processes.
  for  $i \in [m]$  do
    in parallel
      Worker processes  $i$ : compute local gradient
       $V_i^t = \begin{cases} V_{P_i}^t & \text{from benign clients} \\ V_{B_i}^t & \text{byzantine} \end{cases}$  end
    end
  end
  end
  Covariance S:  $(\Sigma(V_i^t - V_i^t) / m)^{-1}$ 
  Mahalanobis Distance  $\sqrt{MD} : (V_i^t - V_i^t)^T S^{-1} (V_i^t - V_i^t)$ 

  Choose and remove some outliers =  $\begin{cases} \text{chi-square distribution} \\ \text{intuitively and empirically} \end{cases}$ 
  Aggregates gradients without outliers and updates  $w^{t+1}$ 

```

3.2 Algorithm

An algorithm of our MDagg is shown in Algorithm 1. Firstly, in each round t , the parameter server sends the global model w^t to all the working processes. When working devices receive the global model, they train their local model using their own dataset. After they trained, they send back the local model V_i^t to the server. In this procedure, Byzantine can send their malicious gradients to the server. To prevent their attack, if the server receives all local models from devices, it calculates covariance and Mahalanobis Distance by using this covariance. Once the server handles outliers, which are susceptible models, it aggregates benign local models and updates the global model. In this paper, we choose and remove outliers empirically for the sake of the simpleness of the experiment.

4 Experiment

4.1 Attack on Aggregation Rules

To prove the ability of MDagg, we perform the omniscient attack [8, 10], which multiplies the update vector with a negative constant. This reverses the gradient descent's direction and reduces the model's performance. We test this attack on existing aggregation rules and MDagg. For experiments, we set up an FL network with ten clients and three Byzantine workers to test the omniscient attack. Also,

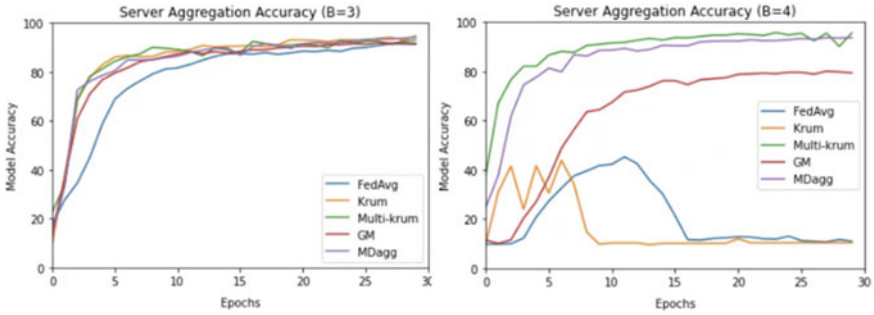


Fig. 2 Server aggregation accuracy

we use MNIST dataset for classification tasks and a slightly changed LeNet for computation performance. The result of the server aggregation is shown in Fig. 2. Then, we change the number of Byzantine workers to four. The result of the server aggregation is also shown in Fig. 2.

4.2 Robustness of the Algorithm

According to the results of Fig. 2, when the number of Byzantine workers is small ($B = 3$), all aggregation rules work correctly, including FedAvg. However, when the number of Byzantine workers is large enough ($B = 4$) to compromise the global model’s performance, most aggregation rules no longer work well. In the case of FedAvg, this averages all gradients. Thus, if there are many compromised devices, FedAvg will also calculate the gradients from these devices. Aside from that, Krum bounds the guaranteed number of Byzantine since this chooses a benign gradient among gradients. Geometric Median is also similar in this case. Thus, we check that large numbers of Byzantine can compromise the FL network, although there is a particular aggregation rule.

In the case of our algorithm, we provide robustness of the system even if there are many Byzantines. However, the model performance of server aggregation between MDagg and Multi-krum is similar since Multi-krum averages all chosen gradients from Krum algorithm. Because it takes one more step than Krum, it has more computation overhead. When we compare the computation overhead between MDagg and Multi-krum, our algorithm is slightly faster than Multi-krum, averaging 0.02 s.

5 Conclusion

In this paper, we present MDagg, a new aggregation rule to consider the distance between gradients and covariance at the same time. We first study existing aggregation rules to analyze the weaknesses of original states. Then, we provide an algorithm to prevent untargeted poisoning attacks on the FL network. Also, we experiment with a poisoning attack on existing GARs to compare the model performance and the computation overhead among GARs.

As a result, our algorithm provides robustness of the system even though there are many Byzantines in the network. Also, MDagg is slightly faster than Multikrum, which is the most accurate algorithm among GARs we researched. Using chi-distribution to handle outliers is also interesting research for future work.

Acknowledgements This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2022-2020-0-01602) supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP).

References

1. Fang M, Cao X, Jia J, Gong NZ (2020) Local model poisoning attacks to Byzantine-robust federated learning, p 18
2. Shejwalkar V, Houmansadr A (2021) Manipulating the byzantine: optimizing model poisoning attacks and defenses for federated learning. In: Network and distributed system security symposium, Virtual
3. McMahan B, Moore E, Ramage D, Hampson S, Arcas BAY (2017) Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th international conference on artificial intelligence and statistics, pp 1273–1282
4. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D (2017) Federated learning: strategies for improving communication efficiency. arXiv
5. Bonawitz K et al (2019) Towards federated learning at scale: system design. System Design, p 15
6. Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantaha A, Srivastava G (2021) A survey on security and privacy of federated learning. Future Gener Comput Syst 115:619–640
7. Lyu L et al (2022) Privacy and robustness in federated learning: attacks and defenses. arXiv
8. Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J (2017) Machine learning with adversaries: Byzantine tolerant gradient descent. In: Advances in neural information processing systems, vol 30
9. Pillutla K, Kakade SM, Harchaoui Z (2022) Robust aggregation for federated learning. arXiv
10. Wan CP, Chen Q (2021) Robust federated learning with attack-adaptive aggregation. arXiv

Graphormer-NAS-X: A Novel Graphormer-Based Neural Architecture Search Method



Hu Pan , Yuan Hu , Ying Yang , and Xu Zhang 

Abstract Neural architecture search (NAS) aims to search for the best neural network model for various tasks. In this paper, an improved Graphormer is used as an encoder to learn the embedding of the neural architecture, called Graphormer-NAS-X, where X is related to the specific search algorithm. Specifically, the centrality encoding is improved by adding an eigenvector centrality that can capture the importance of adjacent nodes. And the position encoding is enhanced by using the shortest distance of the directed graph, which can better reflect neural architecture's single-directional flow of computational information. The experiments are evaluated on two public datasets with two commonly used search algorithms. Graphormer-NAS-X outperforms most existing NAS algorithms, which indicates Graphormer benefits the downstream search.

Keywords NAS · Graphormer · Neural architecture encoder

1 Introduction

Neural architecture search (NAS) is a technology for automatically designing neural network structures, attracting researchers' extensive attention recently. Research shows that the encoding method of neural architecture may significantly change the search result. Recently, based on Transformer, Graphormer [1] introduces a centrality encoding to capture the node importance in the graph and a position encoding to capture the related information among nodes. However, Graphormer is mainly designed for undirected graphs, and cannot be directly applied to learn

H. Pan · Y. Yang · X. Zhang (✉)

Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

e-mail: zhangx@cqupt.edu.cn

Y. Hu

Chongqing Planning and Natural Resources Information Center, Chongqing 401147, China

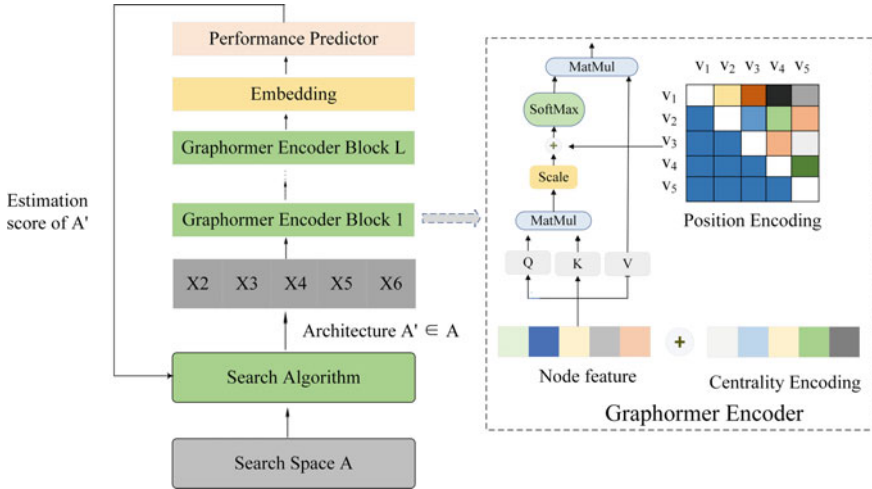


Fig. 1 Pipeline of Graphormer-NAS-X

the embedding of neural architecture, because the neural architecture is a directed acyclic graph.

In this paper, an improved Graphormer is used as an encoder to learn the embedding of the neural architecture, called Graphormer-NAS-X, where X is related to the specific search algorithm. Specifically, the centrality encoding is improved by adding an eigenvector centrality that can capture the importance of adjacent nodes. And the position encoding is enhanced by using the shortest distance of the directed graph because it can better reflect neural architecture’s single-directional flow of computational information. Figure 1 shows the pipeline of Graphormer-NAS-X in detail. The experimental results show that Graphormer can fully extract the spatial information of neural architecture and facilitates the downstream architecture search.

Our contributions are summarized as follows:

1. The eigenvector centrality is added to centrality encoding, and the shortest distance of the directed graph is used as the position encoding, which improves the representation ability of neural architecture information.
2. We proposed Graphormer-NAS-X which uses the improved Graphormer as a neural architecture encoder to learn the embedding of the neural architecture in NAS methods, which improves the downstream search efficiency of NAS tasks.

2 Related Work

Most early NAS methods directly use a discrete one-hot adjacency matrix to encode the structure of neural architectures. However, the size of the adjacency matrix grows

quadratically as the search space grows up. Recent NAS methods widely use dedicated neural networks to learn continuous representations of neural architectures, such as LSTM [2], MLP [3], GCN [4, 5], and Transformer [6].

Graphormer [1] uses several effective structural encoding methods to help Transformer better encode graph information. However, Graphormer is mainly designed for undirected graphs and can't be directly used for neural architecture encoding.

3 Methods

The improvements in Graphormer and the detailed implementation of Graphormer-NAS-X are described in this section.

3.1 Graphormer

Centrality Encoding. Different nodes of a graph may have different importance.

Graphormer uses degree centrality to represent the importance of the node. However, the importance of a node depends not only on the number of neighboring nodes but also on the importance of neighboring nodes. So, the eigenvector centrality is added to the degree centrality encoding of Graphormer, which can capture the importance of adjacent nodes. Now, the calculation formula of centrality encoding is

$$h_i^{(0)} = x_i + z_{\deg^-(v_i)}^- + z_{\deg^+(v_i)}^+ + z_{e_i} \quad (1)$$

where $z^-, z^+ \in R^d$ are learnable embedding vectors based on indegree $\deg^-(v_i)$ and outdegree $\deg^+(v_i)$, respectively. The z_{e_i} is a learnable embedded vector based on eigenvector centrality e_i . The x_i is the feature of node i . The calculation formula of the eigenvector centrality is

$$e_j = c \sum_{j=1}^n a_{ij} e_j \quad (2)$$

where c is a proportional constant. If node i is connected to node j , a_{ij} is 1, otherwise, it is 0.

Position Encoding. In the Transformer model, it is very important to explicitly specify the location relationship of different nodes or positional dependencies. The attention mask of positional dependency of original Graphormer is designed for undirected graphs and cannot be directly used for neural architecture representation learning.

Therefore, an improved position encoding method is proposed. Specifically, in the neural architecture, if there is a reachable path from node i to j , the shortest distance between i and j is encoded as the positional dependency of node i on j . Otherwise, set the location dependency to a very large number, indicating little mutual influence between them. Thus, the spatial position encoding calculation formula is

$$\sigma(v_i, v_j) = \begin{cases} \tilde{A}_{ij}, & \tilde{A}_{ij} \neq 0 \\ +\infty, & \tilde{A}_{ij} = 0 \end{cases} \quad (3)$$

where \tilde{A}_{ij} is the shortest distance matrix calculated according to the adjacency matrix. This position encoding is regarded as a biased term in the self-attention module.

3.2 Graphormer-NAS-X

Based on the improved Graphormer, the Graphormer-NAS-X is proposed by us, where X represents a specific search algorithm. As it is shown in Fig. 1, a performance predictor-based Graphormer is used to estimate the performance of the candidate neural architecture and then feed the evaluated results back to the search strategy to guide the next search. Evolutionary algorithm (EA) and reinforcement learning (RL) are used by us as two representative search algorithms to evaluate Graphormer-NAS-X, and these two methods are called Graphormer-NAS-EA and Graphormer-NAS-RL, respectively.

Graphormer-NAS-EA. The pipeline of EA follows NPENAS [7]. Firstly, the architecture pool is initialized with randomly sampled architectures from the search space, then train and validate these neural architectures. A predefined number of candidate neural architectures are generated using the one-to-many mutation strategy. Then, the neural predictor is used to predict the performance of candidate neural architectures. The top K architectures with the best performance are selected as descendants. Finally, those descendants are added to the architecture pool.

Graphormer-NAS-RL. The RL-based search strategy follows arch2vec [8]. Firstly, an encoder based on the Graphormer is used to compute the embedding of the neural architecture. Then, the embedding is transmitted to the controller-based LSTM to sample the action and get the next state based on the nearest neighbor search using L2 distance to maximize the accuracy as a reward.

4 Experiments

The proposed method is evaluated on two public datasets, NASBench-101 [9] and NASBench-201 [10]. The Graphormer is composed of 12 Transformer layers. The

number of attention heads in the self-attention module is 32. The hidden state size is 64. The hidden dimension is 256 for all feed-forward layers. A dropout layer is used after the feed-forward layer, and the dropout rate is 0.1. AdamW is used as the optimizer. The initial learning rate is 1×10^{-4} . The momentum parameters are set to 0.9 and 0.999. All experiments were run on NVIDIA V100 GPUS and took around 11 GB GPU memory for NASBench-101 and 6 GB GPU memory for NASBench-201.

As it is show in Fig. 2, the proposed method is compared with five architecture encoding methods using the predictor performance, including LSTM [2], MLP [3], GIN [8], GCN [4], and Transformer [6], and the parameters of these encoders follow the original paper. The Kendall’s Tau(KTau) coefficient was employed as the evaluation criterion to assess the correlation between the predicted values of the model and the actual values of the samples.

The prediction results are exhibited in Fig. 2, in which the solid line and filling width represent the mean and the standard deviation in ten separate experiment runs, respectively. The neural predictor based on the Graphormer encoder is superior to others when the number of training samples is greater than 200.

We compare Graphormer-NAS-X with other predictor-based NAS methods. It should be noted that although NAS research is highly abundant, and it is still difficult to compare fairly and reproduce NAS method. This is because no shared experimental protocol was followed by all. To make the comparison as fair as possible, we try to compare our method with other NAS methods under the same search budget. Tables 1 and 2 show the experimental results of Graphormer-NAS-X and other predictor-based NAS algorithms on NASBench-101 and NASBench-201. The experimental results for Graphormer-NAS-X are presented in bold. It is obvious that the final search

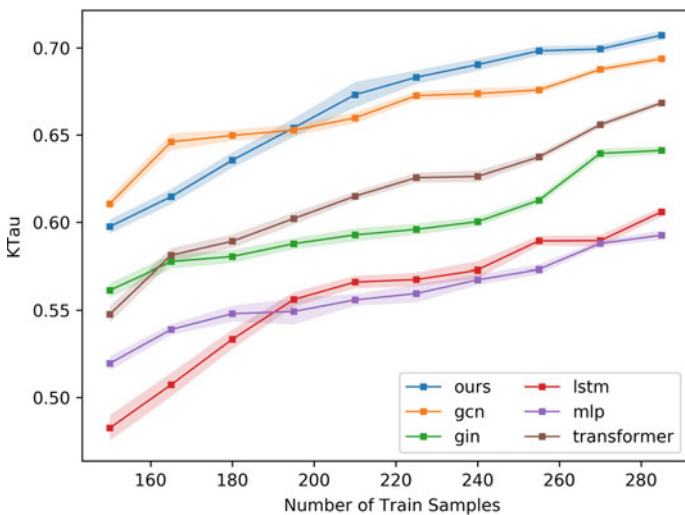


Fig. 2 Kendall coefficient between the model’s predicted and actual performance

Table 1 Search performance of all NAS methods on NASBench-101

NAS methods	Test accuracy [%]	Queries
NPENAS-BO [7]	94.10	150
CATE-DNGO-LS [6]	94.12	150
BONAS [4]	94.09	200
Graphormer-NAS-EA	94.14	150
Graphormer-NAS-RL	94.05	150
arch2vec-RL [8]	94.10	423
CTNAS [5]	93.92	423
Graphormer-NAS-EA	94.15	423
Graphormer-NAS-RL	94.10	423

Table 2 Search performance of all NAS methods on NASBench-201

NAS methods	Test accuracy [%]	Queries
arch2vec-BO [8]	94.18	1.2×10^4 s
NPENAS-NP [7]	94.27	100
BANANAS [3]	94.18	100
Graphormer-NAS-EA	94.24	100
Graphormer-NAS-RL	94.28	1.2×10^4 s

results improved about 0.1%, which indicates that Graphormer-NAS-X outperforms most other algorithms under a similar search budget.

5 Conclusions

In this paper, we propose a novel method named Graphormer-NAS-X, which uses an improved Graphormer as an encoder to learn the embedding of the neural architecture. The experimental results show that Graphormer-NAS-X outperforms most existing NAS algorithms and can search for better neural architecture.

Acknowledgements This research is sponsored by Natural Science Foundation of Chongqing (cstc2018jscx-mszdX0116), China.

References

1. Ying C (2021) Do transformers really perform badly for graph representation? In: Advances in neural information processing systems, p 34
2. Luo R (2018) Neural architecture optimization. In: Proceedings of the 32nd international conference on neural information processing systems, pp 7827–7838

3. White C (2021) BANANAS: Bayesian optimization with neural architectures for neural architecture search. In: Proceedings of the AAAI conference on artificial intelligence, pp 10293–10301
4. Shi H (2020) Bridging the gap between sample-based and one-shot neural architecture search with BONAS, p 33
5. Chen Y (2021) Contrastive neural architecture search with neural architecture comparators. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9502–9511
6. Yan S (2021) CATE: computation-aware neural architecture encoding with transformers, pp 11670–11681
7. Wei C (2020) NPENAS: neural predictor guided evolution for neural architecture search
8. Yan S (2020) Does unsupervised architecture representation learning help neural architecture search? In: Advances in neural information processing systems, vol 33, pp 12486–12498
9. Ying C (2019) NAS-Bench-101: towards reproducible neural architecture search. In: International conference on machine learning. PMLR, pp 7105–7114
10. Dong X (2019) NAS-Bench-201: extending the scope of reproducible neural architecture search. In: International conference on learning representations

Secondary Salient Feature-Based GNN for Few-Shot Classification



Chuang Liu , Wei Hou , Xinyue Wang , and Xu Zhang 

Abstract Few-shot learning aims to use limited labeled samples to achieve effective classification results. To mine the features of images in a limited number of pieces, some researchers proposed to drill salient features to improve the classification effect. However, they ignore the use of salient secondary features. Therefore, we offer to use salient secondary features to supplement the deficiency of salient features. Combining with the foreground extraction network and the graph neural network, a better classification effect is obtained in the experiment.

Keywords Few-shot learning · Salient features · Foreground extraction network · Graph neural network

1 Introduction

With the continuous development of deep learning research, the characteristics of less data accumulation and the high annotation cost of new application scenarios make traditional deep learning more and more powerless. Therefore, people urgently want to find a method that can play a good role in application scenarios with limited samples. Few-shot learning has become an important research direction.

In 2003, the one-shot learning problem [1] was formally raised. They believe that only using one or a few labeled samples in a new class can predict unlabeled samples by relying on the base class with a large number of labels. With the fire of deep learning after 2012, few-shot learning also began to try to explore the field of deep learning. During this period, many famous basic algorithms have been proposed, such as ProtoNet [2] and MAML [3], in few-shot learning. With the popularity of

C. Liu · X. Wang · X. Zhang (✉)

Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: zhangx@cqupt.edu.cn

W. Hou

Chongqing Planning and Natural Resources Information Center, Chongqing 401147, China

graph neural networks in previous years, Satorras and Estrach [4] introduced a graph model to the few-shot problem and used the graph model to increase the classification effect. It took the embedded vector of the image and the label as the nodes of the graph network and constantly updated the node characteristics through the adjacency matrix of the graph model so that the predicted image is closer to the image of the class in the graph. Kim et al. [5] believes that the edges of graph nodes also contain image-related connections, thus proposing the edge-labeled neural network. Gidaris and Komodakis [6] added noise reduction self-coding to the graph neural network. They modified the weights of new classes through noise reduction self-coding and graph neural networks to achieve the effect of optimizing the model. Zhang et al. [7] proposed a significant local image generation network based on the attention mechanism to locate and generate substantial local features.

We found that the above papers ignored the influence of salient secondary features. For example, the most prominent feature of the male peacock is the big bright tail, but the female does not. When you take the male bird as the labeled sample and the female bird as the unlabeled sample, the significant features generated will be very different, and the classification accuracy will certainly be greatly reduced. Therefore, we consider the salient secondary features to achieve better results. We propose a local salient and secondary local salient image generation network to generate local salient and secondary local salient features and classify them according to the Multi-granularity graph neural network. We obtain better results on the standard data set with few-shot learning.

2 Methodology

2.1 Problem Definition

In few-shot learning, training and testing are divided into C-way K-shot problems. The C-way K-shot problem is that in each training or test task, C classes are selected from the training or test set, and K images are selected from each class as the support set. At the same time, M and K images are selected from the C classes as the query set.

2.2 Overall Framework

As shown in Fig. 1, the overall framework is improved according to MRA-GNN [7].

The Pre-trained Extraction Foreground Network. Some saliency detection based on deep learning has achieved good results. Therefore, this paper hopes to obtain the image foreground related to people's attention through relatively mature saliency detection technology in few-shot classification tasks. The algorithm selected is the

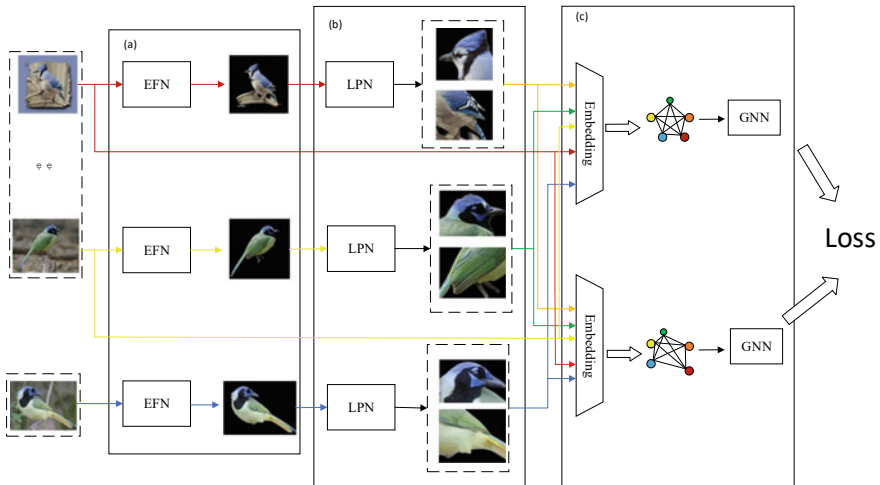


Fig. 1 The pipeline of framework **a** the pre-trained extraction foreground network, **b** local proposal network, and **c** graph neural network

unsupervised public algorithm BasNet [8], which mainly extracts foreground images in two steps: (1) generating the saliency map $S(I)$ by using the saliency network and (2) generating the foreground image F by combining the original image I .

$$F = S(I) \odot I \quad (1)$$

\odot is the Hadamard product.

Local Proposal Network. LPN adaptively locates salient local and secondary local salient images, and the ability to extract salient local features is continuously trained by embedding network and classification network.

After LPN extraction, we will get $[x_1, y_1, l_1, x_2, y_2, l_2]$, and we will get two local salient images through $[x, y, l]$ of these two groups, as follows:

Through a group of $[x, y, l]$, we can get the coordinates of its border,

$$\begin{aligned} x_1 &= x - l, \quad x_r = x + l \\ y_1 &= y - l, \quad y_r = y + l \end{aligned} \quad (2)$$

Then we used mask $M(\cdot)$ to crop the image.

$$\begin{aligned} M(\cdot) &= [h(x - x_1) - h(x - x_r)] \\ &\quad * [h(y - y_1) - h(y - y_r)] \end{aligned} \quad (3)$$

$$h(x) = \frac{1}{1 + e^{-kx}} \quad (4)$$

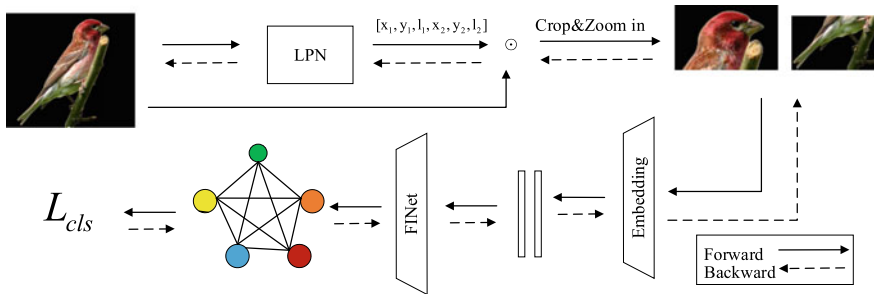


Fig. 2 The pipeline of extract salient features

$$x^{\text{crop}} = x^0 \odot M(\cdot) \quad (5)$$

The mask $M(\cdot)$ ensures that the value outside the cropped border becomes infinitely small while the value inside the border remains unchanged. Then, bilinear interpolation enlarges the cropped image to obtain a salient local image.

After obtaining two locally salient images, they are put into the embedded network to generate embedded features. Then, FINet uses a layer of a neural network to fuse the elements of the two images to output a more characteristic feature, which is put into the classification network. The classification loss obtained is taken as the direction of model optimization, and then the network is trained. The overall architecture is shown in Fig. 2.

Graph Neural Network. Here we use GNN [4] as the classification network and extend GNN to multiple levels to increase the classification effect.

2.3 Loss Function

In this experiment, with the help of an interlude training mechanism, the model parameters are trained by setting classification loss and coarse and fine-scale ranking loss in each epoch.

$$L(x) = L_{\text{cls}}(Y, \tilde{Y}) + \sum_{s=0}^l (L_{\text{rank}}(P_y^{(s)}, P_y^{(s+l)})) \quad (6)$$

\tilde{Y} represents the label finally obtained through the framework and Y represents the real label of the sample. L_{cls} is to update the parameters in the graph neural network and L_{rank} indicates that a more reliable prediction will be generated based on the coarse and fine-scale network. The classification ability of the model can be further improved by minimizing $L(x)$.

L_{cls} is represented as follows:

$$L_{\text{cls}}(Y, \tilde{Y}) = \sum_{i=1}^{C \times K + M} \sum_{j=1}^C -\delta(y_i == j) \log(p(\tilde{y}_i = j | x_i)) \quad (7)$$

$$P(\tilde{y} = j | x_i) = \frac{\exp(y_{ij}^*)}{\sum_{k=1}^C \exp(y_{ik}^*)} \quad (8)$$

$\delta(\cdot)$ is an indicator function, $\delta(\cdot) = 1$ if a is true and 0 otherwise. y_{ij}^* represents the j th component of the predicted label y^* in the label propagation.

L_{rank} denotes pairwise ranking loss, and $P^{(s)}$ represents the prediction probability in scale s .

$$L_{\text{rank}}(P_y^{(s)}, P_y^{(s+l)}) = \max(0, P_y^{(s)} - P_y^{(s+l)} + \text{margin}) \quad (9)$$

In this way, it can combine the two loss functions L_{cls} and L_{rank} to get $L(x)$, and then continuously minimize $L(x)$ to train the whole network model.

3 Experiments

3.1 Datasets

miniImagnet It is a benchmark data set for few-shot learning and is derived from an ILSVRC-12 data set. All images were 84×84 pixel RGB color pictures, a total of 100 different categories, 600 images per category, and 60,000 images.

CUB-200-2011 Caltech proposed the benchmark image data set for fine-grained classification and recognition research in 2010. It is a fine-grained data set of birds with rich categories and has been widely used in the field of few-shot learning.

3.2 Performance

We adopted the episodic training procedure and conducted 5 way-5 shot and 5 way-1 shot experiments for both miniImagenet and CUB datasets, which are the standard few-shot learning settings. The results show in Table 1, and all results have average accuracies with 95% confidence intervals.

The results show that on both miniImagenet and CUB datasets, the 5 way-1 shot experiment has achieved better results. Still, the improvement of 5 way-5 shot investigation could be better, and there is a downward trend on miniImagenet. That is

Table 1 Few-shot classification accuracies on miniImageNet and CUB

Model	Embedding	CUB		miniImageNet	
		5 way 1 shot	5 way 5 shot	5 way 1 shot	5 way 5 shot
GNN (2018) [4]	Conv-4	60.26	75.95	52.33	65.22
AdarGCN (2021) [9]	Conv-4	–	78.04	–	71.48
CSS (2021) [10]	Conv-4	66.01	81.84	50.85	68.08
IEPT (2021) [11]	Conv-4	–	–	58.43	75.07
MRA-GNN (2022)	Conv-4	72.30	87.08	61.99	78.71
Ours	Conv-4	73.21	87.48	62.71	78.32

Bold indicates the best experimental result among the comparative experiments

because the images are more scarce for 5 way-1 shot, and the extraction of salient secondary features is more effective for improving the preparation rate. On the contrary, after the expansion of 5 way-5 shot images, the diversity of images is supplemented by the real labeled images. The complementary effect resulting in salient secondary features is not apparent.

4 Conclusion

We introduce salient secondary features to supplement the deficiency of only using salient features to express image features. The model’s classification will be better when the images are extremely rare. However, as the image number improves, the supplementary effect of salient secondary features will be significantly reduced, or even the classification accuracy will be reduced. For future work, there will be more research prospects for the lightweight design of the overall network and a better combination of the secondary local salient image and salient local image.

Acknowledgements This research is sponsored by Natural Science Foundation of Chongqing (cstc2018jscx-mszdX0116), China.

References

1. Fe-Fei L (2003) A Bayesian approach to unsupervised one-shot learning of object categories. In: Proceedings of ninth IEEE international conference on computer vision. IEEE, pp 1134–1141
2. Snell J, Swersky K, Zemel R (2017) Prototypical networks for few-shot learning. In: Advances in neural information processing systems, pp 4077–4087
3. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of international conference on machine learning, pp 1126–1135.
4. Satorras VG, Estrach JB (2018) Few-shot learning with graph neural networks

5. Kim J, Kim T, Kim S, Yoo D (2019) Edge-labeling graph neural network for few-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 11–20
6. Gidaris S, Komodakis N (2019) Generating classification weights with GNN denoising autoencoders for few-shot learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 21–30
7. Zhang X, Zhang Y, Zhang Z, Liu J (2022) Discriminative learning of imaginary data for few-shot classification. *Neurocomputing* 467:406–417
8. Qin X, Zhang Z, Huang C, Chao G, Dehghan M, Jagersand M (2019) BASNet: boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7479–7489
9. Zhang J, Zhang M, Lu Z, Tao X (2021) AdarGCN: adaptive aggregation GCN for few-shot learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. IEEE Press, Nashville, pp 3482–3491
10. Zhang M, Zhang J, Lu Z, Tao X, Ding M, Huang S (2021) IEPT: instance-level and episode-level pretext tasks for few-shot learning. In: International conference on learning representations
11. An Y, Xue H, Zhao X, Zhang L (2021) Conditional self-supervised learning for few-shot classification. *IJCAI* 2140–2146

Mixup Cluster Contrastive Learning Over Harder Samples for Unsupervised Person Re-identification



Xin He , Hongwen Zhou , Die Dai , and Xu Zhang 

Abstract Learning discriminative features is essential for unsupervised person re-identification (Re-ID). Recently, the unsupervised representation learning-based Re-ID method has gained great progress by conducting contrastive learning with memory dictionaries. However, most existing methods ignored the importance of positive and negative samples. In this work, we proposed a novel mixup cluster contrastive (MCC) learning framework for unsupervised person Re-ID that can fully use negative samples. Furthermore, we proposed to utilize the mixup hard sample strategy to mixture the query and the negative features for even harder samples. Finally, we define a mixup hard sample contrastive loss, which helps to learn discriminative representation features. Experimental evaluations demonstrate that the proposed method outperforms state-of-the-art methods in unsupervised person Re-ID on Market-1501 and DukeMTMC-reID.

Keywords Unsupervised representation learning · Unsupervised person re-identification · Contrastive learning

1 Introduction

Person re-identification (Re-ID) has been widely studied as a specific person retrieval problem across non-overlapping cameras [1], which uses computer vision technology to determine a query person-of-interest whether has appeared in another place at different moment of shooting by distinct cameras, or even the same camera at a different time instant [2].

X. He · D. Dai · X. Zhang (✉)

Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

e-mail: zhangx@cqupt.edu.cn

H. Zhou

Chongqing Planning and Natural Resources Information Center, Chongqing 401147, China

The extant unsupervised person Re-ID methods can be broadly divided into two categories, unsupervised domain adaptation Re-ID (UDA) methods and purely unsupervised learning Re-ID (USL) methods. In UDA setting, some methods consider eliminating or reducing the differences between domains to migrate the discriminative information from the source domain to the target domain [3, 4]. In USL Re-ID setting, the most popular methods exploit high-quality pseudo-labels to train models [5, 6]. Since no annotations are used, the USL is more challenging. Recently, some state-of-the-art methods utilize a memory dictionary to store all instance features, then perform clustering and other operations from these stored features [7, 8].

In this work, we concentrate on contrastive self-supervised visual representation learning, taking the advantage of the principle that a good feature representation model should map the representation close to the positives and away from the negatives. The contributions of this work are threefold:

- We propose a mixup cluster contrastive learning framework for unsupervised person Re-ID which can make full use of negative samples deeply.
- We introduce a novel mixture strategy of negatives thus constructing harder samples, which can explore more discriminative information.
- Extensive experiments on two famous Re-ID benchmarks demonstrate that the method outperforms state-of-the-art methods and improves the performance of unsupervised person Re-ID.

2 Related Work

Unsupervised Person Re-ID. Due to the lack of identity labels, it is difficult for unsupervised person Re-ID model to learn discriminative features. Therefore, some works have given more attention to pseudo-labels that do not require source domain data to supervise the neural network. BUC [5] uses a bottom-up clustering to generate pseudo tags. The bottom-up clustering can access all samples and determine the similarity of samples. HCT [6] combines hierarchical clustering and hard batch triplet loss, using PK sampling to generate new datasets for training in each iteration.

Mixup. Mixup [9] has shown to be an effective data augmentation technique for supervised learning in various domains and tasks. More recently, mixup has been successfully adapted to the field of semi-supervised learning, domain adaption, and self-supervised learning [8]. In this paper, we utilize the mixup strategy to mixture the negative instance features based on dynamic memory dictionaries, which can generate harder instance features and are conducted to help the training process.

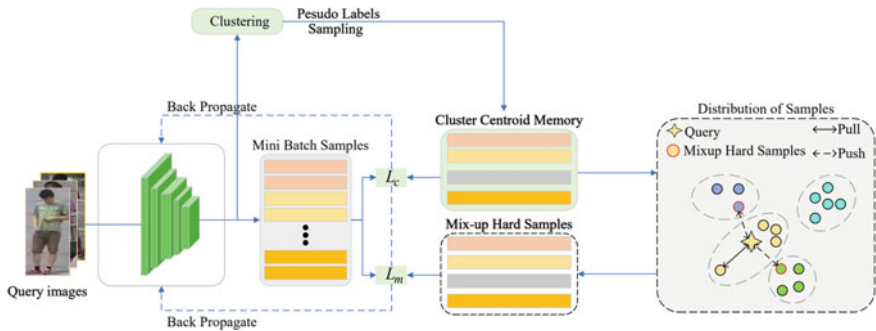


Fig. 1 Overview of our proposed framework for fully unsupervised person Re-ID

3 Proposed Method

3.1 Contrastive Learning Framework

To tackle the challenges in fully unsupervised person Re-ID, we propose a mixup contrastive learning framework (see Fig. 1), which consists of a CNN-based encoder, cluster centroid contrastive loss, and a novel mixup hard sample contrastive loss. Our training scheme alternates between three steps: (1) grouping the input images into clusters in instance memory and calculating the cluster centroid in cluster centroid memory, (2) updating the cluster centroid memory and the mixup hard sample memory, respectively, during every mini-batch, and (3) optimizing the encoder θ with cluster centroid contrastive loss and mixup hard sample contrastive loss.

3.2 Cluster Centroid Contrast

The key point of contrastive learning is to construct positive and negative pairs. Thus, maximizing the mutual information between positive sample pairs and minimizing the mutual information between negative sample pairs.

$$L_c = -\log \frac{\exp(q \cdot c_+ / \tau)}{\sum_{i=1}^N \exp(q \cdot c_i / \tau)}, \tag{1}$$

where q is a query feature and c_+ is the positive cluster centroid feature which belongs to the same identity with q . $(q \cdot c_i)$ indicates a negative sample pair. τ is a temperature hyper-parameter that controls the scale of similarities.

Calculating the cluster centroid of each cluster and store them in a memory bank $M_C[c_1, c_2, \dots, c_K]$ for the cluster centroid contrastive loss. In order to update all cluster centroid in consistent, we update the cluster memory bank as follows:

$$c_i \leftarrow \delta c_i + (1 - \delta)\bar{c}_i, \quad (2)$$

where \bar{c}_i is the average of i th class instance feature in the mini-batch. δ is a momentum coefficient that controls the update speed of the memory bank.

3.3 Mixup for Harder Samples

As mentioned in Sects. 2 and 3, we know that it is necessary to construct positive and negative sample pairs. Obviously, the role of negative samples has been ignored. In order to obtain more hard samples, the current solution is to either increase the batch-size or maintain a large memory bank. However, blindly increasing memory demand will also lead to diminishing performance gains. We construct even harder samples through mixup based on a memory dictionary with our mixup hard sample strategy.

Given a query q , $H = \{c_1, c_2, \dots, c_k\}$ is a queue features composed by negatives of q from cluster centroid memory. $\bar{H} = \{h_1, h_2, \dots, h_k\}$ be the set of mixup hard negative samples to be generated as follows:

$$h_k = \frac{\tilde{h}_k}{\|\tilde{h}_k\|_2}, \quad (3)$$

$$\tilde{h}_k = \alpha c_i + (1 - \alpha)c_j,$$

where $\|\cdot\|_2$ is the ℓ_2 -norm, $\alpha \in (0, 1)$ is a randomly chosen mixing coefficient. c_i and c_j are selected randomly chosen negative features from H .

We further propose to mix the query q with the hardest negatives to get even harder negatives. Then, similar to Eq. (3), we mixup query q and $h_k \in \bar{H}$ according to $h'_k = \frac{\tilde{h}'_k}{\|\tilde{h}'_k\|_2}$, where $\tilde{h}'_k = \beta q + (1 - \beta)h_k$, and h_k is randomly chosen negative feature from \bar{H} , while $\beta \in (0, 0.5)$ is randomly chosen mixing coefficient for the query. Note that $\beta < 0.5$ guarantees that the query's contribution is always smaller than the one of the negative.

We construct a memory bank M_{mix} to store the i th mixup hard negative instance features. We use the mixup hard sample strategy as mentioned above to mixture the query feature and negative samples for the hardest negative samples.

The mixup hard sample contrastive loss strives to distinguish query as positive samples by performing a log loss of k -way softmax-based classifier through K sample pairs constructed by us. We define the mixup hard contrastive loss as follows:

$$L_m = -\log \frac{\exp(q \cdot m^+ / \tau_{\text{mix_hard}})}{\sum_{i=0}^K \exp(q \cdot m^i_{\text{mix_hard}} / \tau_{\text{mix_hard}})}, \quad (4)$$

where $\tau_{\text{mix_hard}}$ is a temperature hyper-parameter, m^+ is the positive feature that belongs to the same identity with q , and $m_{\text{mix_hard}}^i$ is the highest cosine similarity of the mixup hard negative feature that belongs to the different identity with q .

3.4 Overall Contrastive Loss and Optimization

The proposed approach iteratively alternates between the clustering step, cluster centroid contrastive loss, and mixup hard negative contrastive loss. The entire loss for model learning is

$$L = \mu L_c + (1 - \mu) L_m, \quad (5)$$

where μ is a parameter to balance two terms and we set $\mu = 0.5$ by default.

4 Experiments

4.1 Results

We compare our method with state-of-the-arts UDA Re-ID (including MMCL [3] and ABMT [4]). Our method is even better than the unsupervised domain adaption methods using the source domain dataset, which shows the superiority of our method (Table 1). Our approach obtains 7.5% mAP and 0.7% rank-1 gain on the Market-1501 [11] dataset, 6.8% mAP and 2.1% rank-1 in the challenging DukeMTMC-reID [12] dataset.

Under the fully unsupervised setting, we compare our methods to state-of-the-arts USL Re-ID method. Our approach achieves 83.5% in mAP and 93.2% in rank-1

Table 1 Experimental evaluations on Market1501 and DukeMTMC-reID

Method	Market-1501				DukeMTMC-reID			
	mAP	R1	R5	R10	mAP	R1	R5	R10
MMCL [3]	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
ABMT [4]	78.3	92.5	–	–	69.1	82.0	–	–
BUC [5]	29.6	61.9	73.5	73.5	22.1	40.4	52.5	52.5
HCT [6]	56.4	80.0	91.6	95.2	50.7	69.6	83.4	87.4
SpCL (agnostic) [7]	73.1	88.1	95.1	97.0	65.3	81.2	90.3	92.2
CAP [8]	79.2	91.4	96.3	97.7	67.3	81.1	89.3	91.8
MGCE-HCL [10]	79.6	92.1	–	–	67.5	82.5	–	–
Ours	83.5	93.2	97.2	98.1	71.9	84.1	91.7	94.5

accuracy on Market-1501 [11]. We proposed method obtained 4.4% mAP and 1.6% rank-1 gain on the challenging DukeMTMC-reID [12] dataset.

4.2 Ablation Study

Influence of Hyper-Parameter μ . μ is a balancing factor between 0 and 1. When $\mu = 0$, the loss function contains only the mixup hard sample contrastive loss. Only using the mixed hard feature would cause the model cannot learn the real features, resulting in the quality of pseudo-labels worse and worse. In contrast, when μ is equal to 1 and only cluster centroid loss is involved, which cannot make the network to learn more discriminative features of intra-class. And when μ is equivalent to 0.5, we get the best performance 83.5% in mAP and 93.2% in rank-1, indicating that the proposed method can effective learn discriminative features over others methods (Fig. 2; Table 2).

Fig. 2 Different of parameter μ settings on Market1501 dataset

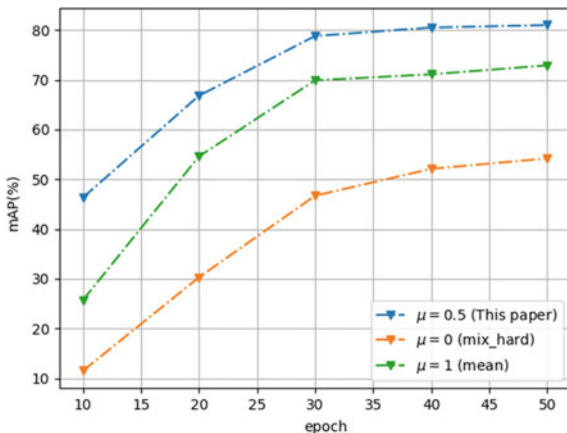


Table 2 Evaluation of parameter μ on Market1501

μ	Market1501			
	mAP	R1	R5	R10
0 (mix_hard)	58.8	79.7	90.8	93.6
0.25	81.0	92.1	96.7	98.0
0.5	83.5	93.2	97.2	98.1
0.75	83.9	92.7	97.0	97.8
1 (mean)	72.9	86.8	93.7	95.8

μ is equivalent to 0.5, we get the best performance 83.5% in mAP and 93.2% in rank-1, indicating that the proposed method can effective learn discriminative features over others methods

5 Conclusion

In this paper, we present a novel mixup cluster contrastive learning framework for fully unsupervised person Re-ID. We proposed to explore affinities between the query and the negative samples and utilize a mixup strategy to mixture the query and negative samples for harder samples based on memory dictionaries. We introduce a mixup hard sample contrastive loss to pull the positives closer and push away the negatives in the feature embedding space, which can learn discriminative representation features. The proposed method achieved better performance as it is shown in the experimental evaluation.

Acknowledgements This research is sponsored by Natural Science Foundation of Chongqing (cstc2018jscx-mszdX0116), China.

References

1. Chen YC, Zhu X, Zheng WS, Lai JH (2018) Person re-identification by camera correlation aware feature augmentation. *IEEE Trans Pattern Anal Mach Intell* 40(2):392–408
2. Ye M, Shen J, Lin G, Xiang T, Shao L, Hoi SC (2021) Deep learning for person re-identification: a survey and outlook. *IEEE Trans Pattern Anal Mach Intell* 44(6):2872–2893
3. Wang D, Zhang S (2020) Unsupervised person re-identification via multi-label classification. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 10981–10990
4. Chen H, Lagadec B, Bremond F (2021) Enhancing diversity in teacher-student networks via asymmetric branches for unsupervised person re-identification. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp 1–10
5. Lin Y, Dong X, Zheng L, Yan Y, Yang Y (2019) A bottom-up clustering approach to unsupervised person re-identification. In: *Proceedings of the AAAI conference on artificial intelligence*, vol 33, pp 8738–8745
6. Zeng K, Ning M, Wang Y, Guo Y (2020) Hierarchical clustering with hard-batch triplet loss for person re-identification. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 13657–13665
7. Ge Y, Zhu F, Chen D, Zhao R et al (2020) Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID. *Adv Neural Inf Process Syst* 33:11309–11321
8. Wang M, Lai B, Huang J, Gong X, Hua XS (2021) Camera-aware proxies for unsupervised person re-identification. In: *Proceedings of the AAAI conference on artificial intelligence*, vol 35, pp 2764–2772
9. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2017) Mixup: beyond empirical risk minimization. *arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412)*
10. Sun H, Li M, Li CG (2022) Hybrid contrastive learning with cluster ensemble for unsupervised person re-identification. In: *Asian conference on pattern recognition*. Springer, pp 532–546
11. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: *Proceedings of the IEEE international conference on computer vision*, pp 1116–1124
12. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: *Proceedings of the IEEE international conference on computer vision*, pp 3754–3762

UWB and IMU Fusion Positioning Based on ESKF with TOF Filtering



Changhao Piao, Houshang Li, Fan Ren, Peng Yuan, Kailin Wan,
and Mingjie Liu

Abstract Focusing on the problem that UWB and IMU fusion localization has a poor resistance to NLOS, we propose a UWB and IMU fusion algorithm based on the error state Kalman filter combined with TOF filtering. IMU errors are used as the system state variables. The error state is always close to 0, which makes the error state Kalman filter work away from singular values, and gimbal lock, and ensures the rationality and effectiveness of linearization. Then, the differences between the UWB ranging values and the pseudo ranges from the IMU solved position to the UWB anchors are used as the systematic observation to obtain the optimal estimation of the IMU errors. However, UWB measurements are constrained by NLOS, which leads to the reduction of positioning accuracy of the fusion positioning system in the NLOS environment. To solve this problem, this paper establishes a Kalman filter to process UWB measurements by using the property of changing UWB range values at adjacent moments during vehicle movement, which effectively reduces NLOS errors. The experimental results show that the error state Kalman filter combined with the TOF filtering has higher localization accuracy and better numerical stability compared with the error state Kalman filter, which enhances the robustness of the localization system under interference.

Keywords Error state Kalman filter · UWB and IMU fusion positioning · NLOS · TOF Kalman filter

C. Piao · H. Li · M. Liu (✉)
Chongqing University of Posts and Telecommunications, Chongqing, China
e-mail: liumj@cqupt.edu.cn

C. Piao
e-mail: piaoch@cqupt.edu.cn

F. Ren · P. Yuan · K. Wan · M. Liu
Chang'an Intelligent Research Institute, Chongqing, China

1 Introduction

With the development of information technology and intelligence, high-precision vehicle real-time positioning is one of the important components of intelligent vehicles. Currently, Global Navigation Satellite System (GNSS) is a commonly used vehicle positioning system. However, GNSS signals are blocked in some areas such as high-rise cities or underground parking lots, making it impossible to achieve accurate vehicle positioning. Therefore, it is necessary to develop a real-time, high-accuracy localization method to meet vehicle localization in complex urban environments.

With the advantages of high precision, low cost, and strong anti-interference ability, UWB can achieve a positioning accuracy of 10 cm, which makes it a mainstream solution for indoor positioning. In practical applications, complex indoor environmental factors will affect the UWB positioning accuracy, among which the main error sources include clock error, multipath error, and non-line-of-sight (NLOS) error.

Multi-information fusion is effective way to improve the indoor positioning precision. Djosic et al. [5] propose a UWB-based multi-algorithm localization (MAL) method for complex indoor environments, which improves the localization accuracy under mixed LoS/NLoS conditions by combining range-based and fingerprinting-based approaches. Wen et al. [6] proposed a novel IPN method based on shoe-mounted micro-electro-mechanical systems inertial measurement unit and ultra-wideband. Data fusion is implemented by quaternion Kalman filter, which does not involve any complex linearization, reducing the system complexity. In indoor environments, the positioning accuracy can reach 10 cm, and the direction drift can be limited to 0.066 rad. To improve the accuracy and robustness of UAV positioning, Xu et al. [7] propose the FEFIRS algorithm fusing INS and UWB, which exploits the distance between UWB reference node and the blind nodes measured by INS and UWB. FEFIRS is more accurate and robust than the Kalman-based solution.

Currently, the filters commonly used in UWB/IMU fusion positioning systems are KF, EKF, and UKF. In this paper, we use the error state Kalman filter (ESKF), which uses the error state as the state vector compared to the conventional Kalman filter. Because the error is a small quantity, the accuracy loss caused by ignoring a large number of higher-order derivatives in the computation is very small, and the Jacobi matrix is very simple, which greatly accelerates the computation.

2 UWB and IMU Related Work

2.1 UWB

UWB is a positioning technology based on ranging, and the choice of ranging method is crucial, which directly affects the positioning accuracy of the positioning system.

Because both TOA and TDOA methods require clock synchronization between devices, this paper chooses to apply TOF ranging method to UWB ranging.

TOF is a two-way ranging method based on time-of-flight. The UWB signal consists of a series of very short pulses, each with a great spectral broadband. The large bandwidth translates into high temporal resolution, making TOF method highly accurate. UWB, as a wireless signal, propagates at $c = 3 \times 10^8$ m/s. Let the time-of-flight of the UWB signal between the tag and the base station be T_{tof} , then the distance between the tag and the base station is expressed as

$$d = cT_{\text{tof}} \quad (1)$$

2.2 IMU

In this paper, we choose ENU coordinate system as the navigation coordinate system, and the vector variation under the carrier coordinate system can be projected under the navigation coordinate system by quaternion method. The quaternion differential equation is as follows:

$$\dot{q} = \frac{1}{2}\Omega(\omega)q = \frac{1}{2}\Omega(q)\omega \quad (2)$$

where ω is the three-axis angle rate of the gyroscope in the carrier coordinate system.

The rotation matrix from the carrier coordinate system to the navigation system is:

$$C_b^e = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \quad (3)$$

The three-axis acceleration measurement value a^b in the carrier coordinate system can be converted to acceleration vector under the ENU coordinate system.

$$a^e = C_b^e a^b - [0 \ 0 \ g]^T \quad (4)$$

where g is the local gravitational acceleration. The current speed and position of the car can be obtained by integrating the acceleration a^e .

3 UWB and IMU Fusion Positioning

In ESKF, we usually take the original state variable as the nominal state and then use the error state as the ESKF state variable. When the IMU measurement data arrives, we put it into the nominal state variable after integrating it. Since this approach does not take noise into account, its results naturally drift quickly. So, we want to put the error in the ESKF, which internally takes into account the effects of various noises and zero biases and gives an optimal estimate of the error state. The ESKF itself, as a Kalman filter, has a prediction process and a correction process, where the correction process relies on sensor observations (UWB) other than the IMU. The structure of the UWB/IMU fusion positioning system based on ESKF is shown in Fig. 1.

3.1 ESKF

For a more precise description of the UWB and IMU fusion positioning system state, position error, velocity error, misalignment angle error, accelerometer drift error and gyroscope drift error are used as the system state vectors, expressed as follows:

$$X_k = \left[\delta x \ \delta y \ \delta z \ \delta v_x \ \delta v_y \ \delta v_z \ \delta \theta_e \ \delta \theta_n \ \delta \theta_u \ \delta a_x \ \delta a_y \ \delta a_z \ \delta g_x \ \delta g_y \ \delta g_z \right] \quad (5)$$

Among them, $\delta x, \delta y, \delta z$ are position errors, $\delta v_x, \delta v_y, \delta v_z$ are velocity errors, $\delta \theta_e, \delta \theta_n, \delta \theta_u$ are misalignment errors, $\delta a_x, \delta a_y, \delta a_z$ are accelerometer drift errors, $\delta g_x, \delta g_y, \delta g_z$ is the gyroscope drift error.

Let the continuous-time state transfer matrix of ESKF be F , the dynamic noise matrix be G , and the process white noise be W . Then the state equation is expressed as

$$\dot{X}(t) = F(t)X(t) + G(t)W(t) \quad (6)$$

Continuous-time state transfer matrix F .

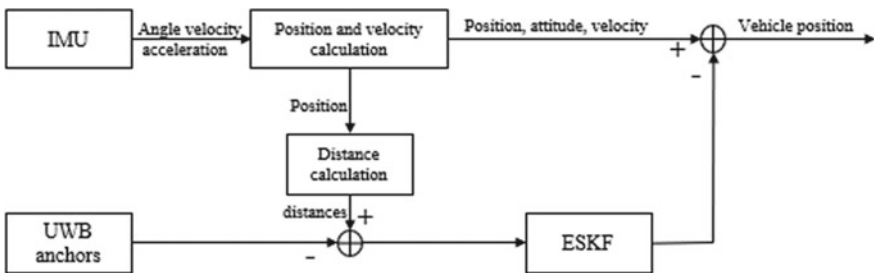


Fig. 1 UWB/IMU fusion positioning system based on ESKF

$$F = \begin{bmatrix} 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & -sk & -C_b^n & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & -C_b^n \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix} \quad (7)$$

where $0_{3 \times 3}$ denotes the 3×3 all-zero matrix, $I_{3 \times 3}$ denotes the 3×3 unit matrix, C_b^n is the rotation matrix from the carrier coordinate system to the ENU coordinate system, and sk denotes the antisymmetric matrix after the projection of the tri-axial acceleration under the carrier coordinate system to the ENU coordinate system.

$$sk = \begin{bmatrix} 0 & -\tau_1 & \tau_2 \\ \tau_3 & 0 & -\tau_1 \\ -\tau_2 & \tau_1 & 0 \end{bmatrix} \quad (8)$$

$$[\tau_1 \ \tau_2 \ \tau_3]^T = C_b^n \times a_m \quad (9)$$

where a_m is the accelerometer measurement. Let Δt be the IMU sampling timestep, then the discretized state transfer matrix Φ_k is expressed as

$$\Phi = \begin{bmatrix} I_{3 \times 3} & \Delta t \times I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & I_{3 \times 3} & -\Delta t \times sk & -\Delta t \times C_b^n & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} & \Delta t \times C_b^n \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \end{bmatrix} \quad (10)$$

The discretized dynamic noise driving matrix G_k is as follows:

$$G_k = \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ C_b^n \times \Delta t & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & -C_b^n \times \Delta t & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \times \Delta t & I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \times \Delta t \end{bmatrix} \quad (11)$$

Thus, the discretized system state equation is as follows:

$$X(k) = \Phi X(k-1) + G_{k-1} W_k \quad (12)$$

Since we use a tight combination of UWB and IMU, UWB does not solve the position information. In order to correct the inertial guidance error using the UWB ranging information, the distance from the IMU solved position to each UWB anchor needs to be calculated, which is then differenced from the UWB ranging value as the observed value. The UWB and IMU observation equation can be expressed as

$$Z(k) = \begin{bmatrix} d_1^{\text{uwb}} - d_1^{\text{imu}} \\ d_2^{\text{uwb}} - d_2^{\text{imu}} \\ \vdots \\ d_n^{\text{uwb}} - d_n^{\text{imu}} \end{bmatrix} \quad (13)$$

where d_i^{uwb} is UWB observation and d_i^{imu} is the distance from the location as derived from IMU trajectory to each UWB anchor. H is the measurement transition matrix.

$$H = \begin{bmatrix} 0_{n \times 3} & J_{n \times 3} & 0_{n \times 3} & 0_{n \times 3} & 0_{n \times 3} & 0_{n \times 3} \end{bmatrix} \quad (14)$$

where $J_{n \times 3}$ is the Jacobi matrix.

$$J_{n \times 3} = \begin{bmatrix} \frac{x - x_1^{\text{anchor}}}{d_1^{\text{imu}}} & \frac{y - y_1^{\text{anchor}}}{d_1^{\text{imu}}} & \frac{z - z_1^{\text{anchor}}}{d_1^{\text{imu}}} \\ \frac{x - x_2^{\text{anchor}}}{d_2^{\text{imu}}} & \frac{y - y_2^{\text{anchor}}}{d_2^{\text{imu}}} & \frac{z - z_2^{\text{anchor}}}{d_2^{\text{imu}}} \\ \vdots & \vdots & \vdots \\ \frac{x - x_n^{\text{anchor}}}{d_n^{\text{imu}}} & \frac{y - y_n^{\text{anchor}}}{d_n^{\text{imu}}} & \frac{z - z_n^{\text{anchor}}}{d_n^{\text{imu}}} \end{bmatrix} \quad (15)$$

$(x_i^{\text{anchor}}, y_i^{\text{anchor}}, z_i^{\text{anchor}})$ are the coordinates of the i th UWB anchor in the Cartesian coordinate system.

Assume that both the process noise vector W_k and the measurement noise vector V_k are Gaussian noise with covariances Q and R . The ESKF process is as follows:
Error state prediction

$$\bar{X}(k) = \Phi \hat{X}(k-1) \quad (16)$$

Predictive covariance matrix

$$\bar{P}_k = \Phi \hat{P}_{k-1} \Phi + G_{k-1} Q G_{k-1} \quad (17)$$

Calculating Kalman gain

$$K_k = \bar{P}_k H^T (H \bar{P}_k H^T + R)^{-1} \quad (18)$$

Status estimation

$$\hat{X}(k) = \bar{X}(k) + K_k (Z(k) - H \bar{X}(k)) \quad (19)$$

State estimation covariance matrix

$$\hat{P}_k = (I - K_k H) \bar{P}_k \quad (20)$$

Updating nominal state

$$X_{\text{nominal}} = X \oplus \hat{X}(k) \quad (21)$$

3.2 UWB Measurements Filtering

In the ESKF-based UWB and IMU fusion positioning system, the observation is derived from the difference between the UWB range value and the IMU solved pseudo-range. The complex environmental factors such as non-visual range and multipath effects in indoor scenes can lead to unstable and reliable observation data with outliers. These outliers can adversely affect fused UWB and IMU positioning accuracy. To suppress the outliers, we use the Kalman filter to deal with the UWB distances.

Using the nature of the change in UWB measurements at adjacent moments during vehicle movement, the state equation can be derived. Assuming that there are n UWB anchors, the distance value d_i^{uwb} and the rate of change of the distance value from the previous moment to the current moment v_i^{uwb} are used as the state vectors.

$$X_m(k) = \left[d_1^{\text{uwb}} \ d_2^{\text{uwb}} \ \dots \ d_n^{\text{uwb}} \ v_1^{\text{uwb}} \ v_2^{\text{uwb}} \ \dots \ v_n^{\text{uwb}} \right]^T \quad (22)$$

Then the equation of state is expressed as

$$X_m(k) = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix} X_m(k-1) + W_m(k) \quad (23)$$

The observation equation is expressed as

$$Z_m(k) = \begin{bmatrix} d_1^{\text{uwb}} \\ d_2^{\text{uwb}} \\ \vdots \\ d_n^{\text{uwb}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} X_m(k) + v_m(k) \quad (24)$$

The NLOS error is always greater than 0 because the NLOS error is caused by the delay in the electromagnetic wave encountering an obstacle on its way to propagation. When the Kalman filter processes the UWB measurements, there is a situation where the predicted value will be greater than the observed value. Due to the positive bias of

the NLOS error, we do not process the data at this time. Instead, when the predicted value in the prediction stage is smaller than the observed value, the Kalman gain is used to obtain the optimal estimate by weighting the sum of the predicted and observed values.

4 Experiment

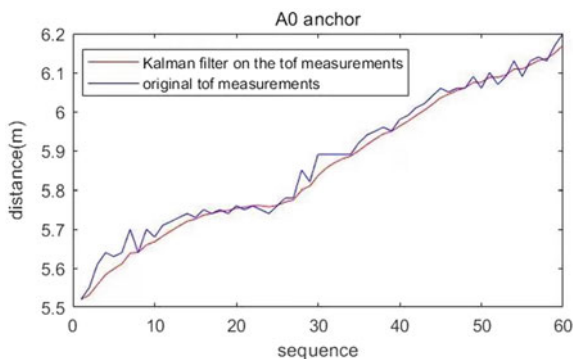
We build a UWB quad anchor indoor positioning scenario where the four anchors are A0 (3, 0, 1.6), A1 (0, 0, 1.6), A2 (0, 6.6, 1.6), and A3 (3, 6.6, 1.6). The trolley equipped with a UWB tag and an IMU sensor follows a rectangular trajectory in the field. At the same time, we collected UWB ranging data and IMU sensor tri-axial acceleration and tri-axial angular rate data with a sampling frequency of 100 Hz.

At first, we filtered the UWB measurements to reduce the TOF errors. From Fig. 2, it can be obtained that the filtered UWB measurements vary more smoothly than the original UWB measurements. Therefore, the TOF Kalman filter method proposed in this paper can effectively reduce the error of TOF.

Next, we compute the localization solution on the measurements to verify the impact of the TOF Kalman filter method on the overall UWB and IMU fusion positioning system.

The positioning trajectories of fused UWB and IMU are shown in Fig. 3, which shows that: the positioning trajectory using only UWB significantly deviate when the ranging values are abnormal. The positioning trajectory of the ESKF-based UWB and IMU fusion is much smoother than that of the only UWB, but it has obvious offset at some places, so it is not as effective as the ESKF with the combined TOF Kalman filter. By comparing the x -axis coordinates and y -axis coordinates of each trajectory with the real trajectory, the average error of each positioning method is obtained, as shown in Table 1. The accuracy of ESKF with the TOF KF method is higher than the other two methods.

Fig. 2 Comparison of TOF measurements before and after filtering



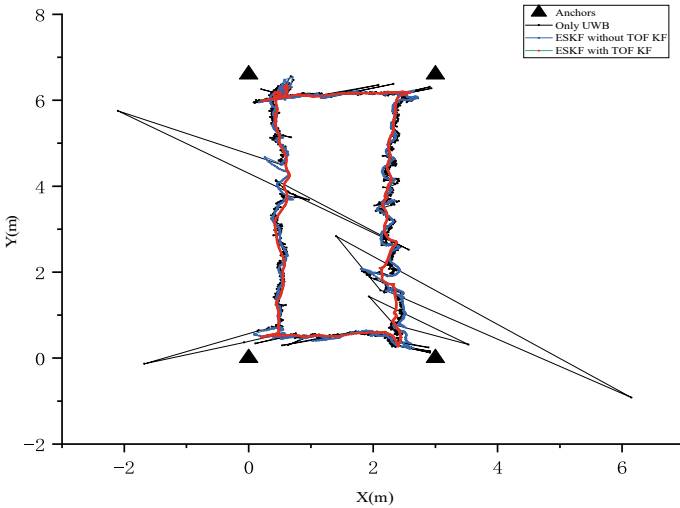


Fig. 3 Trolley positioning trajectories

Table 1 Positioning errors comparison

Method	Error in the x -axis direction (m)	Error in the y -axis direction (m)
Only UWB	0.37	0.33
ESKF	0.20	0.19
ESKF with TOF KF	0.15	0.16

5 Conclusion

For the UWB and IMU fusion positioning system, firstly, the state equation is established based on the IMU motion error state model, and secondly, the UWB ranging information is used as the measurement data of the UWB and IMU fusion positioning system to establish the observation equation. Then, the TOF Kalman filter is designed for the observed outliers. Finally, the experiments show that the ESKF method with the TOF Kalman filter is more suitable for UWB and IMU data fusion, whose localization effect is better than the original ESKF method. The localization trajectory obtained by the method in this paper is closer to the real trajectory, with higher localization accuracy and better localization stability in the face of abnormal observations.

Acknowledgements This work was supported by National Key R&D Program (No. 2022YFE0101000).

References

1. Zhang S et al. Nothing blocks me: precise and real-time LOS/NLOS path recognition in RFID systems. *IEEE Internet Things J* 6(3):5814–5824
2. Zheng Q et al. Channel non-line-of-sight identification based on convolutional neural networks. *IEEE Wireless Commun Lett* 9(9):1500–1504
3. Wang J, Yu K, Bu J, Lin Y, Han S. Multi-classification of UWB signal propagation channels based on one-dimensional wavelet packet analysis and CNN. *IEEE Trans Veh Technol*
4. Huang C et al. Machine learning-enabled LOS/NLOS identification for MIMO systems in dynamic environments. *IEEE Trans Wireless Commun* 19(6):3643–3657
5. Djosic S, Stojanovic I, Jovanovic M, Djordjevic GL (2022) Multi-algorithm UWB-based localization method for mixed LOS/NLOS environments. *Comput Commun* 181:365–373. ISSN 0140-3664
6. Wen K, Yu K, Li Y, Zhang S, Zhang W. A new quaternion Kalman filter based foot-mounted IMU and UWB tightly-coupled method for indoor pedestrian navigation. *IEEE Trans Veh Technol* 69(4):4340–4352
7. Xu Y, Shmaliy YS, Ahn CK, Shen T, Zhuang Y. Tightly coupled integration of INS and UWB using fixed-lag extended UFIR smoothing for quadrotor localization. *IEEE Internet Things J* 8(3):1716–1727

Observation Noise Covariance Matrix Initialization-Based Objective State Estimation for Kalman Filter Using SVR



Junhu Chen, Mingjie Liu, Fan Ren, Peng Yuan, Jianbin Chen, Yiwei Ma, Tai Liu, and Changhao Piao

Abstract Kalman filtering is a filtering algorithm for optimal estimation of the system state. The optimal estimate of the system state is obtained through iteratively updating the mean and variance of the system by establishing the equations of motion and observation equations for the system. In this paper, the Kalman filtering algorithm is used to estimate the state of multiple targets which is collected by vehicle radar. Since the true values are not accurate enough, support vector regression (SVR) is applied to fit the multi-target data collected by the radar to obtain the approximate true values. Then, the observation error and the sample variance of the observation error are calculated. Finally, the mean of the variances of the different physical quantities is formed into a diagonal matrix as the initialized value of the observation noise covariance matrix R_k . The Kalman filtering of mean initialized R_k^{mean} is then compared with the empirically initialized R_k^{emp1} for state estimation, it is found that our method can improve the accuracy of the Kalman filtering for state estimation of the target.

Keywords Kalman filtering · Multi-objective state estimation · Support vector regression · Observation noise covariance matrix

1 Introduction

Target tracking algorithm, which can be divided into single target tracking and multiple object tracking, is a very practical algorithm in the field of computer vision,

J. Chen · M. Liu (✉) · Y. Ma · C. Piao
Chongqing University of Posts and Telecommunications, Chongqing, China
e-mail: liumj@cqupt.edu.cn

M. Liu · F. Ren · P. Yuan · J. Chen
Chang'an Intelligent Research Institute, Chongqing, China

T. Liu
China Academy of Information and Communications Technology, Beijing, China

and its application scenarios are very wide, such as UAVs [1], autonomous driving [2], intelligent monitoring [3], and medical diagnosis [4]. The problem of multi-target tracking for autonomous driving has also received a lot of attention in recent years, and multi-target tracking for autonomous driving involves techniques such as target detection, data association matching of targets, and state estimation of targets. The more accurate the target state estimation is, the better the target tracking will be. State estimation is to predict and observe the state of multiple target motions in front of the vehicle, and then data form the target data observed by different sensors to finally arrive at the optimal estimate of the target motion state. Common used state estimation methods include Bayesian filtering, Kalman filtering, particle filtering, and so on. Kalman filter is the classical state estimation algorithm, which is widely used in navigation, robot motion planning, finance, and many other fields. In multiple target tracking, for example, SORT [5], DeepSORT [6], FairMOT [7], which are some of the more classical multi-target tracking algorithms, all apply Kalman filter to predict and update the state of the target and estimate the optimal state of the target.

Kalman filtering was introduced as a system of equations of state for linear systems [8]. It mainly processes the observation data that needs filtering and noise reduction in order to restore the real data to the maximum extent possible. For a discrete motion system, a motion model and an observation model can be established with containing process noise w_k and observation noise v_k , respectively. In a real motion system, w_k and v_k are variables that change with the dynamics of the system. In the Kalman filter, it assumed that w_k and v_k are mutually independent random variables, which obeys $w_k \sim N(0, Q_k)$ and $v_k \sim N(0, R_k)$ [9]. Q_k and R_k , respectively, are the process noise covariance matrix and observation noise covariance matrix. If the value of Q_k and R_k is closer to the true value of the system, then the state estimate of Kalman filter is also closer to the true value. Therefore, the initialization of Q_k and R_k of Kalman filter then affects the optimal estimation of it.

The true of Q_k and R_k are difficult to obtain. Commonly, Q_k and R_k are set empirically as fixed values, which may cause poor estimation by Kalman filter. As for a measurement system, the observed values are obtained by sensor acquisition, but the data must contain observation noise, and the true values (theoretical values without noise) are difficult to obtain. Many scholars study on the observation noise and propose various Kalman filtering for adaptive noise [10], but the computation effort will increase. In contrast, the initialization of the observation noise covariance matrix has rarely been studied and dealt with, so when the initialization of the observation noise covariance matrix R_k is not reasonable, then for using the non-adaptive Kalman filter for target state estimation, it will have an impact on the state estimation.

The data in this paper mainly contains the longitudinal and lateral distances (nestCenter.x, nestCenter.y) and longitudinal and lateral relative velocities (velocity.x, velocity.y) of multiple targets in front of the radar and the radar; and the number of frames and target IDs corresponding to the above four observed and true values. Since the true values are not accurate enough, we uses support vector regression (SVR) to obtain the approximate true values, so as to calculate the observation error and variance of the observation error. Then, the observation error sample variance is averaged and the mean of the observation error sample variance of nestCenter.x,

nestCenter.y, velocity.x, and velocity.y is formed into a diagonal matrix and used as the observation noise covariance matrix R_k . The processor used for testing in this thesis is Intel Core i5-10500, the programming language is Python 3.8, and the machine learning library is scikit-learn 1.1.

2 Objective State Estimation Model Construction

2.1 Support Vector Regression Principle

Support vector regression (SVR) is a branch of support vector machine (SVM) that can deal with nonlinear, time series, and other regression problems. The core idea is to map the sample data to a high-dimensional feature space through a nonlinear transformation, using a function $f(x) = w^T x + b$ to fit the relationship between the input and output variables so that $f(x)$ and y are as close as possible, the $f(x)$ and y are weights and intercepts, respectively [11].

From the second section, it can be found that the true value data deviates from the observed values mainly in the transverse direction (y-axis) in terms of distance and velocity, and the true value has a low degree of confidence. SVR is used to fit the observed data. The input sample data x are the longitudinal and lateral relative velocities, longitudinal and lateral distances, and the corresponding frame numbers. The input sample data y are the radar observations of longitudinal distance, lateral distance, longitudinal relative velocity, and lateral relative velocity, respectively. The output $f(x)$ is the approximate true value after fitting the above four radar observations. The deviation ε of y and $f(x)$ can be considered as the maximum permissible observation error. Construct an interval strip of 2ε width at the center $f(x)$ and solve the optimization problem [12].

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m l_\varepsilon(f(x_i) - y_i), \quad (1)$$

C is the penalty coefficient, which represents the tolerance of the model to error; and l_ε is the loss function. Then, the optimization problem can be transformed into a pairwise problem.

$$\max_{\alpha, \hat{\alpha}} \sum_{i=1}^m y_i (\hat{\alpha}_i - \alpha_i) - \varepsilon (\hat{\alpha}_i + \alpha)_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_i - \alpha_j) K(x_i x_j), \quad (2)$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \\ 0 \leq \alpha_i, \hat{\alpha}_i \leq C \end{cases}, \quad (3)$$

$K(x_i, x_j)$ in Eq. (2) is the kernel function, through which the inner product of nonlinear variations is obtained. In this paper, the Gaussian kernel function (RBF kernel) is selected to be reflected in the SVR, which gives

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_j) k(x_i, x_j) + b, \quad (4)$$

$$\begin{aligned} K(x, x) &= \exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \\ &= \exp(-\text{gamma}\|x_i - x_j\|^2), \quad \sigma > 0, \end{aligned} \quad (5)$$

gamma is the kernel function parameter, this paper seeks more suitable parameters C and gamma, to improve the fitting effect of SVR by grid search method. The longitudinal distance, lateral distance, longitudinal relative velocity, and lateral relative velocity of each target are fitted to the model using SVR separately, and the fitted true values are $f_{\text{nestCenter.x}}(x(\text{velocity.x, frame}))$, $f_{\text{nestCenter.y}}(x(\text{velocity.y, frame}))$, $f_{\text{velocity.x}}(x(\text{nestCenter.x, frame}))$, and $f_{\text{velocity.y}}(x(\text{nestCenter.y, frame}))$, respectively.

2.2 Mean Initialized R_k^{mean}

After approximating the SVR-fitted data to true values, the observation error and the observation error sample variance are calculated. Then, the nestCenter.x, nestCenter.y, velocity.x, and velocity.y of each target have the corresponding observation error sample variance. Each set of test sample data is 50 frames, and then the observation error sample variance of the target is summed and averaged

$$\bar{s}_j^2 = \frac{1}{m} \left(\sum_{n=1}^{50} s_{i(\text{obj}_1)}^2 + \cdots + \sum_{n=1}^{50} s_{i(\text{obj}_m)}^2 \right), \quad (6)$$

where i represents nestCenter.x, nestCenter.y, velocity.x, and velocity.y, m is the number of detection targets, and the test sample data of different groups are different because the detection targets are dynamic and the targets may disappear or increase with time. The observation noise covariance matrix R_k can be obtained as follows when i represents different physical quantity in Eq. (6)

$$R_k^{\text{mean}} = \begin{bmatrix} \sigma_{\text{nestCenter.x}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\text{nestCenter.y}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\text{velocity.x}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\text{velocity.y}}^2 \end{bmatrix}$$

$$\approx \begin{bmatrix} \bar{s}_{\text{nestCenter}.x}^2 & 0 & 0 & 0 \\ 0 & \bar{s}_{\text{nestCenter}.y}^2 & 0 & 0 \\ 0 & 0 & \bar{s}_{\text{velocity}.x}^2 & 0 \\ 0 & 0 & 0 & \bar{s}_{\text{velocity}.y}^2 \end{bmatrix}. \tag{7}$$

2.3 Kalman Filtering Principle

Kalman filtering is an algorithm for optimal estimation of data processing by iterating over the mean and variance to reach the optimal estimated state. In this paper, the multiple target states of the self-vehicle radar detection motion are shown in Fig. 1.

The number of targets in Fig. 2 is dynamically changing in different time regions, and a discrete state-space model is established for the target state, then the equations of motion and observation are as in Eqs. (8) and (9).

$$X_k = AX_{k-1} + w_{k-1}, \tag{8}$$

$$Y_k = HX_k + v_k, \tag{9}$$

X_k in the formula is the state of the target at the time k , which is expressed as in Eq. (10)

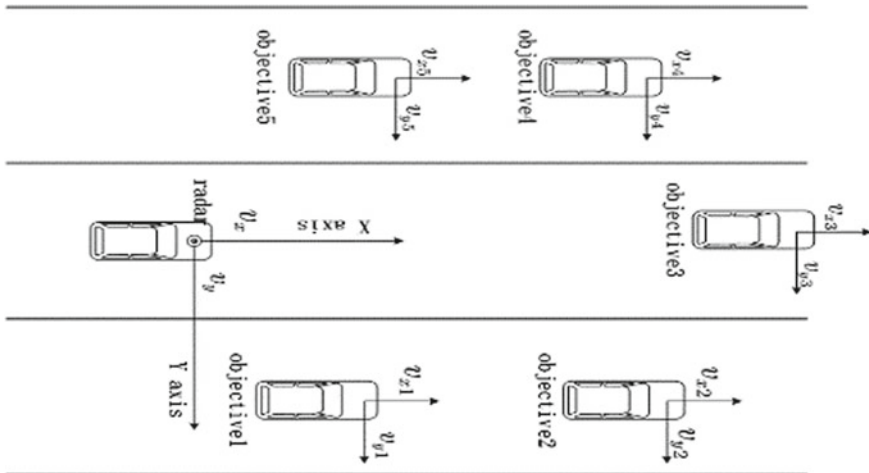


Fig. 1 Multi-target status for radar detection

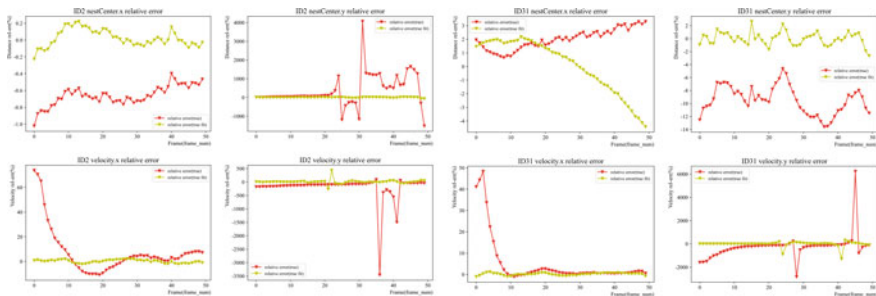


Fig. 2 Relative error between the state observations and the true values and the true values of the SVR fit for the five targets in frames 0–49

$$X_k = \begin{bmatrix} \text{nestCenter_}x_k(\text{obj}_1) & \text{nestCenter_}x_k(\text{obj}_2) & \cdots & \text{nestCenter_}x_k(\text{obj}_m) \\ \text{nestCenter_}y_k(\text{obj}_1) & \text{nestCenter_}y_k(\text{obj}_2) & \cdots & \text{nestCenter_}y_k(\text{obj}_m) \\ \text{velocity_}x_k(\text{obj}_1) & \text{velocity_}x_k(\text{obj}_2) & \cdots & \text{velocity_}x_k(\text{obj}_m) \\ \text{velocity_}y_k(\text{obj}_1) & \text{velocity_}y_k(\text{obj}_2) & \cdots & \text{velocity_}y_k(\text{obj}_m) \end{bmatrix}. \quad (10)$$

The Kalman filtering algorithm is mainly divided into two processes: prediction process is to predict the state and error covariance of the current moment from the target state estimate of the previous moment; while the correction process is to form a correction based on the current observed value and the current predicted value to obtain an estimate closer to the true value.

Prediction

$$X_{k|k-1} = AX_{k-1} \quad (11)$$

$$P_{k|k-1} = AP_{k-1}A^T + Q_{k-1} \quad (12)$$

Update

$$K_k = P_{k|k-1}H^T(H P_{k|k-1}H^T + R_k)^{-1} \quad (13)$$

$$X_k = X_{k|k-1} + K_k(Y_k - HX_{k|k-1}) \quad (14)$$

$$P_k = (I - K_kH)P_{k|k-1} \quad (15)$$

In this paper, the time interval between the data at the last moment and the current data is 0.1 s, $\Delta t = 0.1$; the time interval between the data is small and the change is small, and it can be assumed that the targets are all moving at a uniform speed, then the Kalman filter related parameters can be set as

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (16)$$

The value of R_k as in Eq. (7), and in addition, since the process noise is also unknown, then assuming that the process noise covariance matrix $Q_k = R_k^{\text{mean}}$.

3 Experiment

3.1 Support Vector Regression Fitting Results

The approximate true values are fitted by regression of the observed values on the SVR. Then, the relative errors between the observed and the original true values and the relative errors between the observed and the approximate true values are calculated separately, and the two relative errors are compared as shown in Fig. 2.

The comparison of the relative error in Fig. 1 shows that the relative error means between the observed values and the true values of the SVR fit are smaller, and then the true values of the SVR fit can be considered more plausible.

3.2 Multi-objective State Estimation Results

After the mean initialization matrix R_k^{mean} , the multi-objective state estimation is performed on 20 sample data sets by Kalman filtering for prediction and correction. In addition, the empirical initialized R_k^{emp1} Kalman filter is used to estimate the states of the 20 data samples, and the empirical initialization can be set as R_k^{emp1}

$$R_k^{\text{emp1}} = \text{diag}(\sigma_1^2 = 0.5, \dots, \sigma_4^2 = 0.5). \quad (17)$$

The multi-objective state estimation pairs after mean initialized R_k^{mean} and empirical initialized R_k^{emp1} are shown in Figs. 3 and 4.

The mean square error between the Kalman filter estimates and the true value after mean initialized R_k^{mean} and empirical initialized R_k^{emp1} (ID2 and ID31 in sample data, frames 0–49) is shown in Table 1; the mean square error of the estimated and true values of the Kalman filter after mean initialization R and empirical initialization R (the mean of the mean square error of the overall calculation of 20 sets of data) is shown in Table 2.

From Figs. 3, 4 and Tables 1, 2, it can be seen that the Kalman filter multi-target state estimation after mean initialized R_k^{mean} is better than empirical initialized R_k^{emp1} ;

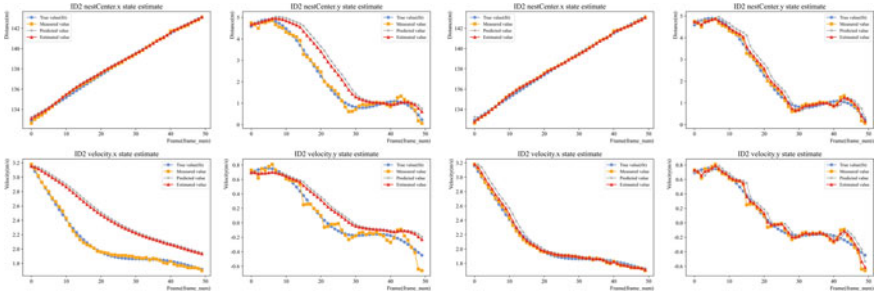


Fig. 3 Comparison of state estimation of the target (ID2) using our method and empirical initialization (left for empirical initialized R_k^{emp1} , right for mean initialized R_k^{mean})

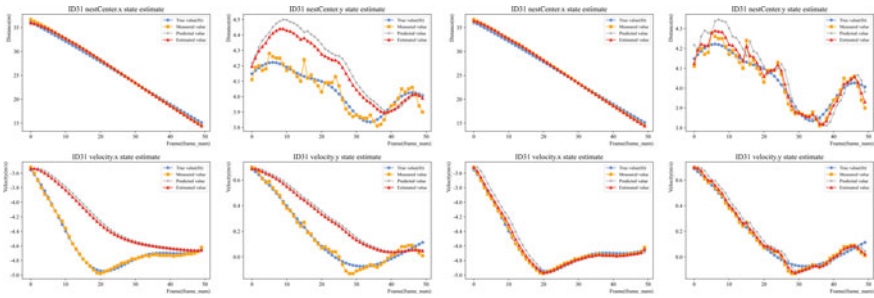


Fig. 4 Comparison of state estimation of the target (ID31) using our method and empirical initialization (left for empirical initialized R_k^{emp1} , right for mean initialized R_k^{mean})

Table 1 Mean squared error of Kalman filter estimates and true values after mean initialized R_k^{mean} and empirical initialized R_k^{emp1} (ID2 and ID31 in sample data, frames 0–49)

Emp1 and mean	MSE (nCx)	MSE (nC _y)	MSE (vx)	MSE (vy)
ID2 (emp1)	0.028455	0.363569	0.135969	0.020626
ID2 (mean)	0.016042	0.024611	0.001339	0.003177
ID31 (emp1)	0.205763	0.025882	0.157672	0.026889
ID31 (mean)	0.205412	0.001449	0.001259	0.000882

Table 2 Mean square error of the estimated and true values of the Kalman filter after mean initialization R and empirical initialization R (the mean of the mean square error of the overall calculation of 20 sets of data)

	Emp1	Mean
MSE (mean)	0.214	0.165

it can also be seen that empirical initialized R_k^{emp1} requires several adjustments to improve the estimation of the Kalman filter, but this requires constant adjustment of the parameters, which increases the workload.

4 Conclusion

In this paper, we focus on the four observations of longitudinal and lateral distance (nestCenter.x, nestCenter.y) and longitudinal and lateral relative velocity (velocity.x, velocity.y) of multiple targets in front of the vehicle and the radar for Kalman filter state estimation. First, the first 1000 frames of data are selected and divided into 20 sets of sample data for testing. Since the true value is not accurate enough and the observation noise is difficult to be determined, the observation error and observation error variance of the corresponding frame are calculated by SVR regression fitting to the true value of the sample data, and then the mean of the observation error variance of the sample data is calculated as the mean initialized R_k^{mean} . Finally, the Kalman filter state estimation is compared with the empirical initialized R_k^{emp1} . By comparing the estimation accuracy and mean square error, we can see that the Kalman filter algorithm after mean initialized R_k^{mean} can avoid repeatedly adjusting the parameter R_k and improve the accuracy of multi-objective state estimation to some extent.

Acknowledgements This work was supported by National Key R&D Program (No. 2022YFE0101000).

References

1. Wang P, Wu L, Qi J, Dai J (2022) Unmanned aerial vehicles object detection based on image haze removal under sea fog conditions. *IET Image Process* 16(10)
2. Li ZX, Liu JM, Li S et al (2015) Group targets tracking algorithm based on box particle filter. *Acta Autom Sin* 41(4):785–798
3. Ciaparrone G, Sánchez FL, Tabik S et al (2019) Deep learning in video multi-object tracking: a survey
4. Cheng H, Hsu C, Hung C et al (2021) A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies. *Biomed J*
5. Bewley A, Ge Z, Ott L et al (2016) Simple online and realtime tracking. In: 2016 IEEE international conference on image processing (ICIP). IEEE
6. Wojke N, Bewley A, Paulus D (2017) Simple online and realtime tracking with a deep association metric. In: 2017 IEEE international conference on image processing (ICIP). IEEE
7. Zhang Y, Wang C, Wang X et al (2020) FairMOT: on the fairness of detection and re-identification in multiple object tracking
8. Ge Z, Chang F, Liu H (2018) Multi-target tracking based on Kalman filtering and optical flow histogram. In: 2017 Chinese automation congress (CAC). IEEE
9. Li Z, Zhang J, Wang J et al (2019) Recursive noise adaptive extended object tracking by variational Bayesian approximation. *IEEE Access* PP(99):1

10. Tripathi RP, Ghosh S, Chandle JO (2016) Tracking of object using optimal adaptive Kalman filter. In: IEEE international conference on engineering & technology. IEEE
11. Tingyu Z (2020) Short-term stock trend prediction based on improved SVR. In: 2020 7th international conference on information science and control engineering (ICISCE), pp 1283–1286. <https://doi.org/10.1109/ICISCE50968.2020.00259>
12. Ye T (2017) Stock forecasting method based on wavelet analysis and ARIMA-SVR model. In: 2017 3rd international conference on information management (ICIM), pp 102–106. <https://doi.org/10.1109/INFOMAN.2017.7950355>

Development of the Fire Analysis Framework for the Thermal Power Plant



Chai-Jong Song and Jae-Yun Park

Abstract We develop the fire analysis framework including dynamic factors of fire hazard to build comprehensive fire protection system that can prevent occurrence of fire in advance in a thermal power plant. The framework contains several engines to acquire, to store, and to process various types of data including the real time sensor data generated by various equipment. In order to process received various types of sensor data, a time series data processing engine is essential. In accordance with these requirements, the framework support the data pipeline with received data, and support various protocols for sharing data with related system like digital twin. We develop dashboard for intuitive monitoring fire hazard index by zone or equipment in real time. The framework is used data-driven IoT system for supporting big data and artificial intelligence to find solution of fire hazard of a thermal power plant.

Keywords Fire hazard index · Fire analysis framework · Time-series data · Thermal power plant

1 Introduction

It is very important to predict and remove the fire risk in a thermal power plant in advance. However, it is the current situation that most thermal power plants have no dynamic fire protection system which the time variable is applied in place. Because all they have to do is comply with the minimum regulations stipulated by law. The existing firefighting system is insufficient to preemptive response since most of the firefighting equipment are installed and operated for the purpose of detecting the

C.-J. Song (✉) · J.-Y. Park
Korea Electronics Technology Institute, Seoul 03924, Korea
e-mail: jcsong@keti.re.kr

J.-Y. Park
e-mail: alboutyou@keti.re.kr

occurrence of a fire and extinguishing fire. Therefore, it is essential to build a comprehensive fire protection system with predicting the fire risk and response preemptive in advance.

There were many researches to monitor and predict environmental condition or disasters using IoT-based prediction models [1, 2]. Data-driven IoT system based on big data and Artificial Intelligence (AI) is used various applications to find solutions of various disaster or risk [3, 4]. The system consisted of big data, AI and IoT is becoming basic options to various application of monitoring or finding solutions.

There are numerous sensors in the thermal power plant to monitor the status of various equipment. These sensors are designed and operated for the purpose of detecting abnormalities of equipment or for efficiency of operation but the out of focus in fire risk. Utilizing these sensors is not appropriate in perspective of fire directly. So, there must be step to select sensors related to fire hazard from among them and then it would be very useful to predict the fire risk. In addition, a better dynamic fire protection system can be built if environmental sensors are installed in zone of operated equipment.

In this paper, we developed the fire analysis framework that can provide Fire Hazard Index (FHI) for each zone and equipment in a thermal power plant based on analyzing various sensor data. The framework is designed to acquire sensor data and provide results per second, and the analyzed results can be used by related systems such as digital twins in real time.

2 Implementation of the Framework

The framework is designed to manage and process various types of data for providing the fire risk of each zone and equipment based on analyzing sensor data generated from various equipment of thermal power plant. To this end, the framework contains a time series data processing engine and static data processing engine. A time series data processing engine is essential to analyze sensor data transmitted in real time. In accordance with these requirements, the framework can be separated by functional engines and each engine is developed independently for the specific purpose.

Separating the framework as microservices not only increases the efficiency of development, but also takes advantages in maintenance. However, microservice could raise issue of concurrency problem in order to share data and functions within the overall framework [5, 6]. To this end, we adopted RMQ (RabbitMQ) as the message queue system and Kafka as streaming process engine. The framework supports various protocols such as REST (RESTful), Message Queuing Telemetry Transport (MQTT), Advanced Message Queuing Protocol (AMQP), Simple Text Oriented Messaging Protocol (STOMP), WebSocket for communication with the related systems.

2.1 Data Acquisition and Refinery Engine

The equipment sensor data such as vibration, temperature, pressure, flow are considered important for the operation and maintenance of the power plant. Composite IoT sensors including heat, smoke, humidity, temperature, and carbon monoxide are considered important for fire detection. The acquisition and refinery engine is designed that can acquire and process IoT sensor and equipment data separately. The engine runs as an independent service and creates pipeline using collected and processed data. That pipeline is used in data analysis and sharing engine. Since data must be processed in real time and asynchronously, the engine adopts MQTT and AMQP which are asynchronous protocols based on Message Queue (MQ), as default.

2.2 Sensor Tag Processing Engine

In order to expand the fire protection system, sensor tag processing engine is needed to automatically extract and manage selected tags from each equipment. Figure 1 shows a block diagram designed to register sensor tags and devices automatically. When request comes in through REST API to register new tag or update existed tag, API Server delivers the request to the message queue broker.

The message queue broker routes the request to the queue registered in topic format and then the queue processes the request. Processed request is applied to the RDBMS and updates the tag information registered in the framework. Through this process, the tag status is always kept up to date in real time.

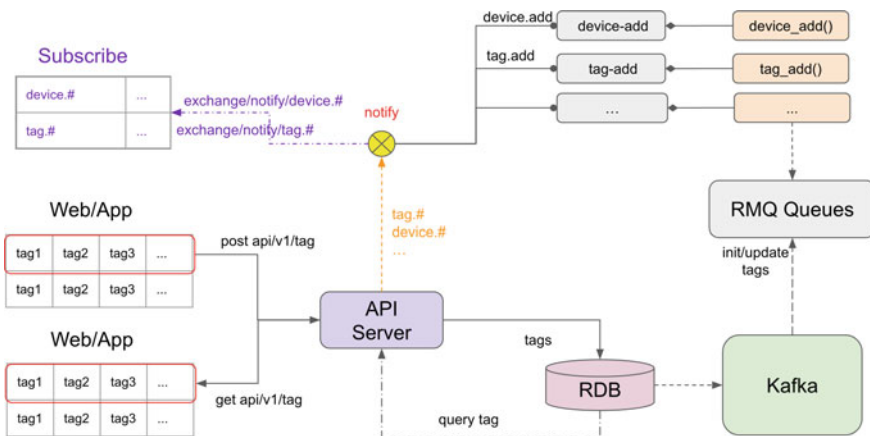


Fig. 1 Sensor tag processing engine diagram

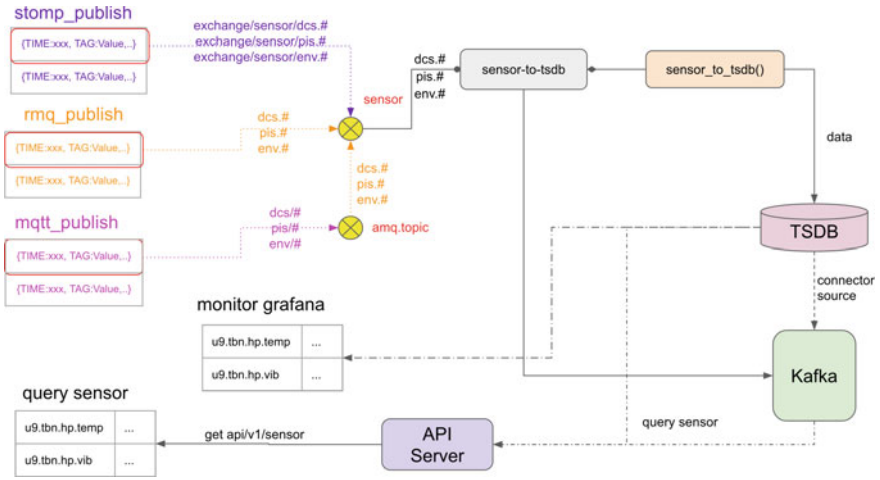


Fig. 2 Sensor data processing engine diagram

2.3 Sensor Data Processing Engine

Sensor data processing engine is designed to acquire, store, and process sensor data based on the registered tag information. The framework must acquire and process more than hundreds of sensor data simultaneously in real time. Therefore, suitable data pipeline is formed by the protocol capable of processing large amounts of data. All of data required in the analysis procedure is streamed into this pipeline. To this end, data pipeline was formed using MQ, TSDB, and Kafka. MQTT, AQMP, and STOMP protocols are supported to transmit sensor data. Data are routed to each topic consisted by zone and equipment and then augmented with information parsed from registered tags. This augmented data is used in analyzing and monitoring process. The framework processes queries coming from API Server and handles streaming process. The sensor data stored in the TSDB can be monitored by the program such as Grafana or Prometheus in real time (Fig. 2).

2.4 Logging Engine

The framework uses various open-source software such as MQ, TSDB, Kafka, and RESTful library. This can make it difficult to trace the cause of the problem when it occurs in the framework. To check the data pipeline is normal or not, we designed a logging engine that can trace data processing at each stage. Framework is able to collect every log generated from overall stages including input, processing, output and request and responses of the API Server. The collected logs are routed to log exchange registered MQ, pushed to console, and stored into NoSQL. Since the routed

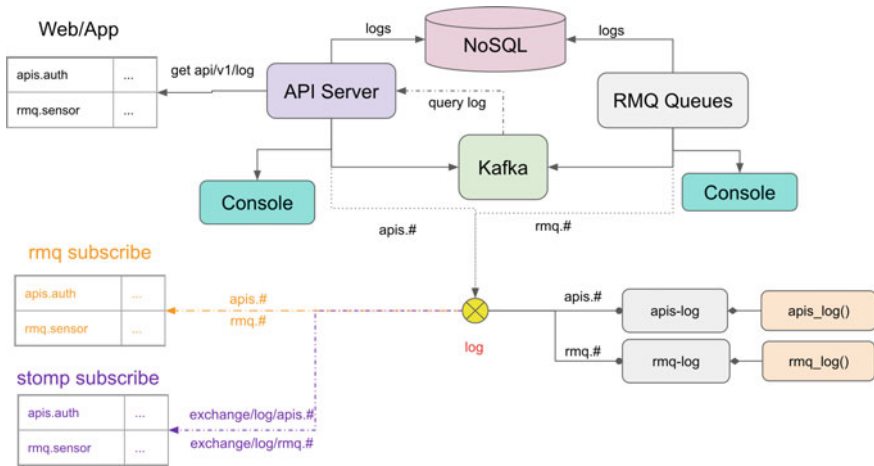


Fig. 3 Logging engine diagram

logs can be monitored with the web client in real time, it helps to trace the cause of the problem in the data pipeline of the framework (Fig. 3).

3 Visualizing Data of the Framework

The dashboard was designed to express the fire risk by zone and equipment. In order to express the fire hazard index by zone and equipment at once, dashboard displays the fire hazard index for each equipment corresponding to the turbine and boiler zone for each unit. ESS was expressed in the common zone. On the comprehensive screen, the fire hazard index was divided into 5 levels and the colors were displayed together for each equipment (Fig. 4).

When the relevant zone is selected on the comprehensive screen, dashboard changes into the selected zone. In the screen of the FHI by zone, since the real time and the time series HFI can be checked together corresponding to the equipment of the zone, it is possible to follow to each equipment state at a glance. User can check more detailed information for each equipment by selecting an equipment component from the screen. It is possible to check the volatility of the FHI and time series, and provide additional information (Fig. 5).

4 Conclusion

In this paper, we designed and developed the framework for predicting fire hazard for thermal power plants. The framework can be divided into three core functions. The



Fig. 4 Dashboard for comprehensive FHI



Fig. 5 Dashboard for zone FHI

first one is technology of acquiring, storing, and processing data related to complex fires. There are various types of data related to fires in thermal power plants; real time data such as sensor, location, and equipment status, information data describing the cause, location, scale of the ignition generated from the occurrence of a fire, and inspection data such as equipment log and video. We are focused real time data for developing dynamic fire protection system including time variable not for static system. The second one is technology for the real time streaming process and pipeline for predicting fire risk. Since refined data should be analyzed in real time, the framework generates a data pipeline so that the results can be monitored. A technology that can provide intuitive visualization to users by forming a real time data pipeline, reprocessing it into the required form for each analysis stage, and provide a fire hazard index for each zone or equipment. The third one is technology for sharing data to expend and support various system. Data sharing technology can be utilized in various services such as digital twin, VR-related training program development, visualization module, and safety manual development.

Acknowledgements This work was supported by Energy Technology Development of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry and Energy (MOTIE, Korea). [Project Name: Development of visualized fire protection system to Thermal Power Plant through IIoT and Digital Twin technology/Project Number: 20206610100060].

References

1. Polymeni S, Athanasakis E (2022) IoT-based prediction models in the environmental context: a systematic literature review. *Internet Things* 100612
2. Ahanger TA, Aljumah A (2022) State-of-the-art survey of artificial intelligent techniques for IoT security. *Comput Netw* 206:108771
3. Karunaratne SM, Angelopoulos CM (2020) A technological framework for data-driven IoT systems: application on landslide monitoring. *Comput Commun* 154:298–312
4. Zhang Y, Muthu BA (2021) Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. *Sustain Energy Technol Assess* 45:100986
5. Newman S (2019) *Monolith to microservices*. O'Reilly Media, Inc.
6. Bellemare A (2020) *Building event-driven microservices*. O'Reilly Media, Inc.; Liang X, Li Y, Wu X, Shen J (2018) Nonlinear modeling and inferential multi-model predictive control of a pulverizing system in a coal-fired power plant based on moving horizon estimation. *Energies* 589

Mortality Forecasting Using Data Augmentation



Kyusuh Park, Ji Su Park, Joosoo Maeng, and Jin Gon Shon

Abstract Recent mortality forecasting studies using artificial neural networks (ANNs) have shown improved forecasting performance compared with previous studies, where annual mortality rates were used. The use of annual mortality rates data leads to a problem in that the data are insufficient. Therefore, in this study, mortality rates were forecast by applying the related time-series data augmentation methods with an ANN, unlike in existing related studies. The experimental results showed that ANNs with augmented data have improved mortality forecasting performance compared to the case without data augmentation.

Keywords Data augmentation · Mortality forecasting · Artificial neural networks · Lee–Carter model · LSTM · Time-series data

1 Introduction

Artificial neural networks (ANNs) have been utilized in various fields, including mortality forecasting [1]. Mortality forecasting is used to predict changes in the future population structure, life expectancy, etc., to price financial instruments and/or to value related liabilities in the insurance and financial industries. In addition, this

K. Park

KS Accounting and Actuarial Consulting Co., Ltd., Seoul, Korea

J. S. Park

Department of Computer Science and Engineering, Jeonju University, Jeonju-si, South Korea

e-mail: jisupark@jj.ac.kr

J. Maeng

WillbeSoft Inc., Seoul, Korea

e-mail: js@wbsoft.kr

J. G. Shon (✉)

Department of Computer Science, Graduate School, Korea National Open University, Seoul, Korea

e-mail: jgshon@knou.ac.kr

methodology can be expanded and applied to various industrial products, including livestock and cultured fishery products.

In previous studies, annual mortality rates were used because of limitations due to the timing of collection and publication thereof [1, 2]. The use of such annual data leads to a problem in that the data are insufficient for the application of ANNs. For example, even if the mortality data period is 100 years, the amount of data is small because the number of mortality rates by sex for a specific age can be 100.

Therefore, this study aims to forecast mortality rates and observe any changes in forecasting performance, if any, by applying time-series data augmentation methods to annual data. Data augmentation is based on the existing domain knowledge and judgment in the field. The ANN was trained using augmented time-series data, unlike in previous studies.

2 Related Research

2.1 Mortality Forecasting Model and Artificial Neural Networks

Among the various mortality forecasting models, the Lee–Carter (LC) model used in this study has been widely used as a benchmark model in the field of mortality forecasting research [1]. The LC model is given as Eq. (1) [3].

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \quad (1)$$

where $m_{x,t}$ are the observed central mortality rates at age x in year t , a_x is the average age-specific mortality pattern, b_x is the sensitivity of each age to k_t , k_t is the mortality index describing the degree of improvement in mortality rates, and $\varepsilon_{x,t}$ is the residual term at age x and time t . Future mortality rates were calculated using Eq. (2) [3].

$$\hat{m}_{x,t} = \exp(\hat{a}_x + \hat{b}_x \hat{k}_t) \quad (2)$$

In recent mortality forecasting research, studies combining mortality forecasting models such as the LC model and deep learning techniques have been published [1, 4]. They showed that forecasting using the LC model and ANNs could obtain better performance results than applying ARIMA to the LC model [1]. In addition, one study showed improved results by applying long short-term memory (LSTM) to a three-factor model considering cohorts [2].

A recurrent neural network (RNN) can be used for time-series data. However, a vanishing gradient can occur in RNNs. An LSTM network was used to deal with the vanishing gradient problem. An LSTM unit consists of a cell, input gate, output gate, and forget gate. In addition, LSTM is operated through interactions [5–8].

Table 1 Training and testing data period

County	Total year	Training set years	Testing set years
Australia	1921–2019	1921–1999	2000–2019

2.2 Time-Series Data Augmentation

The idea of data augmentation is to generate a synthetic dataset for an unexplored input space while maintaining the correct labels. Such data augmentation has played an important role in the successful application of deep learning models and is effective in many areas [9].

However, time-series data augmentation has been relatively less studied compared to that in other fields, such as computer vision. When time-series data augmentation methods are applied, we must consider the properties thereof. Therefore, the data augmentation methods may vary depending on the task. Time-series data augmentation methods include random transformation, pattern mixing, generative models, and decomposition [9, 10].

One of the time-series data augmentation methods is the interpolation method used in this study [10, 11]. In addition, in mortality forecasting, monthly mortality rates can be estimated from annual mortality rates using the assumptions for fractional ages in the field of actuarial science. Typical assumptions are the uniform distribution of deaths, constant force of mortality, and Balducci assumption [12, 13].

3 Data Augmentation

3.1 Datasets for Data Augmentation

As shown in Table 1, we obtained the mortality rates of 55-year-old Australian males from 1921 to 2019 from the Human Mortality Database (HMD) [14]. These were divided into 80% training data and 20% testing data. Data from 0 to 95 years of age were used in this study. Data after the age of 96 years were excluded because they appeared to have an irregular mortality pattern owing to a lack of data.

3.2 Data Augmentation Method

In the experiments, we considered two types of data augmentation. First, the data of k_t were augmented monthly using Eq. (3) based on the linear interpolation method.

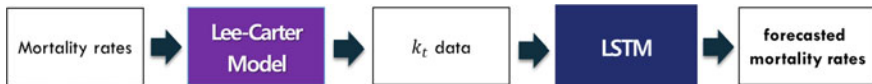


Fig. 1 Experimental process

$$k_{t+\frac{s}{n}} = k_t \frac{(n-s)}{n} + k_{t+1} \frac{s}{n} \tag{3}$$

$$(s = 1, 2, \dots, n-2, n-1)$$

Second, for the annual mortality rates, monthly mortality rates were calculated using Eq. (4), assuming the same mortality rate in the middle of the year, based on the assumptions for fractional ages and judgment.

$$m_{(x,t,\frac{s}{12})} = \frac{1}{12} m_{x,t} \tag{4}$$

$$(s = 1, 2, \dots, 12)$$

4 Experiments

4.1 Experimental Process and Environment

In this study, mortality rates were forecasted by applying the LC model to the mortality data and applying the LSTM to the k_t computed from the LC model, as shown in Fig. 1.

For the experiments, Windows 10, Google Colab, NVIDIA GPU Tesla P100, memory 13.6, Python 3.7, and TensorFlow 2.8.0 were used. The Adam optimizer and activation function of tanh were applied.

4.2 Experimental Results

Experiments were conducted according to whether the data were augmented or not and the types of data augmentation. In Experiment A, without data augmentation, the annual mortality rates were used as in previous studies. The mortality rates were then forecasted by applying the LC model and LSTM. In Experiment B, k_t , the time indices describing mortality trends calculated from the LC model, were augmented by the linear interpolation method of Eq. (3) to forecast mortality rates using LSTM. In Experiment C, the annual mortality rates were augmented by the application of the assumption for fractional ages in the field of actuarial science, as in Eq. (4), and

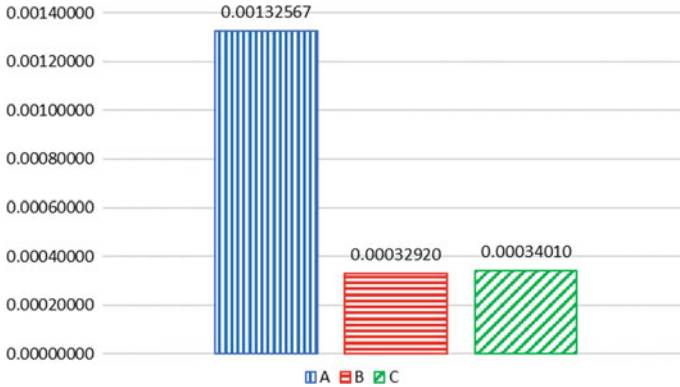


Fig. 2 MAE by experiment

mortality rates were forecasted using the LC model and LSTM. In these experiments, k_t was normalized. A time step of 10 was applied to LSTM. To calculate the goodness of fit measures in the experiments, the mean absolute error (MAE) and root mean square error (RMSE), as in Eqs. (5) and (6), were used.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\text{actual death rate}_i - \text{forecasted death rate}_i| \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{actual death rate}_i - \text{forecasted death rate}_i)^2}{N}} \tag{6}$$

Figure 2 shows the MAEs for Experiments A, B, and C. The MAEs of Experiments B and C with data augmentation were 0.00032920 and 0.00034010, respectively, which were smaller than that of Experiment A (0.00132567). The results indicated that the forecasting performance was improved by data augmentation.

The RMSEs for Experiments A, B, and C were 0.00140166, 0.00044770, and 0.00046879, respectively. The results of the RMSEs have also shown similar results to those of the MAEs.

Figure 3 shows the actual and forecasted mortality rates for the forecasted period in Experiments A, B, and C. This also shows that the results forecasted by augmenting the data are closer to the actual mortality rates than those without augmenting them.

5 Conclusion

Recent mortality forecasting studies using ANNs showed the possibility that the forecasting performance could be improved with ANNs compared to previous mortality forecasting studies. However, annual mortality rates have been used in

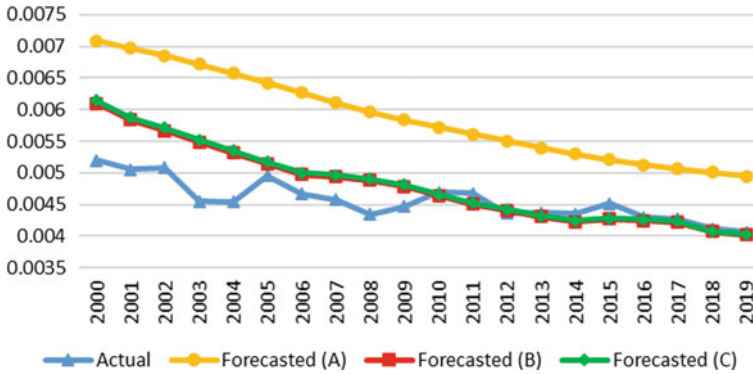


Fig. 3 Actual and forecasted mortality rates by Experiments A, B, and C

related mortality forecasting studies using ANNs. Therefore, we infer that there is a problem in which we do not have sufficient data in mortality forecasting compared to the quantity of data in other fields when applying ANNs.

In this study, we performed experiments to determine whether data augmentation can contribute to the improvement of mortality forecasting performance by augmenting relevant time-series data in mortality forecasting through ANNs. For the experiments, the linear interpolation method was applied to the data of k_t , the mortality index describing the degree of improvement in mortality rates. Furthermore, the annual mortality rates were augmented using the assumption of fractional ages and judgment. From the experiments, it was observed that the ANN trained with time-series data augmentation improved the mortality forecasting performance compared to that in the absence of data augmentation.

Finally, this study focused on whether the mortality forecasting performance would be improved by related time-series data augmentation. In the future, we may extend the research on forecasting performance by considering additional factors, such as augmented data types, data augmentation levels, and methods.

References

1. Nigri A, Levantesi S, Marino M, Scognamiglio S, Perla F (2019) A deep learning integrated Lee–Carter model. *Risks* 7(1):33
2. Odhiambo J, Weke P, Ngare P (2021) A deep learning integrated Cairns–Blake–Dowd (CBD) systematic mortality risk model. *J Risk Financ Manag* 14(6):259
3. Lee RD, Carter LR (1992) Modeling and forecasting US mortality. *J Am Stat Assoc* 87(419):659–671
4. Hainaut D (2018) A neural-network analyzer for mortality forecast. *ASTIN Bull* 48:481–508
5. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
6. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 5(2):157–166

7. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
8. Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM. *Artif Neural Netw* 2(470):850–855
9. Wen Q, Sun L, Yang F, Song X, Gao J, Wang X, Xu H (2002) Time series data augmentation for deep learning: a survey. Working paper. [arXiv:2002.12478v3](https://arxiv.org/abs/2002.12478v3). Accessed 26 Jan 2022
10. Iwana BK, Uchida S (2021) An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE* 16(7):1–32
11. Oh C, Han S, Jeong J (2020) Time-series data augmentation based on interpolation. *Procedia Comput Sci* 175:64–71
12. Bowers NL Jr, Gerber HU, Hickman JC, Jones DA, Nesbitt CJ (1997) *Actuarial mathematics*, 2nd edn. The Society of Actuaries
13. Jones BL, Mereu JA (2000) A family of fractional age assumptions. *Insur Math Econ* 27(2):261–276
14. Human mortality database. Available <http://www.mortality.org>. Accessed 25 Mar 2022

Spatial–Temporal Reservation-Based Multi-vehicle Cooperative Speed Guidance



Changhao Piao, Jia He, and Yongsheng Wang

Abstract A multi-vehicle collaborative speed guidance method is proposed in this paper. Without considering the time and road space occupied by the vehicle ahead, the original signal status information is taken as input during speed planning. The calculated speed will cause multiple vehicles to arrive at the same desired position at the same moment, resulting in vehicles being forced to slow down and stop to avoid a collision. By building the vehicle arrival time model and safe distance model, the moment and recommended speed of each vehicle through the intersection are calculated. The SUMO + Traci simulation platform is built, and experiments show that the proposed method enables vehicles to pass through the intersection without stopping in low, medium and high traffic flow scenarios, while the average vehicle travel time is significantly reduced, reaching over 11% in both medium and high traffic flow scenarios, compared to the vehicle speed guidance method without the above models.

Keywords Spatial–temporal reservation · Multi-vehicle collaboration · Safe distance between cars · Speed guidance

1 Introduction

In recent years, with the impetus of wireless communication technology and computer technology, V2X technology has emerged [1]. In the context of vehicle networking, vehicles rely on high-speed information interaction with the outside world to plan their driving speed, taking into account factors such as signal lights and the state of the vehicle itself.

C. Piao (✉) · J. He
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: piaoch@cqupt.edu.cn

Y. Wang
Tsinghua University, Beijing 100084, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_13

105

Tang et al. [2] introduced a vehicle-following model into a speed-guiding strategy on a single-lane road at a multi-signal intersection to investigate driving behaviours and fuel consumption. Han et al. [3] proposed an optimal control-based vehicle speed guidance strategy to improve traffic safety and efficiency against freeway jam waves. The development of Collaborative Vehicle Infrastructure System (CVIS) has spawned many new strategies, which can solve the congestion wave by implementing micro-vehicle tracking control actions on a single vehicle. Ding et al. [4] Traffic guidance–perimeter control coupled method for the congestion in a macro network. References [5, 6] consider vehicle queues and traffic signal states on arterial routes and propose a multi-stage control model to obtain the optimal vehicle trajectory. Wang et al. [5] proposed a time-loop technique to identify the head vehicle of a convoy and send the suggested acceleration at each time step to optimise the delay over the whole trajectory. Ma et al. [6] used acceleration–deceleration curves as the optimisation objective and proposed a fully and imperfectly obedient speed-guided vehicle queue at signalised intersections energy-efficient driving speed control algorithm, but this speed control algorithm is for convoys travelling at the same speed and headway.

The deficiencies of the current research are mainly manifested as follows: the speed guidance model only considers a single vehicle, and the speed guidance scenarios are roughly delineated. Therefore, this paper integrates the above shortcomings and aims to reduce the average travel time of each vehicle and avoid the phenomenon of forcing vehicles to decelerate or stop. The allowable passing duration is introduced, and the occupation of road space by vehicles is also considered. The spatial–temporal reservation model is established by the above two ideas. The model is applicable at low, medium and high traffic flows.

2 Algorithm Design-Related Work

2.1 Passage Scenario Division

According to the real-time speed v_0 and the green light period $[t_1, t_2]$ when the vehicle arrives at the guidance area, the passage strategy is divided into eight scenarios.

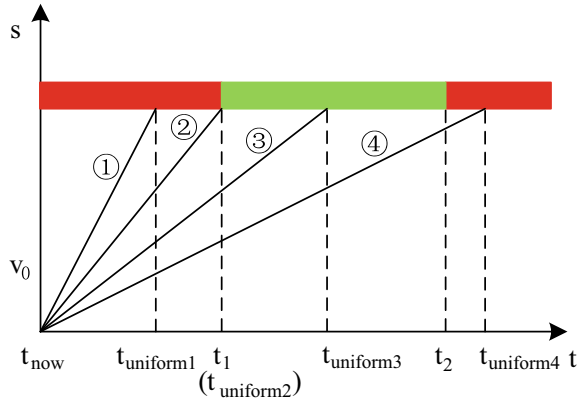
Time spent for the vehicle travelling from the start point of the speed guidance area to the intersection at constant speed:

$$t_{\text{uniform}} = \frac{S}{v_0}, \quad (1)$$

where S is the length of the speed guidance area, and v_0 is the speed of the vehicle when it reaches the guidance area.

As shown in Fig. 1, when the vehicle reaches the guidance area, if it continues to travel at a constant speed, four scenarios will occur when it reaches the intersection.

Fig. 1 Diagram of a vehicle travelling to the intersection at a constant speed



- (1) $t_{uniform} < t_1$: The vehicle arrives at the intersection early. Consideration needs to be given as to whether it can slow down to pass through the intersection. Time spent for the vehicle to slow down to the minimum road speed limit.

$$t_{dec} = \frac{v_0 - v_{min}}{a} \tag{2}$$

If $t_{dec} \geq t_1$ and $v_0(t_1 - t_{now}) + \frac{1}{2}a(t_1 - t_{now})^2 \leq S$ are satisfied, the vehicle passes through the intersection with uniform deceleration, as shown in Fig. 2a.

If $t_{dec} < t_1$ and $\frac{v_{min}^2 - v_0^2}{2a} + v_{min}(t_1 - t_{now} - \frac{v_{min} - v_0}{a}) \leq S$ are satisfied, the vehicle decelerates to v_{out} before maintaining a constant speed through the intersection, as shown in Fig. 2b.

- (2) $t_{uniform} = t_1$: The vehicle is kept at an uniform speed through the intersection.
- (3) $t_1 < t_{uniform} \leq t_2$: The vehicle is travelling at a constant speed and can reach the intersection at the period of the green light. Consider whether it can accelerate pass through the intersection.

Time spent for the vehicle to accelerate to the maximum road speed limit:

$$t_{acc} = \frac{v_{max} - v_0}{a} \tag{3}$$

If $t_{acc} \geq t_1$ and $v_0(t_1 - t_{now}) + \frac{1}{2}a(t_1 - t_{now})^2 \leq S$ are satisfied, the vehicle passes through the intersection with uniform acceleration, as shown in Fig. 2c.

If $t_{acc} < t_1$ and $\frac{v_{max}^2 - v_0^2}{2a} + v_{min}(t_1 - t_{now} - \frac{v_{max} - v_0}{a}) \leq S$ are satisfied, then the vehicle accelerates to v_{out} before maintaining uniform speed to pass through the intersection, as shown in Fig. 2d.

- (4) $t_{uniform} > t_2$: When the vehicle arrives at the intersection, the permitted passing period has ended. Consider whether it can speed up to pass through the intersection or slow down to pass through the intersection at the next permitted passing time.

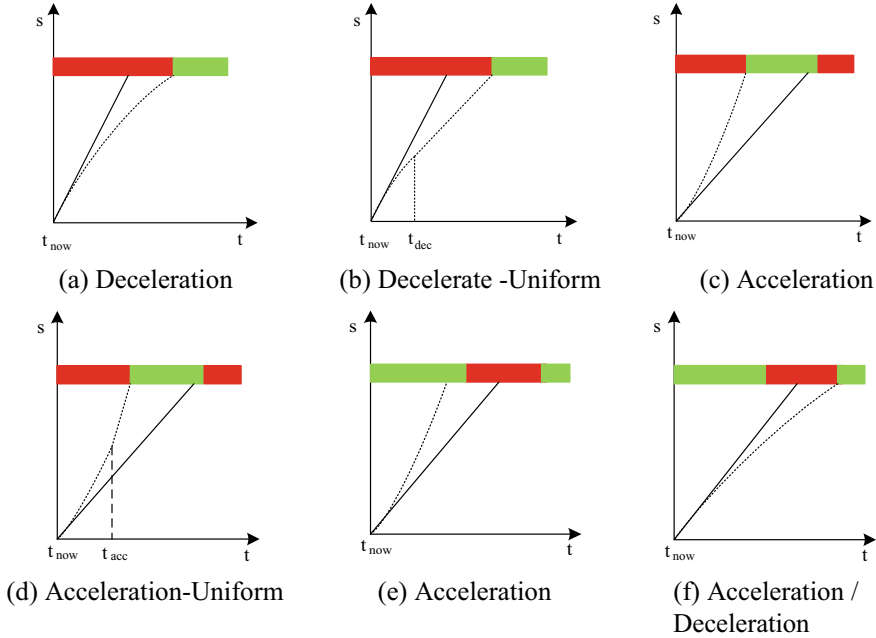


Fig. 2 Diagram of the passage strategy

If $\frac{v_{\max}^2 - v_0^2}{2a} + v_{\max}(t_2 - t_{\text{now}} - \frac{v_{\max} - v_0}{a}) > S$ is satisfied, the vehicle can pass during the current green light and the travel time diagram is shown in Fig. 2e.

If $\frac{v_{\min}^2 - v_0^2}{2a} + v_{\min}(t_2 + T_R - t_{\text{now}} - \frac{v_{\min} - v_0}{a}) \leq S$ is satisfied, the vehicle can pass through the intersection during the next cycle of green light as shown in Fig. 2f.

It is important to divide the situation of crossing as above to lay a foundation for subsequent calculation.

3 Spatial–Temporal Reservation

It is important to consider not only the possible time intersection of two vehicles in terms of the permissible passing period, but also the planning of vehicle speeds with safe vehicle spacing.

3.1 Temporal Reservation Model

Speed guidance with a single vehicle as the object of study, where each vehicle is guided according to the traffic conditions at its disposal, without taking into account

the occupation of road space–time resources by the vehicle ahead, can lead to the following three situations:

- (1) The phenomenon of a vehicle stopping briefly when the green light has just started. As shown in Fig. 3a, vehicle A enters the guidance area and calculates the corresponding guidance speed based on the current green light time; however, a queue of vehicles has been already formed, resulting in vehicle A being forced to stop.
- (2) The phenomenon of vehicles slowing down briefly when the speed of the rear vehicle is too large. As shown in Fig. 3b, both vehicles want to pass through the intersection at the beginning of the green light, and vehicle A’s calculated guiding speed is greater than vehicle B’s. At some position, the distance between the two vehicles will be less than the safe distance, resulting in vehicle A being forced to slow down.
- (3) The phenomenon of vehicles not passing through the intersection when the green light is almost over. As shown in Fig. 3c, because vehicles B, C and D have occupied the green light period of the cycle, vehicle A is driving to the intersection when the green light period has ended. As a result, vehicle A is forced to stop and wait.

To avoid the competition between two vehicles for the same space at a same moment leading to the failure of the speed guidance algorithm, the permissible passing period $[t_{start}, t_{end}]$ should follow the following conditions:

- (1) If there is no vehicle in front, the time when the current signal cycle is green will be used as the permitted period for vehicles to pass through the intersection.
When the current signal is green:

$$t_{start} = t_{now}, \tag{4}$$

$$t_{end} = t_{next}, \tag{5}$$

where t_{now} is the current moment, and t_{next} is the current end-of-phase moment.
When the current signal is red:

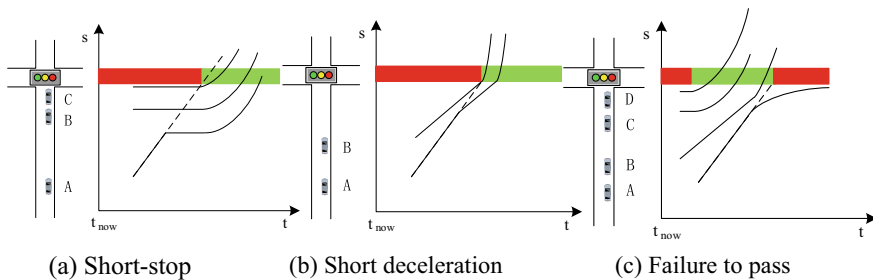


Fig. 3 Diagram of traffic speed influenced by the vehicle in front

$$t_{\text{start}} = t_{\text{next}}, \quad (6)$$

$$t_{\text{end}} = t_{\text{next}} + T_G, \quad (7)$$

where T_G is the duration of the green light.

- (2) If there are vehicles ahead of the road that have not yet passed through the intersection but have not formed a vehicle queue, the moment when the previous vehicle passed through the intersection is used as an input to the lower bound t_{start} for the time allowed for vehicles to pass through the intersection:

$$t_{\text{start}} = t_{\text{last_cross}}. \quad (8)$$

The higher bound on the allowed passing period is the end of the green light for the signal cycle in which $t_{\text{last_cross}}$ is located.

$$t_{\text{end}} = (t_{\text{last_cross}} | (T_G + T_R + T_Y)) * (T_G + T_R + T_Y) + T_G. \quad (9)$$

- (3) If a queue of vehicles has formed ahead of the intersection, the queue dissipation time needs to be calculated.

$$t_{\text{start}} = \sqrt{\frac{2 * [(n - 1) * (l_{\text{vehicle}} + d) + l_{\text{vehicle}}]}{a}} + t_w(n - 1), \quad (10)$$

$$t_{\text{end}} = (t_{\text{start}} | (T_G + T_R + T_Y)) * (T_G + T_R + T_Y) + T_G,$$

where n is the number of vehicles stopping at the intersection, l_{vehicle} is the length of the vehicle, d is the distance between two vehicles when vehicles stop and a is the acceleration, 2 m/s^2 .

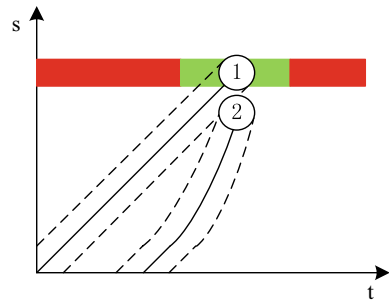
3.2 Spatial Reservation Model

As shown in Fig. 4, at the same moment, if the distance between two vehicles is less than the safe distance, the occupation time of the same space on the road is crossed, which will lead to the failure of speed guidance. To keep the distance between the two vehicles is always greater than the safe distance, the vehicle driving speed strategy is set to variable speed first and then uniform speed. Under this condition, we only need to ensure that at the moment of vehicle 1 travels to the intersection, the driving distance of vehicle 2 is less than the difference of the length of the speed guidance area and the safe distance of two vehicles, can avoid the space–time conflict.

Safe distance:

$$S_{\text{safe}} = 1.2 * v_{\text{host}} + l_{\text{vehicle}} \quad (11)$$

Fig. 4 Schematic diagram of vehicle space location conflicts



where v_{host} is the own speed, and l_{vehicle} is the length of the vehicle.

Maximum distance travelled by the rear vehicle at the moment $t_{\text{lastcross}}$ is:

$$S_{\text{host}} = S - S_{\text{safe}}. \quad (12)$$

3.3 Model Solutions

Solving the model subjected to time–space limited conditions to get recommended speed and passing moment.

(1) Acceleration scenario, the vehicle travels at a speed that satisfies:

$$\frac{v_{\text{out}}^2 - v_0^2}{2a} = S, \quad (13)$$

$$v_0(t_{\text{lastcross}} - t_{\text{now}}) + \frac{1}{2}a(t_{\text{lastcross}} - t_{\text{now}})^2 < S - (1.2v_{\text{out}} + l_{\text{vehicle}}). \quad (14)$$

Solution of the above equation:

$$v_{\text{out}} = v_0 + a * (t_{\text{start}} - t_{\text{now}}), \quad (15)$$

$$t_{\text{cross}} = t_{\text{start}}. \quad (16)$$

(2) Acceleration—uniform scenario, the vehicle travels at a speed that satisfies:

$$\frac{v_{\text{out}}^2 - v_0^2}{2a} + v_{\text{out}} \left(t_{\text{start}} - \frac{v_{\text{out}} - v_0}{a} \right) = S, \quad (17)$$

$$\frac{v_{\text{out}}^2 - v_0^2}{2a} + v_{\text{out}}(t_{\text{lastcross}} - t_{\text{now}}) < S - (1.2 * v_{\text{out}} + l_{\text{vehicle}}). \quad (18)$$

Solution of the above equation:

$$v_{\text{out}} = v_0 + a(t_{\text{start}} - t_{\text{now}}) - \sqrt{(v_0 + a(t_{\text{start}} - t_{\text{now}}))^2 - (v_0^2 + 2aS)} \quad (19)$$

$$t_{\text{cross}} = t_{\text{start}}.$$

(3) Uniform scenario, the vehicle travels at a speed that satisfies:

$$v_0 * (t_{\text{start}} - t_{\text{now}}) < S - (1.2 * v_0 + l_{\text{vehicle}}). \quad (20)$$

Output speed is equal to v_0 :

$$v_{\text{out}} = v_0, \quad (21)$$

$$t_{\text{cross}} = t_{\text{now}} + \frac{S}{v_0}. \quad (22)$$

(4) Deceleration scenario. The vehicle travels at a speed that satisfies:

$$\frac{v_{\text{out}}^2 - v_0^2}{2a} = S \quad (23)$$

$$v_0 * (t_{\text{last_cross}} - t_{\text{now}}) + \frac{1}{2} * a * (t_{\text{last_cross}} - t_{\text{now}})^2 < S - (1.2 * v_{\text{out}} + l_{\text{vehicle}}).$$

Solution of the above equation:

$$v_{\text{out}} = v_0 + a * (t_{\text{start}} - t_{\text{now}}), \quad (24)$$

$$t_{\text{cross}} = t_{\text{start}}. \quad (25)$$

(5) Deceleration—uniform scenario. The vehicle travels at a speed that satisfies:

$$\frac{v_0^2 - v_{\text{dec}}^2}{2a} + v_{\text{out}} \left(t_{\text{start}} - \frac{v_{\text{out}} - v_0}{a} \right) = S, \quad (26)$$

$$\frac{v_{\text{out}}^2 - v_0^2}{2a} + v_{\text{out}}(t_{\text{last_cross}} - t_{\text{now}}) < S - (1.2v_{\text{out}} + l_{\text{vehicle}}). \quad (27)$$

Solution of the above equation:

$$v_{\text{out}} = v_0 - a(t_{\text{start}} - t_{\text{now}}) + \sqrt{(v_0 - a(t_{\text{start}} - t_{\text{now}}))^2 - (v_0^2 - 2aS)}, \quad (28)$$

$$t_{\text{cross}} = t_{\text{start}}.$$

(6) Slow down and stop scenario. The vehicle is travelling at the minimum speed:

$$v_{\text{out}} = v_{\text{min}}, \quad (29)$$

$$t_{\text{cross}} = (t_{\text{lastcross}} | (T_G + T_R + T_Y) + 1) * (T_G + T_R + T_Y) + \sqrt{\frac{2 * (n - 1) * (l_{\text{vehicle}} + d)}{a}}. \quad (30)$$

4 Experiment

4.1 Experimental Condition Settings

The opposite six-lane single-point intersection was modelled in SUMO, and the middle straight lane of the northern inlet of the intersection was selected for the study. The total length of the road is 500 m and the speed guidance area is from 250 m upstream of the stop line to 0 m. The intersection signal control scheme is a fixed duration, with a period of 50 s, including 27 s for green, 20 s for red and 3 s for yellow. The simulation time is set to 500 s, with a simulation step size of 0.01 s, and the position of each vehicle in the lane is recorded every 0.01 s.

4.2 Comparison of Vehicle Travel Maps

As shown in Fig. 5, compared to single-vehicle guidance, the spatial–temporal reservation-based multi-vehicle collaborative speed guidance model can effectively solve the phenomenon of vehicles being forced to slow down and stop. Moreover, the effectiveness of this model is not affected by intersection saturation and is applicable to different traffic flows.

4.3 Analysis of Average Vehicle Travel Time

Compared to single-vehicle guidance, multi-vehicle guidance reduces average journey time of each vehicles in low, medium and high traffic flow scenarios and is particularly effective in medium and high traffic flow scenarios, where the time is reduced by more than 11%. As the traffic flow increases, the average vehicle stopping

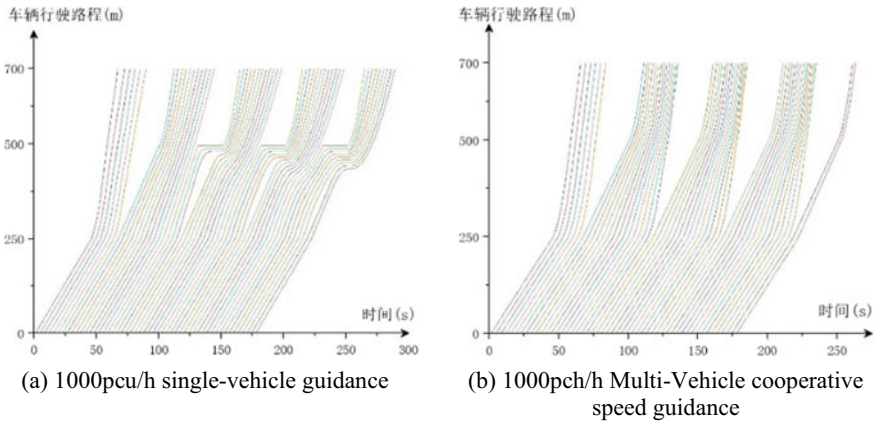


Fig. 5 Spatial and temporal diagrams of different traffic flows

Table 1 Difference of two different guidance methods

Traffic flow (pcu/h)	Multi-vehicle cooperative speed guidance		Single-vehicle guidance		Reduction of average travel time (%)
	Average travel time (s)	Average parking time (s)	Average travel time (s)	Average parking time (s)	
700	66.59	0.00	68.90	0.00	3.35
800	66.63	0.00	72.61	1.42	8.24
900	68.59	0.00	77.18	1.65	11.13
1000	68.16	0.00	82.85	2.56	17.46
1100	79.88	0.00	90.55	3.13	11.78
1200	83.31	0.00	96.94	3.58	14.06

time under the single vehicle guidance method gradually increases, while the spatial-temporal reservation-based multi-vehicle cooperative speed guidance method allows vehicles to pass through the intersection without stopping in low, medium and high traffic flow scenarios due to the inclusion of allowable passing period and space occupation model (Table 1).

5 Conclusion

Compared with the single-vehicle guidance model, the spatial-temporal reservation-based multi-vehicle cooperative speed guidance model can avoid the occurrence of forced vehicle slowdown and stopping and significantly reduce the average vehicle

travel time while ensuring safe vehicle spacing. The spatial–temporal reservation-based multi-vehicle collaborative speed guidance model is effective in low, medium and high traffic flow scenarios, and the model is highly robust.

References

1. Habibovic A, Chen L (2021) Connected automated vehicles: technologies, developments, and trends. In: International encyclopedia of transportation. Elsevier, pp 180–188
2. Tang T-Q, Yi Z-Y, Zhang J, Wang T, Leng J-Q (2018) A speed guidance strategy for multiple signalized intersections based on car-following model. *Phys A Stat Mech Appl* 399–409
3. Han Y, Yu H, Li Z, Xu C, Ji Y, Liu P (2022) An optimal control-based vehicle speed guidance strategy to improve traffic safety and efficiency against freeway jam waves. *Accid Anal Prev*
4. Ding H, Guo F, Zheng X, Zhang W (2022) Traffic guidance–perimeter control coupled method for the congestion in a macro network. *Transp Res Part C Emerg Technol*
5. Wang X, Liu M, Ci Y, Yang Y (2017) Effectiveness of driver’s bounded rationality and speed guidance on fuel-saving and emissions-reducing at a signalized intersection. *J Clean Prod*
6. Ma F, Yang Y, Wang J, Li X, Wu G, Zhao Y, Wu L, Aksun-Guvenc B, Guvenc L (2021) Eco-driving-based cooperative adaptive cruise control of connected vehicles platoon at signalized intersections. *Transp Res Part D Transp Environ*

A Research on Quantum Repeater Platform Using Drone: Utilization Methods and Considerations



Tae Woo Kim, Seo Yeon Moon, Il Chul Shin, and Jin Sang Park

Abstract Communication technology has evolved into quantum communication with improved security and speed by utilizing quantum information technology. However, it is difficult to use it in special situations such as rugged mountainous terrain and the state of war. In order to configure a wireless quantum communication environment in an environment where it is difficult to utilize wired quantum communication, a plan to effectively utilize quantum repeater is needed. In this paper, we discuss how to utilize quantum repeater platform using drones and the considerations of drone for effective quantum communication. Finally, we propose double-layer-based drone prototype model for quantum repeater platforms. The proposed model enables secure quantum communication in rugged terrain. This research will actually produce drones and conduct quantum communication using quantum repeater platforms in the future.

Keywords Octocopter drone · Double-layer drone · Quantum repeater platform

1 Introduction

With the spread of mobile communication in the late twentieth century, communication technology developed rapidly. Currently, communication technology has

T. W. Kim · S. Y. Moon · I. C. Shin · J. S. Park (✉)
Gangwon Technopark, Gangwon SW Industry Promotion Center, 24220 Chuncheon-si,
Gangwon-do, Republic of Korea
e-mail: pjs_1129@gwtp.or.kr

T. W. Kim
e-mail: tang_kim@gwtp.or.kr

S. Y. Moon
e-mail: seoyeon@gwtp.or.kr

I. C. Shin
e-mail: shin172@gwtp.or.kr

evolved beyond optical communication using optical fiber to quantum communication using quantum information technology, which is a next-generation technology. Quantum communication is a communication method using quantum entanglement phenomena and has excellent speed and security [1]. Various methods such as optical fiber-based quantum communication and optical photomagnetic-based quantum communication are being utilized for quantum communication. Quantum wireless communication is difficult to provide smooth communication because the beam becomes wider when the transmission exceeds a certain distance, making it difficult for the receiving point to receive correctly [2]. Quantum repeater is required to configure a seamless wireless quantum communication environment in various situations, and methods for activating and maintaining repeater are needed.

In this paper, we present a method for enabling quantum repeater platform using drones and discuss considerations for effective quantum communication based on drones. We then propose a double-layer structure-based drone prototype model for the quantum repeater platform, which can accommodate considerations. The proposed model uses a double-layer structure to provide larger rotors than typical octocopter drones. This increases the payload of the drone and enables stable operation in weather conditions such as strong winds and rain. It also uses a hybrid battery module with drone batteries to enable long flights. The key components, such as rotor configuration and hybrid battery module, will be described later. In the conclusion, we introduce the discussion of research and future research.

2 Related Work

This chapter introduces various researches and relay stations for wireless quantum communication. The major researches are how to use satellites and drones. Chu et al. [3] conducted a feasibility study to deploy quantum key distribution (QKD) in high altitude platforms (HAP) as a way to secure future communication applications and services. This study evaluated the link ability to deliver QKD in stratospheric HAPs flying at an altitude of 20 km. Sidhu et al. [4] reviewed space-based quantum technology communication for wireless quantum communication overcoming the criminal limitations of terrestrial networks and presented a roadmap for development directions. Liorni et al. [5] proposed a scheme to achieve entanglement distributions in space using quantum repeaters and satellite-based links. The proposed scheme uses quantum iterators based on quantum error correction.

As a study on wireless quantum communication using aerial relays, research using satellite communication currently being used is underway. Quantum communication using satellites is considered a next-generation technology that will enable fast data communication. However, in the case of quantum communication using satellites, long quantum wireless communication should be provided. It is a technology that is still difficult to use because of its very low stability and astronomical cost for testing.

Researchers studied wireless quantum communication using drones as a way to reduce the long-range, astronomical cost problem of quantum satellite communication. Since the quantum repeater platform is constructed using miniaturized drones, the cost of environmental construction is relatively low with satellite communication, and various studies can be conducted. Conrad et al. [6] proposed quantum QKD communication between drones using a single photon detector, an FPGA-based time tagger, and a novel statistical post-processing synchronization algorithm. The proposed paper analyzed the conditions for quantum communication between drones, such as drone weight and power conditions during construction. Kumar et al. [7] proposed the Internet of Quantum Drones (IoQDs) using quantum drones. In the proposed paper, the district applying IoQDs presented economic and social problems and reviewed the main problems, risks, and advantages of technology practical use from a future-oriented perspective. Liu [8] proposed a photo-repeater platform using drones. The proposed platform reconstructed the wave front of the photon to minimize the diffraction loss in free-space transmission and proceeded with quantum entanglement distribution at a distance of 1 km using two drones acting as relay nodes.

3 Utilization Methods and Considerations

Quantum repeater platforms utilized in wireless quantum communication can distribute and transmit quantum signal entanglement between ground stations and drones using quantum entanglement light sources and quantum signal tracking modules. As in the previous studies, various studies have been conducted using flying objects for safe wireless quantum communication over long distances.

In order to effectively maintain the network in rugged terrain, war, and wide seas such as the Pacific Ocean, the quantum repeater platform can be installed on a small unmanned aerial vehicle as shown in Fig. 1 to be used as a communication relay station. The following requirements are required for the quantum repeater platform using the unmanned aerial vehicle as introduced.

Stability: Quantum wireless communication uses ultra-precise beam tracking devices to track quantum signals and transmit quantum entanglement light sources. Currently fixed ground-to-air quantum signal transmission has advanced to some commercially available stages. However, in drone-to-drone communication, it is difficult due to the great influence of weather environments such as wind and rain because it is hovering in the air. Therefore, it is essential to develop a drone to minimize vibration by minimizing the influence of the weather climate.

Payload Capacity: The quantum repeater platform is a high-weight device because it consists of various state-of-the-art equipment modules, including small quantum entanglement light sources, ultra-precise beam tracking devices, and quantum entanglement distribution devices, which are key elements. For high-weight transportation, the size of the drone will gradually increase, and the power of the drone must also

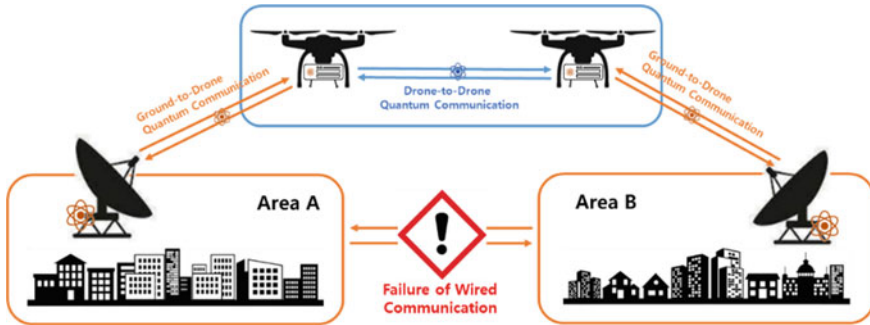


Fig. 1 Drone-based wireless quantum communication overview

increase. Currently, quantum relay platforms are about 25–40 kg and require drones weighing more than 130 kg to fly drones.

Hovering Time: Drone needs long hours of hovering to act as a radio relay station. As discussed above, there shall be an innovation in the battery to hovering high-weight drones for a long time. Currently, most drones use lithium battery batteries. For the long-term use of batteries, various materials such as hydrogen, lithium ions, and potassium ions are being studied, and they are more than three times more efficient than existing batteries.

4 Proposed Drone Prototype Model

This chapter proposes a drone prototype model using the sliding parallel rotor method based on the requirements of the drone-based quantum repeater platform derived from the battlefield. First, we explain the conceptual diagram of the proposed model and then explain the key elements, rotor configuration, and hybrid battery module.

4.1 Design Overview

The drone for the proposed quantum repeater platform has an octocopter shape suitable for high-weight payload transport and precision work, as shown in Fig. 2. The rotor of the proposed drone uses the sliding parallel rotor scheme, which allows for improved aerodynamic performance. The hovering time of a drone was increased using a hybrid battery module that uses both a hydrogen cell and a lithium battery as a battery. In addition, it provides gas stability by arranging the hybrid battery module and the quantum repeater platform according to the payload weight. The key elements of the proposed prototype are introduced in detail in the following sections.

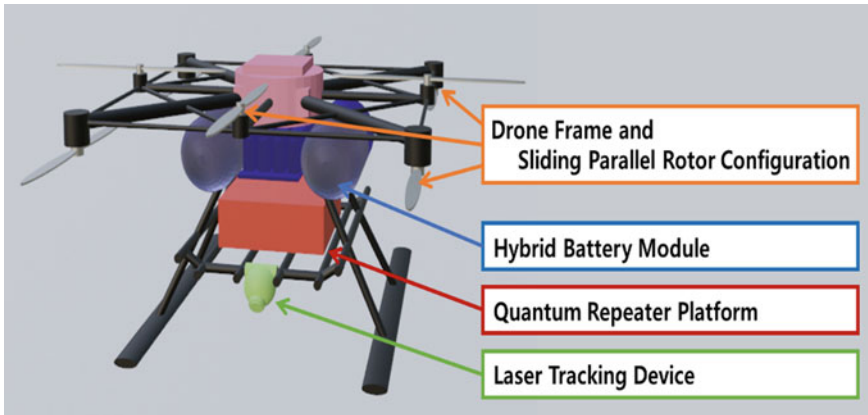


Fig. 2 Overview of proposed drone prototype model

4.2 Rotor Configuration

The rotor arrangement of octocopter-type drones can be largely divided into biaxial, coaxial, and sliding parallel as shown in Fig. 3 [9]. Figure 3d–f shows the aerodynamic interference according to the rotor arrangement of the octocopter drone. (d) In the biaxial rotor configuration, the rotor is horizontally placed, resulting in a decrease in the rotor thrust due to interference between rotors. (e) In the coaxial rotor configuration, the rotor interference is the highest as the air passing through the rotor on the second floor enters the rotor on the first floor. (f) In the sliding parallel rotor configuration, the second and first floor rotors are arranged diagonally, so there is less thrust reduction than the coaxial arrangement.

Octocopter drones for the proposed quantum repeater platform use the sliding parallel rotor configuration to load high-weight payloads. Since the sliding parallel rotor configuration uses a double-layer structure, it is possible to increase the size of the rotor and improve the power of the gas. In addition, it is possible to minimize shaking and thrust reduction due to aerodynamic interference, so it has excellent stability.

4.3 Rotor Configuration

Currently, most commercially available drones use lithium batteries. Table 1 is the result of a survey of drones released in the “2022 Unmanned Mobile Industry Expo” conducted in Korea in Aug. 2022 [10]. The investigated drone is a drone that transports high-weight payloads or is manufactured for precise work and has an octocopter shape. Among the several drones, drones using lithium batteries could fly for 30 min on a flat basis, but drones using hydrogen fuel cells can fly for more than 1 h.

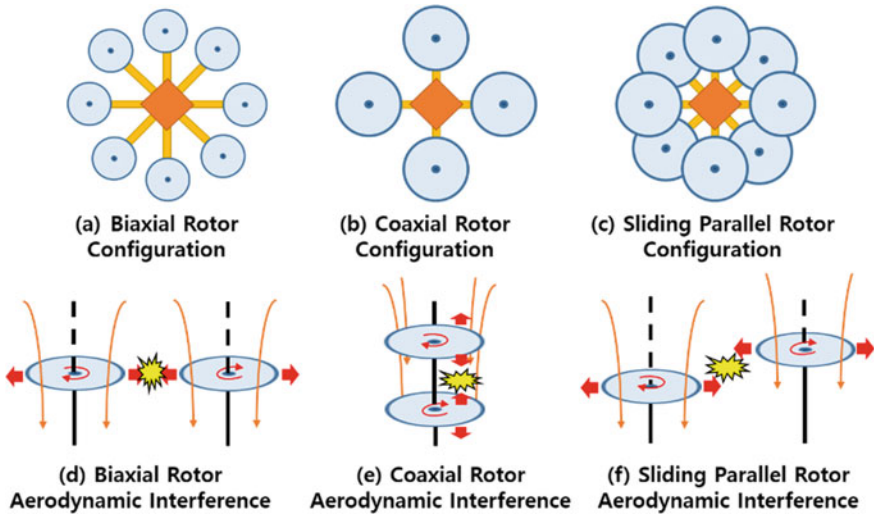


Fig. 3 Configuration according to rotor arrangement for octocopter [9]

Table 1 Resources of commercially available octocopter drones [10]

Model name	Weight of drone (kg)	Payload of drone (kg)	Battery type	Hovering time
ND-840Z	67	23	Lithium	30 min
XD-ISD4	47	10	Lithium	30 min
X-CARRIER	56	30	Lithium	30 min
Korean Navy drones	29	20	Hydrogen	1 h 30 min

The drone for the proposed quantum repeater platform uses a hybrid battery module that uses lithium battery and hydrogen fuel cell together. The proposed hybrid battery module utilizes a lithium battery in high-power activities such as takeoff and landing and uses a hydrogen fuel cell in low-power activities such as air hole and left-right movement. In addition, during low-power activities, spare energy can be utilized to charge a lithium battery to store some of the wasted energy. A charged lithium battery is capable of exerting strong power in emergency situations, such as rapid wind and fast start.

5 Conclusions and Future Work

In this paper, we propose a double-layer structure-based drone prototype model for the quantum repeater platform. The proposed model consists of a sliding parallel rotor configuration suitable for high-weight payload transport and precision work.

The proposed model uses a double-layer structure to increase the size of the rotor compared to a typical octocopter drone, and as a result, the power of the drone air is improved, resulting in an increase in payload loading. In addition, the sliding parallel arrangement was used to provide a reduction in shaking and thrust due to aerodynamic interference, thereby improving stability. In the battery part, hovering time was greatly improved by using a hybrid battery module in which a lithium battery and a hydrogen fuel cell were combined. We are planning to conduct simulations by specifying the drone design as a future study. Additionally, based on the simulation results, we produce a payload drone of 30 kg or more. The manufactured drone loads a quantum repeater platform to conduct an actual wireless quantum communication test and specify the utilization plan.

Acknowledgements This work was supported by Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-01014, Technology development of lightweight free-space quantum repeating platform).

References

1. Chen J (2021) Review on quantum communication and quantum computation. J Phys Conf Ser 022008. IOP Publishing
2. Mahdy A, Deogun JS (2024) Wireless optical communications: a survey. In: IEEE wireless communications and networking conference, vol 4. IEEE, pp 2399–2404
3. Chu Y et al (2021) Feasibility of quantum key distribution from high altitude platforms. Quant Sci Technol 6(3):035009
4. Sidhu JS et al (2021) Advances in space quantum communications. IET Quant Commun 2(4):182–217
5. Liorni C et al (2021) Quantum repeaters in space. New J Phys 23(5):053021
6. Conrad A et al (2021) Drone-based quantum key distribution (QKD). In: Free-space laser communications XXXIII, vol 11678. SPIE
7. Kumar A et al (2022) Futuristic view of the internet of quantum drones: review, challenges and research agenda. Veh Commun 100487
8. Liu HY (2021) Optical-relayed entanglement distribution using drones as mobile nodes. Phys Rev Lett 126(2):020503
9. Otsuka H, Nagatani K (2016) Thrust loss saving design of overlapping rotor arrangement on small multirotor unmanned aerial vehicles. In: 2016 ICRA. IEEE, pp 3242–3248
10. Korea Ministry of Science and ICT (2022) 2022 unmanned system world congress, Aug 2022

Implementation of Virtual Sensor for Semiconductor Process Verification Using Machine Learning



Jeong Il Shin, Ji Su Park, and Jin Gon Shon

Abstract The facility that performs the exposure process during semiconductor process verification includes more than 2500 sensors, and if an abnormal value is detected in any of these sensors, the wafer is treated as an exception. Wafers treated as exceptions undergo an additional process of determining wherein it is determined through physical measurement whether they are normal or abnormal. In this study, to reduce the time and cost of physical measurement, machine learning is used to learn sensor values to predict measurement results and to implement virtual sensors based on predicted values. The method for determining whether an exposure process is normal or abnormal is suggested using the implemented virtual sensor. The virtual sensor implements a single virtual sensor and multiple virtual sensors, and the algorithms used are extreme gradient boosting (XGBoost) and multi-output classifier. As a result of the experiment, the single virtual sensor detected an average of 93% of defects, furthermore, when measuring the detected defective wafers, an average of 99% of the actual defects were detected. Multiple virtual sensors also exhibited an average performance of 93%. Based on the findings of this study, 2500 real sensors were implemented as 250 virtual sensors, enabling the reduction of the time required to verify semiconductor exposure process results.

Keywords XGBoost · Multi-output classifier · Semiconductor

J. I. Shin · J. G. Shon (✉)

Department of Computer Science, Graduate School, Korea National Open University, Seoul, South Korea
e-mail: jgshon@knou.ac.kr

J. S. Park

Department of Computer Science and Engineering, Jeonju University, Jeonju-si, South Korea
e-mail: jisupark@jj.ac.kr

1 Introduction

Semiconductor circuits are complex, and miniaturization is continuously progressing. In the process, many sub-processes are included, and various facilities for the additional sub-processes are also developed. When the number of such processes increases in semiconductor manufacturing, the time from wafer input to finished product, that is, turnaround time (TAT) increases, resulting in an increase in manufacturing cost. To solve this problem, various studies are conducted to reduce the process time and simplify the process itself [1–5]. Semiconductors have eight main processes, among which five processes, called pre-processes, are repeated hundreds of times. Among the five processes, the photo process for engraving circuits on the wafer consists of a photoresist coating process, an exposure process, an overlay measurement process, a threshold measurement process, and a surface inspection process. The facility used for the exposure process includes over 2500 real sensors and uses lithography fault detection and classification (LFDC) to monitor each sensor and classify outliers. LFDC detects and classifies outliers in each sensor value and measures the sorted wafers in the measurement process to confirm actual defects. The problem with LFDC is that the outlier detection performance is less than 15% owing to numerous real sensors, causing many wafers to wait for the metrology process, reducing the TAT. In addition, there are frequent cases of process accidents, implemented by arbitrarily skipping the measurement process or changing the outliers of the sensors. In this study, to solve the problem of the corresponding real sensor, the actual sensor value is learned by machine learning and the output value is implemented as a virtual sensor to verify the defect during the exposure process.

2 Related Research

2.1 *Semiconductor Process*

A single semiconductor chip comprehensively includes eight major processes as shown in Fig. 1.

In the oxidation process, oxygen and water vapor chemically reacts with the wafer surface at high temperatures to form a uniform silicon oxide film that acts as an insulating film for the wafer or as a diffusion barrier in the ion implantation process. The photo process uses light to transfer the mask containing the circuit pattern onto the wafer, and the etching process peels off the light and non-contact parts of the wafer engraved in the circuit. The deposition and ion implantation processes are applied as thin films on the wafer to afford electrical characteristics, and the metal wiring process is a process of connecting metal lines along a circuit pattern of a semiconductor. These five processes are specifically called pre-processes and are repeated hundreds of times to completely implement thousands of chips on one wafer.

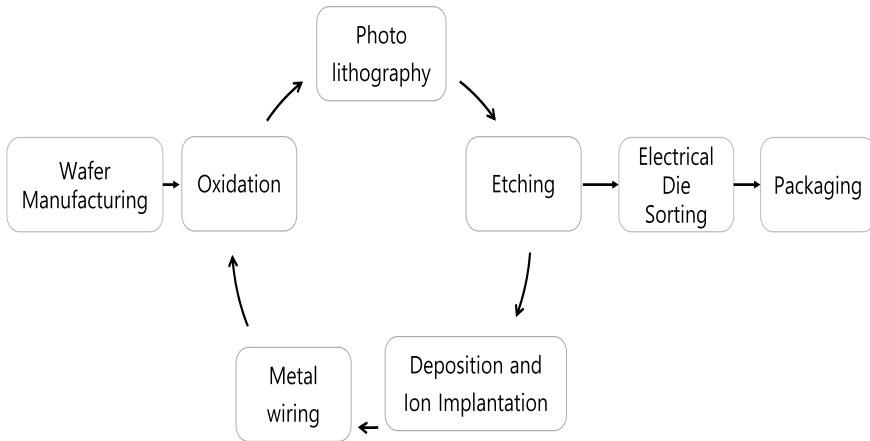


Fig. 1 Semiconductor process

2.2 Machine Learning

XGBoost [6] is an algorithm implemented to support parallel learning in gradient boost that is a representative algorithm implemented using the booting method that is one of the ensemble methods [7] that uses a combination of several weak decision trees. It supports both regression and classification problems and exhibits the advantage of speedier execution compared with gradient boost based on the application of parallel processing function and automatic pruning algorithm. It incorporates an overfitting regulation function, self-cross-validation algorithm, and missing value processing function, to learn the model in a balanced tree partitioning method, thereby forming a symmetrical tree, and it also includes an early termination function. This feature can be used for both regression and classification problems. The multi-label problem is a problem in which y includes a matrix of 0 and 1 as many as n classes, and one sample can include multiple categories. Representative algorithms include one versus rest classifier and multi-output classifier. One versus rest classifier considers an indicator matrix as input, and this matrix includes the label of j in i samples. Multi-output classifier is a method that can solve multi-label classification problems in any classification problem and learns one classifier per response variable. In other words, it is possible to classify multiple response variables and measure the objective function that predicts the response variables by extending the predicted values of the independent variables.

3 Suggestion Method

3.1 Single Virtual Sensor

Single virtual sensor in Fig. 2 classifies 2500 real sensors as sensors for each module of the facility as input and uses 200 classified sensors for each module as learning data. For the output, a virtual sensor that finally assesses whether the process result is normal or abnormal is implemented by combining six key management results among the measurement results in the measurement process. Since the semiconductor data have a severely unbalanced data form [8], oversampling was performed with SMOTE [9] as the preprocessing method, and the measurement result was set as an appropriate label for supervised learning by setting a threshold value. The learning algorithm used for machine learning used the XGBoost algorithm that demonstrated the best performance and consistency through experiments, and the model's performance was evaluated from the learning results in terms of accuracy and $F1$ -score. The model with the best performance among the learned models is set as a virtual sensor, and it outputs zero and one, by directly receiving the sensor value from the facility as input and judging it as normal or abnormal, they are implemented as an interlock in the LFDC, and if the output is one, it is judged as inappropriate. The measurement is conducted on an actual measurement equipment.

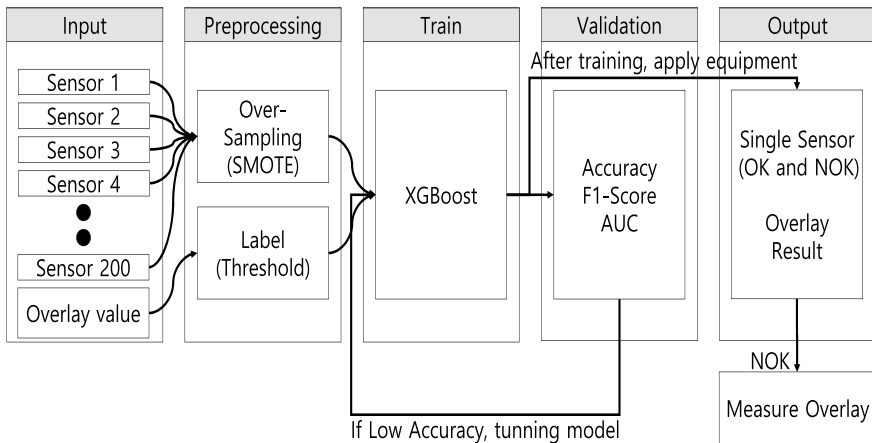


Fig. 2 Single virtual sensor

3.2 Multiple Virtual Sensors

Multiple virtual sensors in Fig. 3 also use 200 sensors for each facility module among the actual 2500 input sensors.

The input is the same as that to a single virtual sensor, and the classification algorithm uses XGBoost. However, the label used for supervised learning is used as a label by setting threshold values for each of the six key management values, average x/y , 3sigma x/y , and residual x/y , among the results of the measurement process. To classify six multiple labels, a multi-output classifier algorithm was added to classify six result values with the same input as a single virtual sensor. A virtual sensor that judges normal or abnormal with respect to the six results was implemented by setting the six classified results as virtual sensors. To solve data imbalance, oversampling was performed using SMOTE. To verify the performance of the multi-output classifier, the classification report algorithm was applied to accurately verify the performance and overall performance of each virtual sensor. Since the single virtual sensor learns each key management value one by one, six individual models are required for six virtual sensors, but it can be implemented with only one model for six virtual sensors by using the multi-output classifier. Although it is possible to implement a virtual sensor with a smaller load, two methods are proposed because the load is higher when the core management value of the measurement result is one.

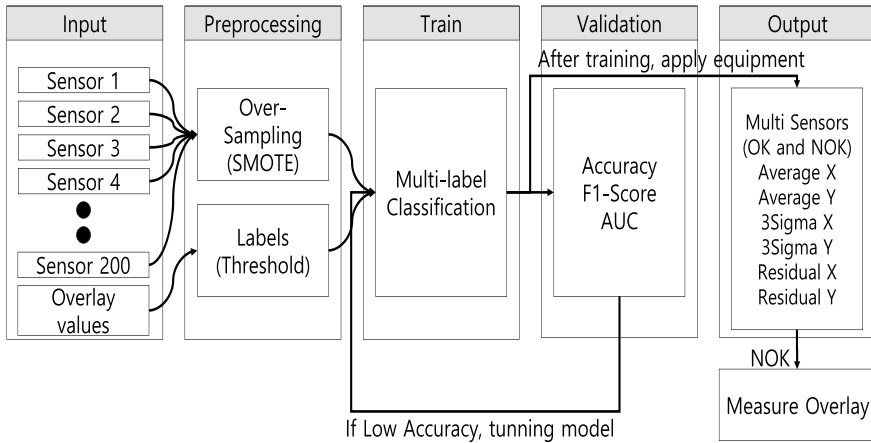


Fig. 3 Multiple virtual sensor

Table 1 Single virtual sensor

Hyper parameter	Accuracy	F1-score
No optimization	0.917	0.955
Grid search	0.927	0.960
GS + random optimization	0.930	0.962

Table 2 Real equipment test results

	Equipment A	Equipment B
Total wafers	32,734	10,000
Normal wafers	32,730	9832
Abnormal wafers	4	168
Defect rate (%)	0.01	1.7

4 Experiments

4.1 Single Virtual Sensor

Table 1 shows the performance of single virtual sensor. The accuracy was 0.917 without hyperparameter optimization, and 0.927 was obtained after optimizing the hyperparameter based on grid search. In addition, after arbitrary fine optimization of the results of grid search, the final accuracy of 0.930 was obtained. The $F1$ -score also improves sequentially.

Table 2 is a model implemented as a virtual sensor, and it is the result of receiving and classifying the sensor values of real facilities. Equipment A judged 32,730 wafers as normal out of a total of 32,734 wafers and only four wafers were judged as defective. Equipment B judged 9832 wafers as normal out of 10,000 wafers and 168 wafers as abnormal, showing a defect rate of 1.7%.

4.2 Multiple Virtual Sensors

Table 3 is the result of multi-output classifier. As a result of the classification report, the accuracy of individual labels is not shown, but the $F1$ -score, precision, and recall of each individual virtual sensor all exhibit a high performance of 0.90 or more. The accuracy of the model was 0.93% that was like that of a single virtual sensor.

Table 3 Multiple virtual sensors

	Precision	Recall	F1-score
Average x	0.90	0.99	0.94
Average y	0.91	0.99	0.94
3sigma x	0.89	1.00	0.94
3sigma y	0.91	1.00	0.95
Residual x	0.88	0.99	0.93
Residual y	0.92	0.98	0.95
Accuracy			0.93
Macro-avg.	0.90	0.99	0.94
Weighted avg.	0.90	0.99	0.94

5 Conclusion

Recent mortality forecasting studies using artificial neural networks indicate the possibilities of forecasting performance improvement using artificial neural networks, compared with the previous mortality forecasting studies. However, annual mortality rates have been used in the related mortality forecasting studies using ANNs. Therefore, the in-sufficient data in mortality forecasting compared to the quantity of data in other fields in the application of ANNs is problematic.

In this study, we have performed experiments on data augmentation in order to contribute to the improvement of mortality forecasting performance by augmenting relevant time series data in mortality forecasting through artificial neural networks. For the experiments, the linear interpolation method has been applied to the data of k_t , the mortality index describing the degree of improvement in mortality rates. The annual mortality rates have been augmented using the assumption for fractional ages and judgment. As a result of the experiments, it was observed that the artificial neural network trained with time series data augmentation improved the mortality forecasting performance compared with that without data augmentation.

This study focused on whether the mortality forecasting performance could be improved by incorporating the related time series data augmentation. In the future, we may extend the research on the forecasting performance considering additional factors such as augmented data types, data augmentation levels, and methods.

References

1. Ding D, Wu X, Ghosh J, Pan D-Z (2009) Machine learning based lithographic hotspot detection with critical-feature extraction and classification. In: 2009 IEEE international conference on IC design and technology, pp 219–222
2. Jang S, Jo M, Cho S, Moon B (2018) Defect prediction using machine learning algorithm in semiconductor test process. J Korean Inst Electr Electron Mater Eng 31(7):450–454

3. Kim HG, Han YS, Lee J-H (2015) Package yield enhancement using machine learning in semiconductor manufacturing. In: 2015 IEEE advanced information technology, electronic and automation control conference (IAEAC). IEEE
4. Moyne J, Iskandar J (2017) Big data analytics for smart manufacturing: case studies in semiconductor manufacturing. *Processes* 5(3):39
5. Mathirajan M, Sivakumar AI (2006) A literature review, classification and simple meta-analysis on scheduling of batch processors in semiconductor. *Int J Adv Manuf Technol* 29(9):990–1001
6. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (KDD'16). Association for Computing Machinery, pp 785–794
7. Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33(1):1–39
8. Abd Elrahman SM, Abraham A (2013) A review of class imbalance problem. *J Netw Innov Comput* 1(2013):332–340
9. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357

An Analysis of the Longitudinal Changes in the Determining Factors for Adolescents' Self-esteem with Random Forests



Hyejoo Lee and Euihyun Jung

Abstract Self-esteem is an important development factor in adolescence. Therefore, in the field of education, efforts have been made for a long time to find variables that affect self-esteem. This study aims to examine the longitudinal trend of variables related to self-esteem using the mean decrease impurity (MDI) of random forests with large-scale longitudinal data, and how the trend appears differently in the upper elementary and middle school years. The results showed that important variables change as the grades went up. Also, there were differences in important variables for self-esteem according to gender. These results suggest that a different approach is needed for improving self-esteem by gender. This study shows that MDI of random forests can play a big role in understanding the importance of variables that explain and predict self-esteem in the field of education.

Keywords Adolescents' self-esteem · Random forests · Determining factors

1 Introduction

Adolescence is a period of many changes through all developmental areas of physical, mental, and social aspects [1]. How adolescents spend this time has a huge impact on their later life. Self-esteem is a subjective judgment of how much an individual considers himself or herself to be a worthy being, and it induces positive characteristics such as promoting individual competence and confidence, positive thinking, and happiness. Therefore, it is very important to explore how self-esteem is formed and what factors are related and influencing it.

H. Lee

Department of Education, Chung-Ang University, Seoul, Korea
e-mail: ladyzen@cau.ac.kr

E. Jung (✉)

Department of AI Convergence, Anyang University, Anyang, Kyunggi-do, Korea
e-mail: jung@anyang.ac.kr

According to studies on the change pattern of self-esteem [2, 3], various arguments coexist in the direction of change. However, all the studies agree about the change itself, and in general, there is a difference in self-esteem between childhood and adolescence. From childhood to adolescence, the center of relationships shifts from family to school, so factors that influence individual thoughts, perceptions, and attitudes change [4]. In addition, the perception as a social being and the judgment of one's own abilities make it hard to analyze the change.

Self-esteem was found to be related to various variables such as individual, family, and school. However, results according to the existing studies were inconsistent. Some studies identify positive variables for self-esteem while others identify negative, or vice versa. Traditional statistical methods such as regression analysis or structural equation model have been very useful in educational domain, but their results often conflict because they showed limitations in exploring the influence, the longitudinal changes, and reciprocal relationships of vast variables related to individuals, families, and schools from multiple perspectives.

This study aims to examine the longitudinal trend of variables' importance related to self-esteem using random forests [5] with large-scale longitudinal data, and how they appear differently in the upper elementary and middle school years. Random forests does not assume a linear relationship between the independent variable and the dependent variable. It enables flexible variable exploration and shows high predictive power even when many variables are simultaneously input. Unlike the general use of random forests, this study seeks to find the importance of features using mean decrease impurity (MDI) [6]. MDI computes the total reduction in loss or impurity contributed by all given features. This method is computationally very efficient and has been widely used in a variety of applications to decide the relative importance of features.

The rest of this paper is organized as follows. Section 2 describes a dataset for this study. In Sect. 3, the authors explain the results of MDI from random forests analysis and interpret the educational implications of the results. Section 4 concludes the paper.

2 Dataset

The study used the dataset originated from the Korean Children and Youth Panel Survey (KCYPS), a longitudinal study of a nationally representative sample of South Korean children and adolescents collected by the National Youth Policy Institute (NYPI) of South Korea [7]. The KCYPS conducted six follow-up surveys from 2010 (Wave 1) to 2016 (Wave 6). For this study, data from the second wave for the fifth graders in elementary school and the sixth wave for the third graders in middle school were analyzed. In data preprocessing, 18 input variables and one target variable were selected. The target variable, *self-esteem* was categorized into two values of high and low depending on the degree of self-esteem. The second wave and the sixth wave were distinguished by the suffixes of “_e42” and “_e46”, respectively. The selected

input variables and correlation coefficients are summarized in Table 1. Also, negative correlations were marked in bold.

Table 1 Overview of input variables and correlation coefficients to the target variable

Variable	Description	Correlation (r)
<i>health</i>	Health status	0.266*** (_e42), 0.297*** (_e46)
<i>grasat</i>	Grades satisfaction	0.375*** (_e42), 0.313*** (_e46)
<i>caneg</i>	Parental neglect (4 questions, Cronbach's $\alpha = 0.77$)	- 0.426*** (_e42), - 0.339*** (_e46)
<i>caabu</i>	Parental abuse (4 questions, Cronbach's $\alpha = 0.81$)	- 0.375*** (_e42), - 0.330*** (_e46)
<i>schact</i>	Learning activities in school (5 questions, Cronbach's $\alpha = 0.72$)	0.467*** (_e42), 0.422*** (_e46)
<i>schrul</i>	School rule compliance (5 questions, Cronbach's $\alpha = 0.78$)	0.360*** (_e42), 0.274*** (_e46)
<i>schfri</i>	Friend relationship (5 questions, Cronbach's $\alpha = 0.73$)	0.490*** (_e42), 0.486*** (_e46)
<i>schtea</i>	School teacher relationship (5 questions, Cronbach's $\alpha = 0.86$)	0.294*** (_e42), 0.356*** (_e46)
<i>local</i>	Community awareness (6 questions, Cronbach's $\alpha = 0.71$)	0.343*** (_e42), 0.270*** (_e46)
<i>commu</i>	Community sprit (4 questions, Cronbach's $\alpha = 0.79$)	0.331*** (_e42), 0.272*** (_e46)
<i>comgame</i>	Computer games	- 0.143*** (_e42), - 0.056* (_e46)
<i>phdep</i>	Reliance on mobile phone (7 questions, Cronbach's $\alpha = 0.88$)	- 0.238*** (_e42), - 0.258*** (_e46)
<i>pecom</i>	Peer communication (3 questions, Cronbach's $\alpha = 0.83$)	0.431*** (_e42), 0.403*** (_e46)
<i>petru</i>	Peer trust (3 questions, Cronbach's $\alpha = 0.81$)	0.439*** (_e42), 0.398*** (_e46)
<i>peiso</i>	Peer alienation (3 questions, Cronbach's $\alpha = 0.70$)	- 0.348*** (_e42), - 0.376*** (_e46)
<i>papeknow</i>	Parents' recognition of my friends	0.222*** (_e42), 0.152*** (_e46)
<i>papemeet</i>	Parents' meeting with my friends	0.163*** (_e42), 0.098*** (_e46)
<i>papelike</i>	Parents' preference for my friends	0.219*** (_e42), 0.153*** (_e46)

* $p < 0.05$, *** $p < 0.001$

3 Analysis with Random Forests

3.1 Overview

The study used MDI of random forest to find most affecting variables. First, the entire data was analyzed, and the analysis was performed again by gender. The accuracies are 78.44% (total), 79.27% (male), and 80.1% (female), respectively. However, since the purpose of this study is not to obtain higher accuracy for the prediction, we focused more on obtaining MDI. This is because MDI is a useful way to understand the importance of input variables.

In each analysis, the three most important features were obtained by estimating MDI, and how these features changed according to the year was confirmed. Also, it was confirmed whether this change differed by gender. Of course, MDI does not look for positive or negative influences of each variable, but for variables with high contribution. Therefore, to determine the direction of influence, the correlation coefficient was checked.

3.2 Analysis Results

Longitudinal Comparison. According to Fig. 1, in the case of the fifth grader in elementary school, “peer alienation (*peiso_e42*, MDI = 0.092)” was found to be the most important variable in predicting self-esteem. Next, “learning activities in school (*schact_e42*, MDI = 0.086)” and “parents’ recognition of my friends (*pape-know_e42*, MDI = 0.075)” were most important. In case of the third grader in middle school, “learning activities in school (*schact_e46*, MDI = 0.126)” was found to be the most important variable in predicting self-esteem. Next, “reliance on mobile phone (*phdep_e46*, MDI = 0.104)” and “community awareness (*local_e46*, MDI = 0.093)” were important.

These results suggest that peer alienation and parental involvement with friends were important at a young age. Since peer alienation has a negative correlation to the target variable, children who experience peer alienation are more likely to have low self-esteem. On the other hand, as the grades increased, school studies and the reputation became more important. That is, in adolescence, it was found that social awareness plays an important role for self-esteem.

Longitudinal Comparison of Two Female Groups. According to Fig. 2, in the case of the fifth grader in elementary school, “computer games (*comgame_e42*, MDI = 0.099)” was found to be the most important variable in predicting self-esteem. Next, “parental neglect (*caneg_e42*, MDI = 0.093)” and “community awareness (*local_e42*, MDI = 0.076)” were most important. In the case of the third grader in middle school, “peer trust (*petru_e46*, MDI = 0.181)” was found to be the most important variable in predicting self-esteem. Next, “parents’ meeting with my friends

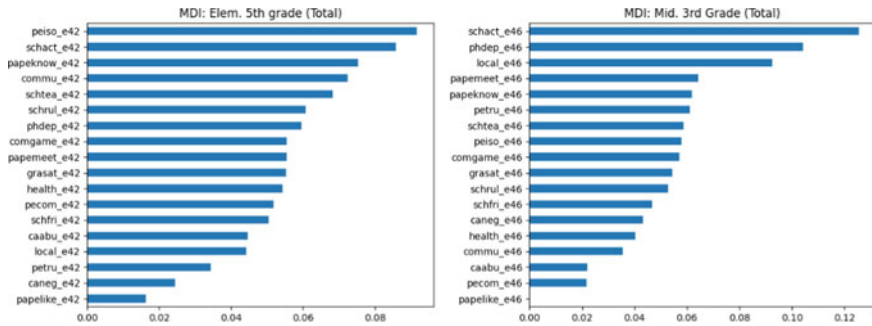


Fig. 1 Left graph is the MDI result of the fifth grader in elementary school, and the right graph is the MDI result of the third grader in middle school

(papemeet_e46, MDI = 0.133)” and “reliance on mobile phone (*phdep_e46*, MDI = 0.103)” were important.

For female students, computer games and parental neglect were important when they were young. Unusually, these two features were found to have a negative correlation with self-esteem as shown in Table 1. It means that girls who play computer games a lot and are neglected by their parents are more likely to have low self-esteem. In contrast, in adolescence, “peer trust” and “parents’ meeting with my friends” were found to be the most important factor. It means that the relationship between peers has a greater influence on self-esteem as the grade increases.

Longitudinal Comparison of Two Male Groups. According to Fig. 3, in the case of the fifth grader in elementary school, “peer alienation (*peiso_e42*, MDI = 0.136)” was found to be the most important variable in predicting self-esteem. Next, “community spirit (*commu_e42*, MDI = 0.106)” and “teacher relationship (*schtea_e42*, MDI = 0.104)” were important. In case of the third grader in middle school, “grades satisfaction (*grasat_e46*, MDI = 0.131)” was found to be the most important variable in predicting self-esteem. Next, “reliance on mobile phone (*phdep_e46*, MDI =

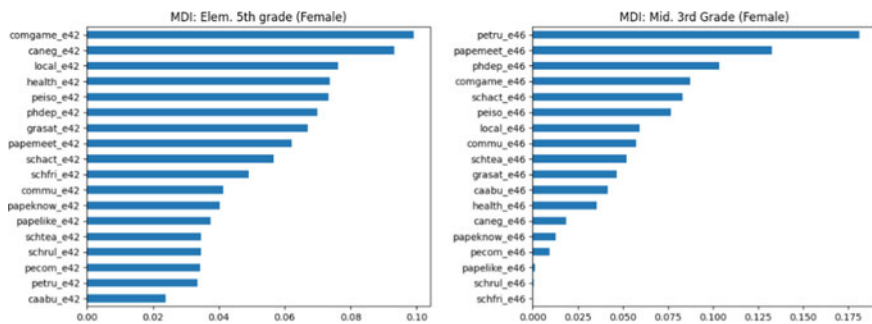


Fig. 2 Left graph is the MDI result of the fifth female grader in elementary school, and the right graph is the MDI result of the third female grader in middle school

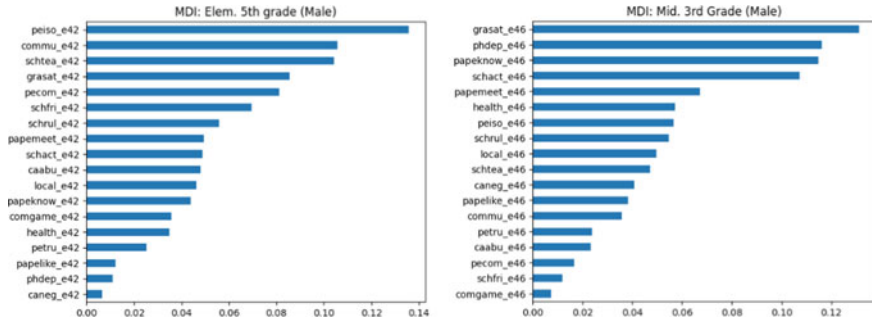


Fig. 3 Left graph is the MDI result of the fifth male grader in elementary school, and the right graph is the MDI result of the third male grader in middle school

0.116)” and “parents’ recognition of my friends (*papeknow_e46*, MDI = 0.115)” were important.

For male students, peer alienation and relationships with teachers were important when they were young. Since peer alienation has a negative correlation to the target variable, boys who experience peer alienation are more likely to have low self-esteem. In this period, social activities in school life seem to affect self-esteem. In contrast, in adolescence, grade satisfaction was found to be the most important factor. These results show that social reputation becomes more important during this period.

4 Conclusion

Self-esteem is an important developmental factor in adolescence. Therefore, in the field of education, efforts have been made for a long time to find variables that affect self-esteem. However, with the existing statistical methods, it was difficult to find variables other than hypotheses assumed in advance, and there were conflicts among the variables found. In this study, an approach was taken to determine the importance of variables using the MDI of random forests. Through this method, it was possible to analyze the factors affecting the self-esteem of adolescents from various angles. Also, this study shows that MDI of random forests can play a big role to determine the affection of variables for self-esteem in the education field. However, this study has a limitation to figure out the direction of the affection, because MDI can only tell the degree of affection. To complement this, we examined the correlation coefficients of variables in the study. In the future, the prediction values from random forests will be examined to check not only the degree of the affection but also the directions of it.

Acknowledgements This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5B5A17047090).

References

1. Zeigler-Hill V (2013) Self-esteem. Psychology Press
2. Erol RY, Orth U (2011) Self-esteem development from age 14 to 30 years: a longitudinal study. *J Pers Soc Psychol* 101(3):607
3. Robins RW, Trzesniewski KH (2005) Self-esteem development across the lifespan. *Curr Dir Psychol Sci* 14(3):158–162
4. Anderman EM, Midgley C (1997) Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemp Educ Psychol* 22(3):269–298
5. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
6. Louppe G, Wehenkel L, Sutter A, Geurts P (2013) Understanding variable importances in forests of randomized trees. In: *Advances in neural information processing systems*, vol 26
7. National Youth Policy Institute (2010–2016) The 2010–2016 Korean children and youth panel survey project report. Seoul, Korea

Design of Multi-channel Ethernet System Based on FPGA



Yougu Ale, Yang Xu, and Chao Wu

Abstract In some autonomous vehicle test scenarios, signal generators are required for generating the testing signals to simulate the real surrounding environment of the vehicle. This paper demonstrated a design scheme of multi-channel Ethernet system which can transmit the testing data through the multiple channels, a FPGA chip is used as the core device of the system, it can control the system to read the UDP frames stored in the two DDR4 DRAM chips integrated on board and processing them, by precisely time control, then the UDP data frames will be transmitted through corresponding Ethernet channel with a programmable time interval between the Ethernet channels, and each UDP frame transmission can be controlled. The result shows that each Ethernet channel of the system can transmit all the data correctly and timely, and its performance is quite stable and can be used as the multiple Ethernet signal generator in the autonomous vehicle testing scenarios.

Keywords Ethernet · FPGA · Multi-channel · Programmable time interval

1 Introduction

In recent years, the auto-driving technology has been developing rapidly, and Ethernet has been widely used in the in-vehicle communicating system because it can achieve higher transmission bandwidth than traditional in-vehicle communication protocols such as CAN, LIN, or MOST [1]. Therefore, many vehicle-mounted devices which are used to help achieving autonomous driving, such as the LiDAR, an important sensor which is used to detect the surrounding obstacles, are using the Ethernet UDP protocol to transmit data. Considering the lack of safety and stability of autonomous vehicles in the early stage of development, it is necessary to develop a device to

Y. Ale (✉) · Y. Xu
Chongqing University of Post and Telecommunications, Chongqing, China
e-mail: aileyogu@qq.com

C. Wu
Department of Network Business, China Mobile Group Co., Ltd., Chongqing, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_17

simulating the functions of multiple vehicle-mounted Ethernet sensors, sending test data to the unit under test during the auto-driving test stage. This paper [2] particularly described the advantage of simulation test, such as low cost, high efficiency.

In order to simulate the real driving environment, the device is required to have the ability of transmitting the test data through multiple Ethernet channels in the same time. There are a few related designs we can find, for instance, Lu's team presented a design of multi-channel data acquisition system based on FPGA in this paper [3], and this design used a FPGA chip to acquire data by controlling multiple ADCs and transmit the acquired data to the computer through the Ethernet lane, but there are no data synchronize between the channels. Johansson presents a way to implement the ethernet based on FPGA [4], above the physical layer all other layers of ethernet OSI module are implemented inside the FPGA through RTL coding and design in this design, and the physical layer was realized by using a PHY chip; between the two layers above mentioned, it was connected by a set of MII interface.

2 System Design

By referring the related design above, we proposed our basic modules of this design, and the modules consisted of data buffer block, data process block, and data transmission block. The main design work is composed of hardware design and software design.

2.1 System Hardware Design

The hardware system is mainly composed of three parts: core control module, Ethernet transmission module, and basic support module. The core control module is composed of Xilinx XCZU5EV chips, which are designed through RTL code to control the entire system workflow. The main function of the Ethernet transmission module is to realize the Ethernet based on the FPGA, in this scheme, in order to consider the needs of system flexibility and scalability, the choice of using the FPGA internal MAC IP core and external Ethernet PHY chip through the RGMII expansion interface connection to achieve the way to achieve [5]; the module of the PHY chip selected can achieve 10/100/1000 Mbps transmission of KSZ9031RNX chip. The basic support module is mainly composed of various components that support the normal operation of core devices such as power supply modules and crystal oscillator clocks. The system hardware structure and physical objects are shown in Fig. 1.

And, the core PCB board of the system is shown as in Fig. 2.

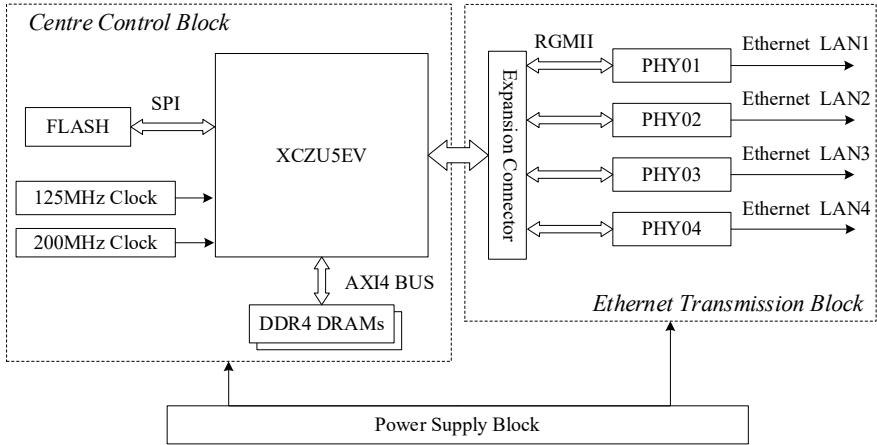
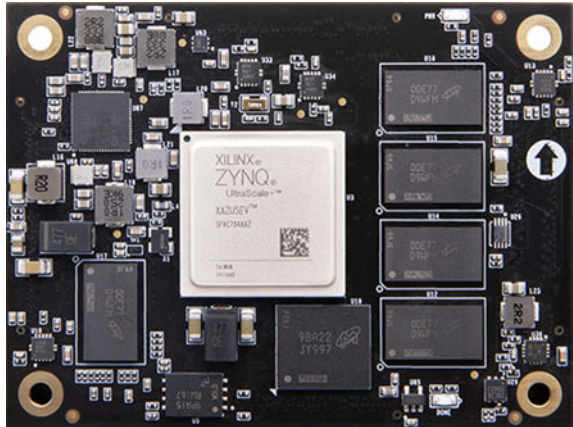


Fig. 1 System hardware architecture

Fig. 2 System core PCB board



2.2 System Software Module Design

This paper uses Verilog, one of the RTL level languages, to design the system software modules with the top-down design method, and it can be divided into three submodules: data processing module, time control module, and data transmission module, and the block diagram is shown in Fig. 3.

The data processing submodule is used to achieve the parameter modification of the MAC layer message data used for testing, as well as the identification of channel information and the re-CRC verification. Since the device needs to be directly connected to the test host when working, the device and the tested host cannot obtain each other's MAC address and IP address information, so it is necessary to modify

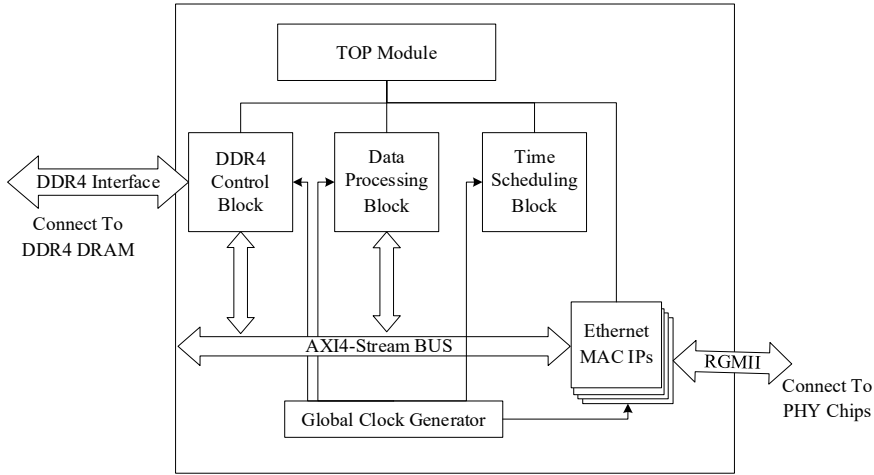


Fig. 3 System software module architecture

the address parameter information and port information inside the test message to the address information of the test host before the test [6].

As Fig. 4 illustrated, the data processing module will replace the parameters such as the destination MAC address, source MAC address, destination IP address, and source IP address in the MAC layer message with the provided parameters, and at the same time, after the replacement parameters modify the data, the MAC frame is re-CRC checked, and the CRC check value is replaced with the recalculated value [7].

The time control module is responsible for adjusting the transmission time of each frame of data packets on each channel. In this design, each time control module will send a frame of messages to a specified channel at intervals of a fixed period of time, so that the messages are sent out through the selected channel, and the module internal state switching diagram is shown in Fig. 5.

As Fig. 5 shows, the module has three states, and when the module is in the status of IDEL, it means that the system has not been initialized yet; when the module is in the status of waiting, it means that the last data transmission was completed, waiting for new data; and when it is in the status of CHx_sending, it means that the

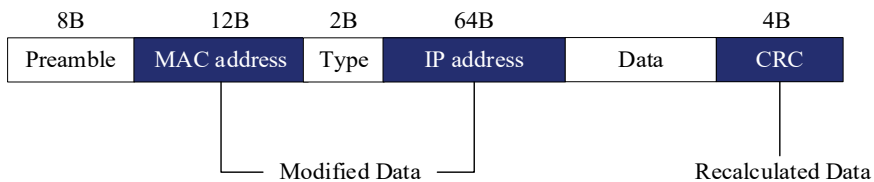


Fig. 4 Data need to be modified in the UDP packet

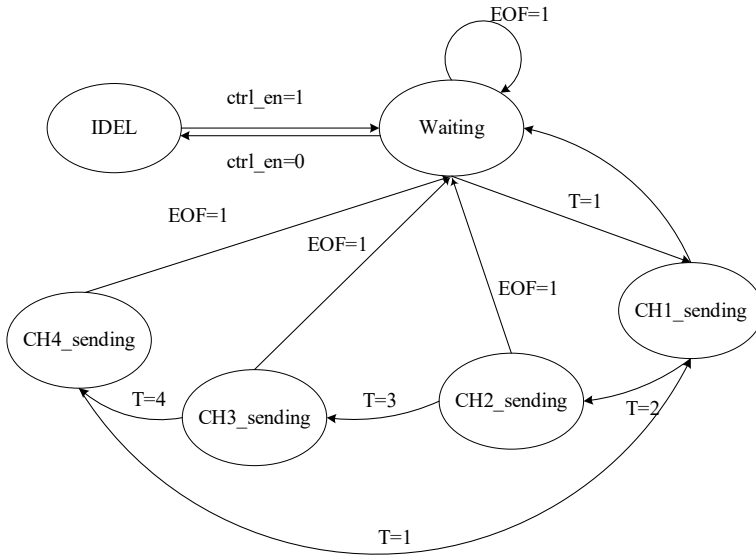


Fig. 5 Status change diagram in time control block

corresponding x channel is sending data. And as T changes, the system module will switch the channel sent, and if the data sent by any channel is the last frame of the message, it will switch to the waiting state to wait for new data.

After completing the module design, the simulation results are shown in Fig. 6, the four Ethernet MAC sub-IP modules of the system output data normally, and the message transmission interval on each two channels is the set time value.

3 Result and Conclusion

After the design was complete, we have stored a set of UDP frames in the block ram of the FPGA and transmit the frames through the Ethernet channels; meanwhile, we use different computers to capture the UDP frames sent by the FPGA, and Fig. 7 shows the physical connection diagram of the system and the test result.

The test results showed that all channels have correctly received all UDP packets, and the received packet frame time slot is $16 \mu\text{s}$; this is because the device will send packets to four Ethernet channels at $4 \mu\text{s}$ intervals each time, so each channel will receive the next frame after $16 \mu\text{s}$, and the surface design of the test results meets the requirements.

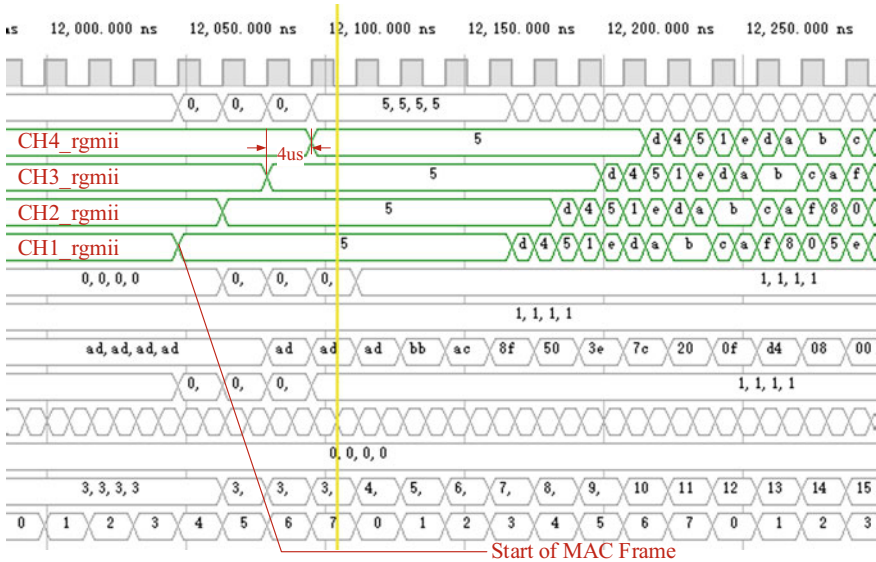


Fig. 6 Simulation



Time	Source	Destination	Protocol
1 0.000000	192.168.1.200	192.168.1.102	UDP
2 0.000016	192.168.1.200	192.168.1.102	UDP
3 0.000032	192.168.1.200	192.168.1.102	UDP
4 0.000048	192.168.1.200	192.168.1.102	UDP
5 0.000064	192.168.1.200	192.168.1.102	UDP
6 0.000080	192.168.1.200	192.168.1.102	UDP

net 5: 1748 bytes on wire (9984 bits), 1748 bytes captured (9984 bits)

```

54 e1 ac bb ac 8f 50 3e 7c 20 0f d4 08 00 45 00 | [E]...P> | ...E-
04 d2 25 04 40 00 80 11 4c 98 c0 a8 01 c8 c0 a8 | ...%@...L.....
01 66 09 41 09 40 04 be fc 9c ff ee e8 62 18 00 | ...fA@...b...
52 1e 00 80 15 00 53 20 00 84 0e 00 5c 1b 00 84 | R...S... \...
12 00 53 1c 00 80 0b 00 68 1f 00 90 19 00 54 1d | ...S...h...T...
00 98 11 00 64 19 00 8c 00 00 58 10 00 8c 11 00 | ...d...X...
60 20 00 90 15 00 54 1b 00 8c 16 00 58 1f 00 98 | ...T...X...
18 00 53 00 00 8c 08 00 58 1e 00 9c 11 00 58 1d | ...S...X...X...
00 8c 10 00 58 1e 00 98 0d 00 54 1e 00 8c ff ee | ...X...S...T...
fa 62 17 00 52 1f 00 80 16 00 53 20 00 80 0f 00 | ...bR...S...
5c 1b 00 84 13 00 53 1c 00 80 0a 00 68 1f 00 90 | \...S...h...
19 00 54 1d 00 98 11 00 64 19 00 8c 00 00 58 0f | ...T...d...X...
00 8c 0f 00 60 21 00 90 16 00 54 1d 00 8c 15 00 | ...T...l...d...X...
58 1f 00 98 18 00 53 00 00 8c 08 00 58 1f 00 9c | X...S...T...X...
11 00 58 1d 00 8c 10 00 58 1e 00 98 0d 00 54 1e | ...X...X...T...
00 8c ff ee 0c 63 18 00 52 1f 00 80 16 00 53 20 | ...c...R...S...
00 80 0f 00 5c 1c 00 84 13 00 53 1d 00 80 0a 00 | ... \...S...
    
```

Fig. 7 Physical connection diagram and the test result

From the test result showed above, we can know that the time accuracy of the system can reach to microsecond level and can be programmed; therefore, by assigning different values of time interval to different channels, the system can simulate a variety of different test scenarios for autonomous driving test. The clock of the system is calibrated and compensated by the MMCN of the FPGA, so the system

can operate in a very stable state regardless of the change of temperature. The multi-channel system is also designed with many standard communication interfaces, for instance, PCIe3.0, USB, and so on; with the interface, the system can be integrated into the other test system easily. With the features of high-stability, convenience, and superior performance, the system can be used in all kinds of autonomous driving vehicle test scenes.

References

1. Kim H (2021) Development of an Ethernet-based heuristic time-sensitive networking scheduling algorithm for real-time in-vehicle data transmission. *Electronics* 10(2):157–158
2. Yu C-H, Chen YZ (2020) The benefit of simulation test application on the development of autonomous driving system. In: 2020 international automatic control conference (CACCS), Hsinchu, Taiwan, pp 1–5
3. Jian L, Fan L (2022) Design of multi-channel data acquisition and transmission system. In: 2022 4th international conference on intelligent control, measurement and signal processing (ICMSP), Hangzhou, China, pp 729–732
4. Johansson C (2021) An FPGA-based Ethernet switch. In: 2021 29th telecommunications forum (TELFOR), vol 10, no 1, pp 1–4
5. Pahl D (2021) A low-noise and scalable FPGA-based analog signal generator for quantum gas experiments. In: 2021 IEEE international conference on quantum computing and engineering (QCE), Broomfield, CO, pp 450–451
6. Wassim M (2019) FPGA implementation of RDMA-based data acquisition system over 100-Gb Ethernet. *IEEE Trans Nucl Sci* 66(7):1138–1143
7. Hema M, Savitha U, Nilashree W (2020) Ethernet implementation on FPGA. In: 2020 international conference for emerging technology (INCET), Belgaum, India, pp 1–4

Design of 2D to 3D Pose Estimation Using NeRF Image View Synthesis



Chan Park , Hyungju Kim , and Nammee Moon 

Abstract Skeleton construction and pose estimation through two-dimensional (2D) images may have reduced estimation accuracy due to truncation and occlusion that may occur within the image. In this paper, a 3D model is created through NeRF (Neural Radiance Fields) based Instant-NGP (Neural Graphics Primitives), a skeleton is constructed, and the artificial intelligence is trained to perform pose estimation in a 2D image. Create a 3D model by performing image view synthesis using Instant-NGP on images taken from 4 or more different angles. After that, we use DeepLabCut to build joint coordinates and skeletons. By learning the built skeleton to artificial intelligence, a pre-trained AI model is created. Using an artificial intelligence model learned in advance through a 3D model, joint coordinate recognition in a 2D image, skeleton construction, posture estimation, and classification are carried out. Through the test image dataset, the posture estimation and classification accuracy of the artificial intelligence model trained in advance with the 3D model and the artificial intelligence model learned using the existing 2D image are compared.

Keywords Pose estimation · Image synthesis · View synthesis · NeRF

1 Introduction

A study that detects a skeleton built with artificial intelligence and recognizes posture is called posture estimation. Posture estimation research is being used in various fields such as behavioral recognition, object tracking, and human–computer interaction [1]. However, existing posture estimation studies use 2D photo or video data. Therefore,

C. Park · H. Kim · N. Moon (✉)

Department of Computer Science and Engineering, Hoseo University, 20, Hoseo-ro 79 Beon-gil, Baebang-eup, Asan-si, Chungcheongnam-do, Republic of Korea
e-mail: nammee.moon@gmail.com

C. Park

e-mail: chan.park941003@gmail.com

there is a limitation in that the accuracy of posture estimation decreases due to problems such as ambiguity of joint points, detection errors, photo cropped, and occlusion [2]. To compensate for the decrease in accuracy of 2D posture estimation, a study on posture estimation through a 3D model has been proposed. After photographing a person using a binocular stereoscopic imaging camera that reflects depth and generating a 3D model, posture estimation was performed through this [3]. However, the proposed study needs to be photographed through a special camera, and there are a limitation of overlapping objects and people due to an error in depth recognition. To overcome these limitations, we propose the generation of 3D models through 2D image view synthesis. As a study that proposed 2D image view synthesis, a study to implement a frontal image as a side image using Generative Adversarial Networks (GANs) was proposed [4]. A study to create a 3D model by synthesizing the images has been proposed [5]. Therefore, in this paper, a 3D model is created by synthesizing 2D images taken from various angles, and then posture estimation and classification are performed based on this.

2 Related Works

2.1 3D Pose Estimation

Pose estimation aims to build a skeleton by specifying moving points such as joints and skeletons, and to find animal or human body parts in 2D images or videos by learning them in artificial intelligence [6]. Various studies have been proposed to perform 3D posture estimation using 2D images or videos. A study was proposed to produce a 2D skeleton into a 3D skeleton by simultaneously filming with multiple cameras [7], and a study to construct a 3D skeleton using DeepLabCut in video data taken with multiple cameras was proposed [8]. Therefore, in this paper, we intend to build the skeleton of a 3D model generated through image synthesis using DeepLabCut used in a 3D animal posture estimation study and propose pose estimation using it.

2.2 NeRF Based Image View Synthesis

3D modeling through image view synthesis is a method of composing many image data into a 3D model through image view synthesis through GANs and deep learning [9]. Among the proposed image view synthesis methods, NeRF derives $RGB\sigma$ from the spatial position (x, y, z) in the viewing direction (θ, φ) , that is, the 5D coordinates $(x, y, z, \theta, \varphi)$ as input values. Deep learning is carried out. It is a technique that generates a 3D model through 3D rendering through the $RGB\sigma$ values derived thereafter [5] (Fig. 1).

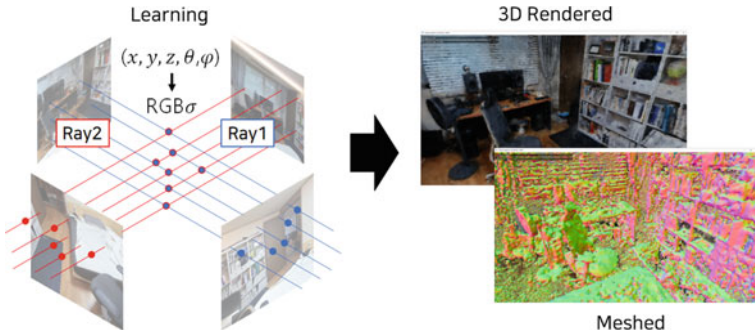


Fig. 1 Image view synthesis process using NeRF

3 Pose Estimation Using NeRF Image View Synthesis

See Fig. 2.

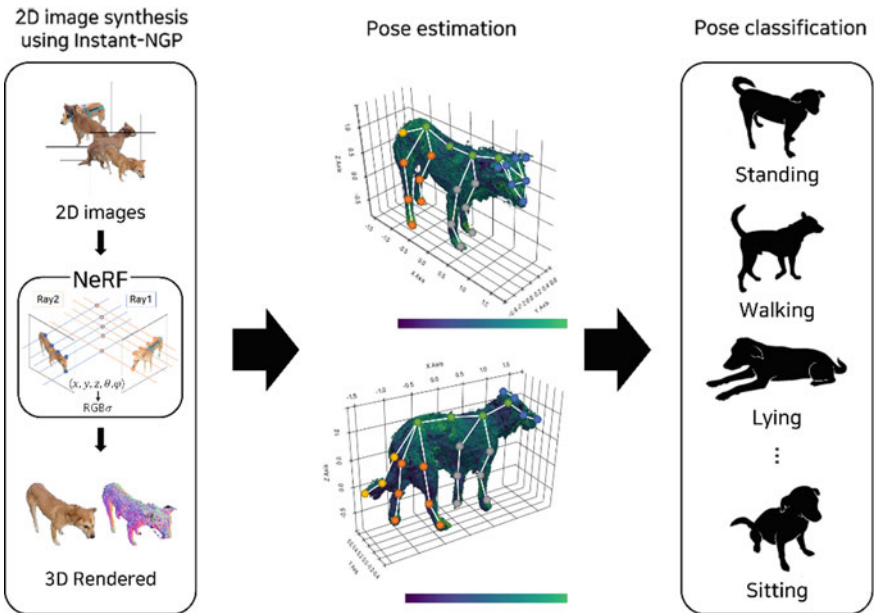


Fig. 2 Process of pose estimation using NeRF image view synthesis

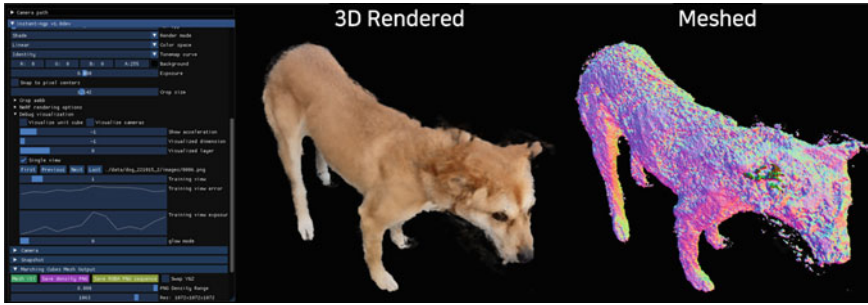


Fig. 3 Image view synthesis process using Instant-NGP

3.1 2D Image View Synthesis Using Instant-NGP

In this paper, as shown in Fig. 2, conduct a study of posture estimation through NeRF-based image view synthesis. First, image data obtained by shooting 360° centered on a dog is used. Instant-NGP (Neural Graphics Primitives) uses hash encoding and linear interpolation to convert the existing 5-dimensional coordinates $(x, y, z, \theta, \varphi)$ of NeRF into input values [10]. Through this, it is a deep learning model in which the learning time until 3D modeling is shortened by quickly deriving 5D coordinates. Therefore, image view synthesis is performed using Instant-NGP, a 3D model is created, and the mesh of the 3D model is extracted (Fig. 3).

3.2 Mesh Model Preprocessing and Skeleton Construction

After the mesh extraction process, the model is pre-processed. Due to the nature of NeRF-based image view synthesis, unnecessary meshes may be generated around the model. Due to this, the center coordinate value of the model may be misaligned, or the size of the model may be reduced. Use MeshLab to remove unnecessary meshes around the model and reduce the number of polygons through remeshing, simplification and reconstruction processes. Model optimization proceeds through this preprocessing process (Fig. 4).

After preprocessing, Pyvista is used to derive the coordinate values of the model. Pyvista is a package that can visualize 3D mesh data, and it is possible to derive coordinate values through it. Then, using DeepLabCut, the 3D model is labeled with more than 10 coordinates such as eyes, nose, head, spine, and patella to build a skeleton. Artificial intelligence learning is carried out based on the built-up skeleton. Posture estimation in a general 2D image is performed through the learned model, and posture classification accuracy is measured through this. Thereafter, joint coordinate labeling is performed using existing 2D images, and posture classification accuracy is compared with the artificial intelligence model learned (Fig. 5).

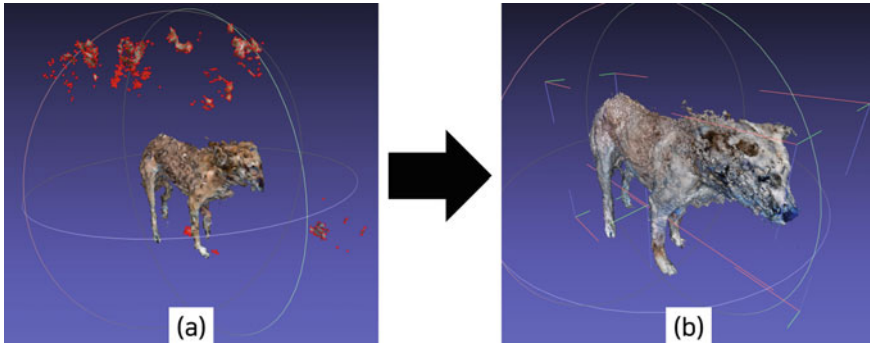


Fig. 4 Mesh model preprocess using MeshLab. **a** Before mesh model preprocessing. Red meshes are unnecessary meshes. **b** After mesh model preprocessing

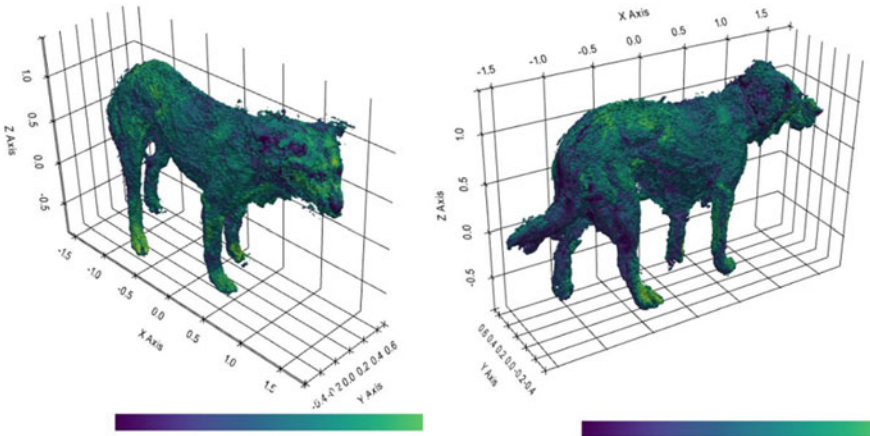


Fig. 5 Coordinate grid configuration through Pyvista

4 Conclusion

We proposed a method of generating a 3D model through NeRF-based 2D image view synthesis and estimating and classifying posture through it. It is expected that the accuracy will be higher than that of previous studies through comparison with the existing 2D posture estimation and classification studies using 2D images, and it is expected to overcome the limitations of 2D posture estimation and classification such as occlusion and photo clipping. It can be applied to abnormal behavior detection and interaction with objects in the future.

Acknowledgements This paper was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2011966).

References

1. Zhang Z, Hu L, Deng X, Xia S (2021) Sequential 3D human pose estimation using adaptive point cloud sampling strategy. In: Proceedings of the 30th international joint conference on artificial intelligence. IJCAI, Montreal, pp 1330–1337
2. Xiu Y, Li J, Wang H, Fang Y, Lu C (2018) Pose flow: efficient online pose tracking, pp 1–12. arXiv preprint [arXiv:1802.00977](https://arxiv.org/abs/1802.00977)
3. Colombel J, Daney D, Bonnet V, Charpillet F (2021) Markerless 3D human pose tracking in the wild with fusion of multiple depth cameras: comparative experimental study with Kinect 2 and 3. In: Activity and behavior computing, vol 204, pp 119–134
4. Shen X, Plested J, Yao Y, Gedeon T (2020) Pairwise-GAN: pose-based view synthesis through pair-wise training. In: International conference on neural information processing, vol 1332. Springer, Bangkok, pp 507–515
5. Mildenhall B, Srinivasan P, Tancik M, Barron J, Ramamoorthi R, Ng R (2021) NeRF: representing scenes as neural radiance fields for view synthesis. *Commun ACM* 65(1):99–106
6. Dang Q, Yin J, Wang B, Zheng W (2019) Deep learning based 2D human pose estimation: a survey. *Tsinghua Sci Technol* 24(6):663–676
7. Dong J, Fang Q, Jiang W, Yang Y, Huang Q, Bao H, Zhou X (2021) Fast and robust multi-person 3D pose estimation and tracking from multiple views. *IEEE Trans Pattern Anal Mach Intell* 44(10):6981–6992
8. Nath T, Mathis A, Chen A, Patel A, Bethge M, Mathis W (2019) Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* 14(7):2152–2176
9. Meng Q, Chen A, Luo H, Wu M, Su H, Xu L, He X, Yu J (2021) GNeRF: GAN-based neural radiance field without posed camera. In: Proceedings of the IEEE/CVF international conference on computer vision. IEEE, Montreal, pp 6331–6341
10. Müller T, Evans A, Schied C, Keller A (2022) Instant neural graphics primitives with a multiresolution hash encoding, pp 1–15. arXiv preprint [arXiv:2201.05989](https://arxiv.org/abs/2201.05989)

Pet Multiple Behavior Recognition Through Sensor Data Synthesis



Hyungju Kim , Chan Park , Jaehyeon Park , and Nammee Moon 

Abstract In behavioral recognition research based on sensor data, variables that may occur in data collection devices must be considered. Typical examples include data frequency, outliers, missing values, and noise. This paper about to pet multiple behavior recognition through 9-axis sensor data. The wearable device collects data using 9-axis sensor module (gyro, accelerator, and magnetometer). Then, after z -score normalization, preprocessing for missing values and outliers, sequence is constructed to be used as an input value for the deep learning model. At this time, when a bias occurs in a specific data class, data is augmented through a GAN-based model. Finally, we want to recognition the pose and behavior based on the hybrid model of CNN-LSTM.

Keywords Multiple behavior recognition · CNN-LSTM · Sensor data · Data augmentation

1 Introduction

Behavior detection is a study that recognizes behavior by extracting patterns or features using sensor data or image data. In image data-based research, sensor data-based research with relatively few restrictions is being actively conducted because the shooting angle must be fixed or the object must not deviate from the recording area [1, 2].

Most of the sensor data is being studied centered on Human Activity Recognition (HAR), and various deep learning and machine learning techniques are being studied to improve the accuracy of activity recognition [3, 4]. Although the configuration

H. Kim · C. Park · J. Park · N. Moon (✉)

Department of Computer Science Engineering, Hoseo University, 20, Hoseo-ro 79 Beon-gil, Baebang-eup, Asan-si, Chungcheongnam-do, Republic of Korea
e-mail: nammee.moon@gmail.com

H. Kim

e-mail: kimhyungju01@gmail.com

of the learning model is important, preprocessing of sensor data is very important. Studies that show different performance according to frequency resampling or various preprocessing methods based on the same dataset are also being conducted [5].

The collection period of sensor data is also greatly affected. Most of the previous studies of sensor data targeting humans are conducted at 20 ~ 50 Hz. However, in the case of pets, there are studies with higher frequency than previous studies due to their relatively irregular characteristics.

In the process of collecting sensor data through a wearable device, data loss or outliers may occur. In order not to burden the subjects participating in the experiment, they are collected based on a wearable device that is made light or minimal. Due to this, data loss or outliers occur because it is greatly affected by communication between hardware.

In this paper, we propose multiple behavior recognition of pets by augmenting 9-axis time series sensor data through a deep learning model. A sequence is created to use the pre-processed data based on the collected 9-axis sensor data as input data for the learning model. At this time, when data bias occurs, performance is improved through GAN-based time series data augmentation.

2 Related Works

2.1 GAN-Based Data Augmentation

When making a learning model based on deep learning, studies are underway to improve the performance of the model through data augmentation. In the case of video data, data augmentation using frame interpolation or GAN is used. In the case of sensor data, there are methods that artificially create noise or augment the data using methods that consider the rotation angle of the device [6].

A typical example of neural network-based data augmentation is a Generative Adversarial Network (GAN). GAN is a neural network technique in which the generator opposes the discriminator hostilely to create data that is as close to real as possible. There are several types of GANs, and stackGAN, which creates images by inputting text, which is one of the latest trendy methods, is a representative example. stackGAN receives text data as input and stabilizes it through Conditioning Augmentation (CA) through word embedding process. After that, a low-resolution image is generated through the stage-1 GAN, and a high-resolution image is generated with the stage-2 GAN. Therefore, in this paper, we propose a model that generates sensor data when text data is received as input by fusing the existing GAN structure and the stackGAN structure.

2.2 Behavior Recognition Using Sensor Data

Previous research on behavior recognition using sensor data is being conducted with HAR, and UCI Dataset or HAR Dataset are generally used a lot. Research to increase the accuracy of behavior recognition through sensor data using machine learning or deep learning is in progress. When the provided dataset is not used, most of the data is collected using the IMU sensor built into the smartphone or a wearable device.

In the case of pet, if products such as Fitbark or Whistle on the market are used, there is a limit to the data that users can access, such as providing only a simple amount of activity [7]. In this paper, we intend to perform multiple behavior recognition of pets with sensor data collected based on the manufactured wearable device.

3 Pet Multiple Behavior Recognition

The process of Pet multiple behavior recognition proposed in this paper is shown in Fig. 1. The research steps are data collection, data preprocessing, GAN-based data augmentation model, and multiple behavior recognition.

3.1 Data Collection

Wearable devices and ODOIDS are used to collect multiple behavior data. The frequency of the dataset used in the experiment is 50 Hz. 9-axis sensor data was collected through the manufactured wearable device and transmitted through Bluetooth communication. At this time, the data collection device used Android OS-based ODOID N2+. Data collected through Bluetooth is stored in webserver DB. The table of stored data consists of time in milliseconds, pet name, and 9-axis data. Based on

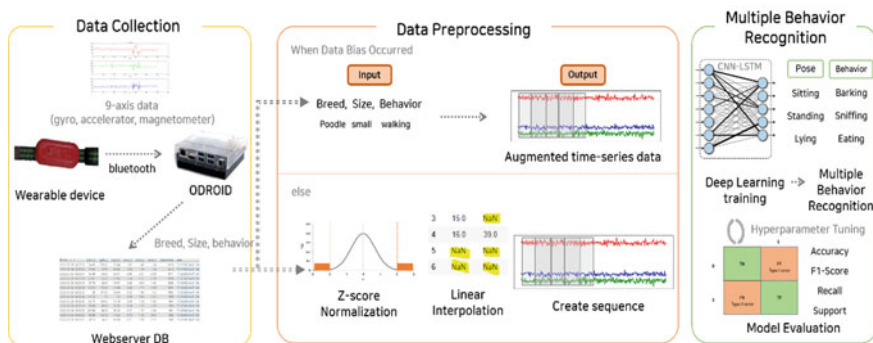


Fig. 1 The process of pet multiple behavior recognition

the collected data, roll and pitch are obtained using the accelerometer, and yaw is obtained using the magnetometer sensor and added to the column.

3.2 Data Preprocessing

Determine the number of behaviors to be recognized based on the constructed dataset. After that, it is determined whether data is augmented based on GAN according to whether data bias has occurred. The biased data is created through a GAN-based augmented model and combined with the original data to construct a new dataset. After that, missing values are replaced through linear interpolation for missing values, and Z-score normalization is performed. After that, create a sequence to use as an input for the learning model. At this time, the length of one sequence is calculated as 4 s, and consists of 200 rows.

3.3 GAN-Based Data Augmentation Model

The proposed GAN-based data augmentation model is shown in Fig. 2. Data augmentation proceeds when a specific behavioral bias occurs, and receives text data as input. The type of input data consists of breed, age, and weight for determining the size of the dog breed, and the behavior and pose of the sensor data to be generated. Based on the input features, the pet size is estimated, and augmentation is performed using the most similar breed among the collected data. The process up to upsampling refers to the structure of stackGAN. Input data enters the CA layer through word embedding. After that, CA vectors are used as input for upsampling of the generator, and the subsequent process is the same as the general GAN structure. Repeat the process to generate fake sensor data similar to the real one.

3.4 Multiple Behavior Recognition

Based on the data that has been pre-processed, deep learning is carried out. The model to be used for learning uses CNN-LSTM, which is a hybrid model of a Convolutional Neural Network (CNN) for extracting features and a Long Short-Term Memory (LSTM) for reflecting time series characteristics, in a one-dimensional form. The structure of the model is shown in Fig. 3.

Based on the learned model, we want to proceed with multi-recognition by dividing poses and behaviors. We want to recognize both the static behavior-oriented pose and the dynamic behavior-oriented behavior. For example, it may be behavior such as barking in a sitting position or eating. Based on the recognized behavior,

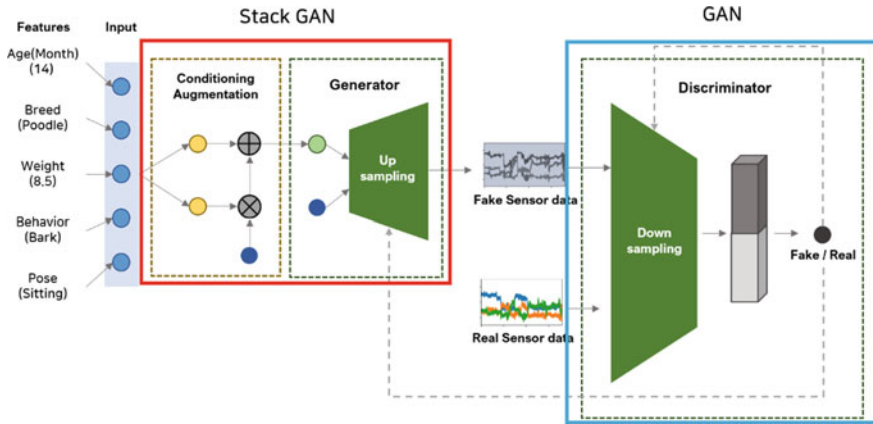


Fig. 2 The proposed GAN-based data augmentation model

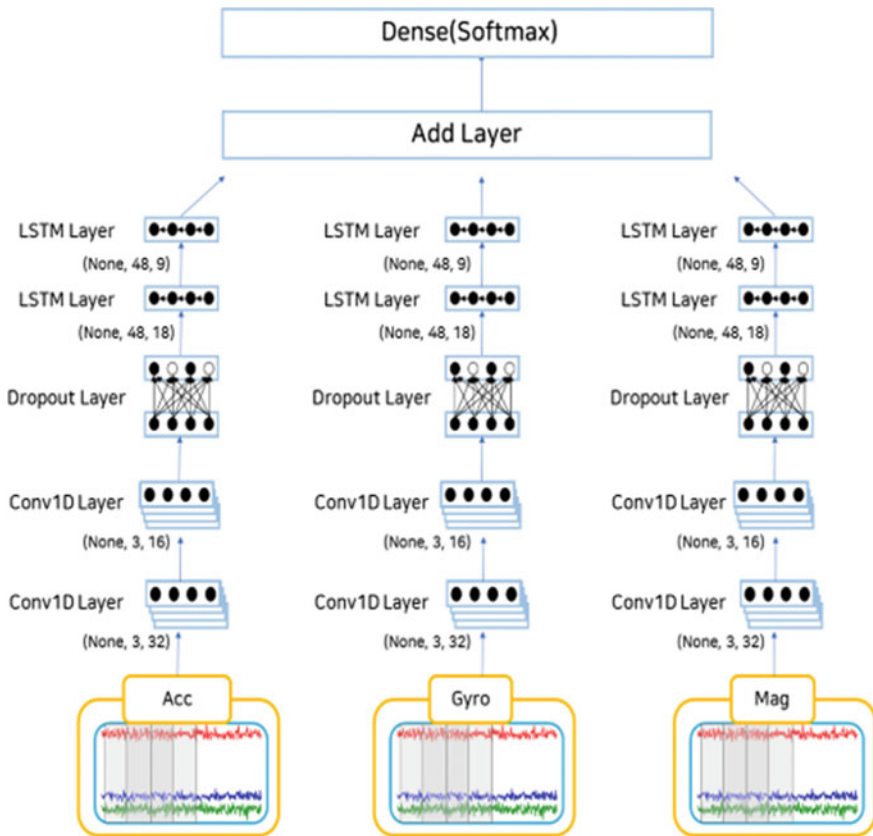


Fig. 3 The structure of CNN-LSTM hybrid model

evaluation of the learning model is conducted using the indicators of accuracy, $F1$ -score, recall, and support. After that, a high-performance learning model is created through iterative hyperparameter tuning.

4 Conclusion

This paper is about multiple behavior recognition of pets using 9-axis sensor data. The collected sensor data identifies the number of each action, determines whether to augment according to data bias, and then uses GAN-based data augmentation. After that, preprocessing for outliers, missing values, and normalization is performed. A sequence to be used as an input of training data is generated, and multi-action recognition based on a hybrid model of training CNN-LSTM is performed. An optimal model was implemented through iterative hyperparameter tuning. In the future, the proposed method expects to extend beyond multi-behavioral recognition to abnormal behavior recognition if behavior classification is subdivided.

Acknowledgements This research was supported by MISP (Ministry of Science, ICT and Future Planning), Korea, under the National Program for Excellence in SW (2019-0-01834) supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation).

References

1. Gupta S (2021) Deep learning based human activity recognition (HAR) using wearable sensor data. *Int J Inf Manage Data Insights* 1(2):1–18
2. Wan S, Qi L, Xu X, Tong C, Gu Z (2020) Deep learning models for real-time human activity recognition with smartphones. *Mob Netw Appl* 25(2):743–755
3. Nguyen HD, Tran KP, Zeng X, Koehl L, Tartare G (2019) Wearable sensor data based human activity recognition using machine learning: a new approach, pp 1–5. arXiv preprint [arXiv:1905.03809](https://arxiv.org/abs/1905.03809)
4. Masum M, Bahadur H, Shan-A-Alahi A, Chowdhury Z, Uddin R, Al Noman A (2019) Human activity recognition using accelerometer, gyroscope and magnetometer sensors: deep neural network approaches. In: 2019 10th international conference on computing, communication and networking technologies. IEEE, Kanpur, pp 1–6
5. Gu F, Chung MH, Chignell M, Valaee S, Zhou B, Liu X (2021) A survey on deep learning for human activity recognition. *ACM Comput Surv* 54(8):1–34
6. Um TT, Pfister F, Pichler D, Endo S, Lang M, Hirche S, Fietzek U, Kulić D (2017) Data augmentation of wearable sensor data for Parkinson’s disease monitoring using convolutional neural networks. In: Proceedings of the 19th ACM international conference on multimodal interaction. ACM, Glasgow, pp 216–220
7. Van Der Linden D, Zamansky A, Hadar I, Craggs B, Rashid A (2019) Buddy’s wearable is not your buddy: privacy implications of pet wearables. *IEEE Secur Priv* 17(3):28–39

Research on AFS/DYC Coordinated Control Strategy for Four-Wheel Independently Driven Electric Vehicle Based on In-Wheel Motor



Changhao Piao, Hao Wang, Ping Liu, Junren Shi, and Jianmin Dang

Abstract Active front steering (AFS) and direct yaw moment control (DYC) are proposed for four-wheel in-wheel motors independently driven electric vehicles to further improve the lateral stability of four-wheel drive vehicles during high-speed driving. An active front wheel steering system and a direct yaw moment control system based on sliding mode control (SMC) are designed, and the driving torque of each wheel is distributed by the dynamic load. The integrated control factor and controller weight are determined using the fuzzy PID controller with the real-time slip rate as the control target. The control algorithm is simulated and verified under high-speed driving conditions. The results show that: under the same typical open-loop conditions, compared with the uncontrolled vehicle, the average error of the vehicle's yaw rate is reduced by 65.14%, and the vehicle's yaw stability and driving state are improved under the controlled state.

Keywords AFS · DYC · SMC · Fuzzy control · Dynamic load · Slip rate

1 Introduction

The emergence of new energy vehicles has currently sparked a wave of energy savings and emission reduction in the automotive industry, and research and development in the automotive industry is focusing on improving vehicle safety and driving stability [1]. Distributed drive electric vehicles [2] have unique advantages in vehicle driving stability control. To elevate the comfort and driving safety of them, it is crucial to research the stability control system. In the process of their research, many scholars has laid a foundation for solving vehicle stability control [3–7]. Metzler et al. [8] designed a lateral and stability controller ground on the display model prediction,

C. Piao · H. Wang · P. Liu · J. Shi (✉)
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: shijr@cqupt.edu.cn

J. Dang
Chongqing Changan Automobile Co. Ltd., Chongqing 400023, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_20

which greatly improves the stability of the vehicle, but reduces the real-time performance of the control system. According to Yan et al. [9], the direct yaw moment decision algorithm of neural network PID control strategy is used for torque distribution. However, the vehicle must adapt to various complex driving conditions, and it is difficult for PID control to achieve precise yaw moment control under extreme conditions. Furthermore, other scholars have adopted relatively novel methods, such as designing a fault-tolerant control strategy for multi-mode switching [10], control methods based on driving intent recognition [11]. It has made a certain contribution to improving the driving stability of them. However, there are still some deficiencies in the stability control of them under high-speed driving conditions. Ultimately, it will have an influence on the lateral stability of them at high speeds.

In order to improve the lateral stability of the vehicle under the limited conditions of high-speed driving, the active front wheel steering system and the direct yaw moment control system based on SMC are designed in this paper. At the same time, the coupling between AFS and DYC is considered. With the real-time slip rate as the control objective, the fuzzy PID controller is used to obtain the integrated control factor, and the dynamic load distribution of the driving torque of the four wheels is carried out, then, the lateral stability of the vehicle under the limit condition of high-speed driving is improved, and the effectiveness of the coordinated control strategy is verified on the simulation platform.

2 Vehicle Dynamics Modeling

2.1 *Vehicle Dynamics Model with Seven Degrees of Freedom (DOF)*

Two assumptions are put forward when establishing the vehicle dynamics model because this study mainly considers the longitudinal, lateral and yaw motions of the vehicle in the horizontal plane [12]: (1) Suppose the vehicle is running on a flat two-dimensional road, the vertical movement of each wheel is not considered, and the impact in the vertical direction is ignored; (2) The vehicle body and chassis are simplified as only the rotation in the pitch and roll directions; establish a seven-DOF steering dynamic model [13] as shown in Fig. 1a.

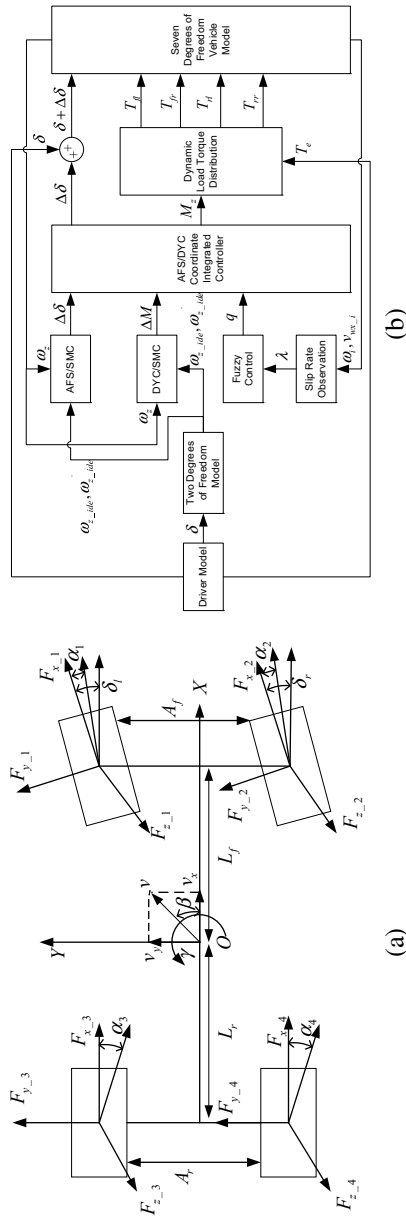


Fig. 1 a Seven degrees of freedom car model; b vehicle control strategy

$$\left\{ \begin{array}{l} M(\dot{v}_x - v_y \omega_z) = F_{x_{-1}} \cos \delta_l + F_{x_{-2}} \cos \delta_r - F_{y_{-1}} \sin \delta_l \\ \quad - F_{y_{-2}} \sin \delta_r + F_{x_{-3}} + F_{x_{-4}} \\ M(\dot{v}_y + v_x \omega_z) = F_{x_{-1}} \sin \delta_l + F_{x_{-2}} \sin \delta_r + F_{y_{-1}} \cos \delta_l \\ \quad + F_{y_{-2}} \cos \delta_r + F_{y_{-3}} + F_{y_{-4}} \\ I_z \dot{\gamma} = \frac{A_r}{2} (F_{x_{-4}} - F_{x_{-3}}) + \frac{A_f}{2} \left(\begin{array}{l} F_{x_{-2}} \cos \delta_r - F_{x_{-1}} \cos \delta_l \\ -F_{y_{-2}} \sin \delta_r + F_{y_{-1}} \sin \delta_l \end{array} \right) \\ \quad - L_r (F_{y_{-3}} + F_{y_{-4}}) + L_f \left(\begin{array}{l} F_{x_{-1}} \sin \delta_l + F_{y_{-1}} \cos \delta_l \\ +F_{x_{-2}} \sin \delta_r + F_{y_{-2}} \cos \delta_r \end{array} \right) \end{array} \right. \quad (1)$$

In Eq. (1): M is the mass of the vehicle; I_z is the moment of inertia of the vehicle around the z -axis; δ_i is the front wheel turning angles; β is the vehicle center of mass slip angle; L_i is the distances from the center of mass of the vehicle to the axles; A_i is the wheelbase of the wheels; γ is the yaw rate; v_x is the longitudinal speed; v_y is the lateral speed; F_x and F_y are the longitudinal and lateral force on the tire, respectively.

During the driving process of the vehicle, the tire load will be transferred, that is, the vertical load force of the tire. The calculation equation is expressed as follows:

$$\left\{ \begin{array}{l} F_{z_{-1}} = m_2 g + \frac{m_1 (g L_r - h a_x)}{2(L_f + L_r)} - \frac{m_1 h a_y}{2A_f}, \quad F_{z_{-2}} = m_2 g + \frac{m_1 (g L_r - h a_x)}{2(L_f + L_r)} + \frac{m_1 h a_y}{2A_f} \\ F_{z_{-3}} = m_2 g + \frac{m_1 (g L_f + h a_x)}{2(L_f + L_r)} - \frac{m_1 h a_y}{2A_r}, \quad F_{z_{-4}} = m_2 g + \frac{m_1 (g L_f + h a_x)}{2(L_f + L_r)} + \frac{m_1 h a_y}{2A_r} \end{array} \right. \quad (2)$$

In Eq. (2): $F_{z_{-i}}$ is the vertical load force exerted by the ground on the wheel of the vehicle, respectively; a_i is the accelerations of the vehicle's center of mass; h is the height of the center of mass; m_1 and m_2 are the body weight and tire weight, respectively; g is the gravitational acceleration.

2.2 Drive System Model

The distributed drive electric vehicle drive system uses a permanent magnet synchronous in-wheel motor, and the torque balance equation of the rotational motion of each nonlinear wheel is expressed as follows:

$$I_w \dot{\omega}_i = T_{di} - F_{z_{-wi}} f R_w - F_{x_{-wi}} R_w \quad (3)$$

In Eq. (3): I_w is the moment of inertia of the tire; f is the rolling resistance coefficient of the tire.

The whole vehicle model is driven by the permanent magnet synchronous wheel motor. Because the research content of this study is the yaw stability of the vehicle in the running state, the driving motor and battery are not the focus of this study. The motor model used in this study is a simplified model that can mirror the features of the motor. The motor characteristic curve will determine the motor output.

2.3 Tire Model

The tire model will use magic formula tire model [14]. Through the driving status of the vehicle, the longitudinal force and lateral force of the four wheels are obtained, and then the side slip angle and wheel slip ratio of the wheels are calculated. The mathematical equation of the slip angle is as follows:

$$[\alpha_1, \alpha_2, \alpha_3, \alpha_4] = \left[-\left(\delta_1 - \arctan \frac{v_{y_1}}{v_{x_1}} \right), -\left(\delta_2 - \arctan \frac{v_{y_2}}{v_{x_2}} \right), \arctan \frac{v_{y_3}}{v_{x_3}}, \arctan \frac{v_{y_4}}{v_{x_4}} \right] \tag{4}$$

In Eq. (4): α_i represents the side slip angle, v_{x_i} and v_{y_i} represent the longitudinal and lateral speeds of the tire in the vehicle coordinate system, respectively.

The longitudinal slip rate λ_i of each wheel can be defined as follows:

$$\lambda_i = (\omega_i R_w - v_{wx_i}) / \max(\omega_i R_w, v_{wx_i}) \tag{5}$$

In Eq. (5): λ_i represents the slip rate λ ; ω_i represents the rotational angular velocity; R_w represents the radius of rotation of the wheel.

3 Vehicle Controller Design

The vehicle control strategy is shown in Fig. 1b.

3.1 Ideal Reference Model

In this study, the two-DOF vehicle model can be used to calculate the ideal parameters. When the vehicle is in the steering state, the tire will produce a limit value due to the influence of the ground adhesion limit, and then the vehicle's yaw rate will be limited. Therefore, the expected yaw rate must satisfy Eq. (6).

$$|\omega_{\max}| = \kappa(\mu g/v_x) \quad (6)$$

In Eq. (6): κ is the safety factor; μ is the road adhesion coefficient; ω_{\max} is the maximum yaw rate.

The corrected ideal yaw rate can be obtained:

$$\omega_{z_ide} = \min\left(\left|\frac{v_x}{L(1 + Kv_x^2)}\delta_f\right|, |\omega_{\max}|\right) \cdot \text{sgn}(\delta_f) \quad (7)$$

3.2 Design of AFS and DYC Controller Based on SMC

The AFS and DYC controllers will design a SMC controller using the expected yaw angular velocity from Sect. 3.1 and the actual yaw angular velocity from the seven-DOF model of the vehicle. They will then solve for a stable additional front wheel rotation angle $\Delta\delta$ and additional yaw moment ΔM .

There are four typical reaching laws in sliding mode control, that is, the moving point reaches the switching surface in a finite time. In this study, the constant velocity approach law with the advantages of good robustness, strong real-time performance and small amount of computation will be selected [15].

$$\dot{s} = -\varepsilon \cdot \text{sgn}(s), \quad \varepsilon > 0 \quad (8)$$

In Eq. (8): ε is the approach law constant.

The yaw rate tracking error and its derivative and the sliding mode switching surface function are defined as follows:

$$e_{\omega z} = \omega_z - \omega_{z_ide}, \quad \dot{e}_{\omega z} = \dot{\omega}_z - \dot{\omega}_{z_ide}, \quad s_{\omega z} = c_{\omega z}e_{\omega z} + \dot{e}_{\omega z} \quad (9)$$

In Eq. (9): $c_{\omega z}$ is the relative weight coefficient, and $c_{\omega z} > 0$.

Replace sign function $\text{sgn}(s_{\omega z})$ with saturation function $\text{sat}(s_{\omega z})$ can reduce chattering in sliding mode control, and $\Delta\delta$, ΔM can be expressed as follows:

$$\begin{aligned} \Delta\delta = & \int \frac{I_z}{L_f k_f} \left(c_{\omega z} \dot{e}_{\omega z} + \frac{L_f k_f - L_r k_r}{I_z} \dot{\beta} + \frac{L_f^2 k_f + L_r^2 k_r}{I_z v_x} \dot{\omega}_z \right. \\ & \left. - \frac{L_f k_f}{I_z} (\dot{\delta}_f + \Delta \dot{\delta}) - \ddot{\omega}_{z_ide} + K_\omega \text{sat}(s_{\omega z}) \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \Delta M = & \int -I_z \left(c_{\omega z} \dot{e}_{\omega z} + \frac{L_f k_f - L_r k_r}{I_z} \dot{\beta} + \frac{L_f^2 k_f + L_r^2 k_r}{I_z v_x} \dot{\omega}_z \right. \\ & \left. - \frac{L_f k_f}{I_z} \dot{\delta}_f - \ddot{\omega}_{z_ide} + K_\omega \text{sat}(s_{\omega z}) \right) \end{aligned} \quad (11)$$

3.3 AFS/DYC Integrated Control Weight Distribution

The AFS/DYC integrated control prioritizes the wheel longitudinal force for stability control before considering the lateral force control. Its joint control strategy consists of passing the slip rate observation module, using the slip rate to control the target, and using the fuzzy PID controller to obtain the integrated control factor. This allows it to perform AFS compensation, control the use ratio of DYC, and finally obtain the additional yaw moment and additional front turning angle. In this study, the fuzzy and PID controllers will be combined, and a parameter self-tuning fuzzy PID controller with strong robustness will be used to control the slip rate. The adaptability and precise control of the system can be achieved at the same time.

3.4 Dynamic Load Torque Distribution

In this study, the dynamic load method is used to distribute the torque of the wheels. The driving force and additional yaw moment are distributed to each wheel in accordance with the dynamic vertical load proportion of each wheel while driving. The distribution method is shown in Eq. (12).

$$\begin{aligned} T_{fl} &= \tau_1 \left(T - \frac{2\Delta Mr}{B} \right), & T_{fr} &= \tau_2 \left(T + \frac{2\Delta Mr}{B} \right), \\ T_{rl} &= \tau_3 \left(T - \frac{2\Delta Mr}{B} \right), & T_{rr} &= \tau_4 \left(T + \frac{2\Delta Mr}{B} \right) \end{aligned} \quad (12)$$

In Eq. (12): T_e is the torque; r is the rolling radius of the vehicle tire; B is the distance between the tires; ΔM is the additional yaw moment.

4 Simulation and Result Analysis

Under the MATLAB/Simulink simulation platform, the vehicle model and stability controller are built, and offline simulation analysis is performed to verify the effectiveness of the controller. The initial conditions are as follows: the longitudinal vehicle speed is 100 km/h, the road adhesion coefficient is 0.9, the sine delay is selected as the steering condition, the front wheel rotation angle amplitude is 0.105 rad, and the simulation time is 14 s. When the vehicle is steered under the sinusoidal delay condition, the simulation results are shown in Fig. 2.

Figure 2a shows that the additional front wheel angle output under AFS system is smaller than the additional front wheel angle under SMC, and has the same trend with the front wheel angle steering input, indicating that compared with SMC, the control strategy proposed in this paper can have better tracking control performance

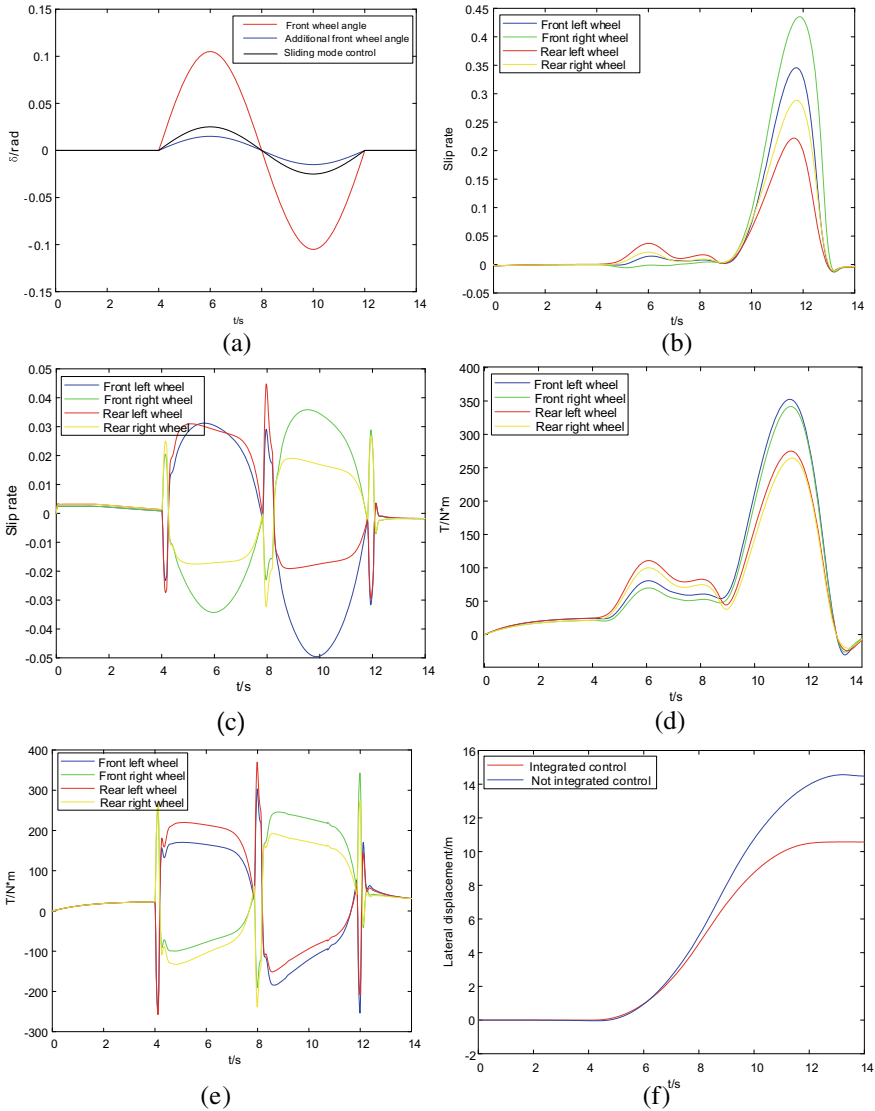


Fig. 2 **a** Front wheel corner and additional front wheel corner; **b** slip rate without applying integrated control; **c** slip rate under integrated control; **d** torque distribution without applying integrated control; **e** torque distribution under integrated control; **f** lateral displacement; **g** longitudinal speed; **h** comparison of yaw rates under various controllers

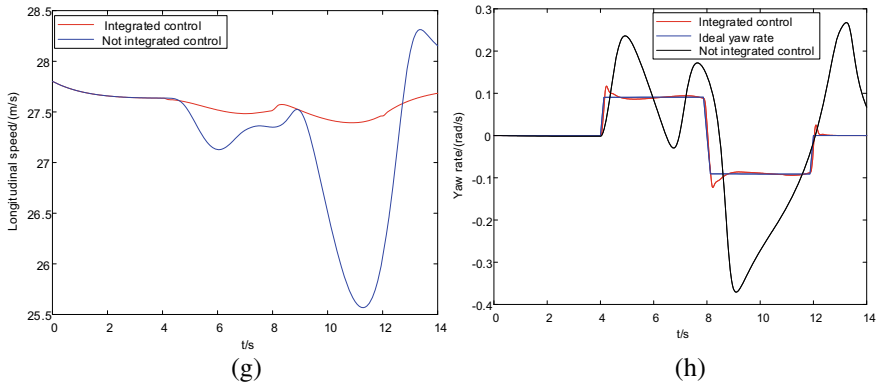


Fig. 2 (continued)

when the external angle input changes little. In order to achieve the goal of tracking control of steering angle control compensation and correction. Figure 2b–e shows that without integrated control, the slip ratio and torque distribution do not match the driving state conditions of the actual vehicle. In particular, at the tenth second, the slip ratio and the torque of each wheel increase sharply, and the slip ratio value increases. Evidently, the slip rate is larger than that under the integrated control. The vehicle becomes unstable at this time, which has a great impact on the overall driving safety of the vehicle. Under the integrated control system, the slip ratio of each wheel changes according to the driving state of the vehicle, and the value of the slip ratio is significantly smaller than that under the condition of no integrated control. The driving force of each wheel are actively distributed in real-time to meet the requirements of stable driving of the vehicle and to improve the lateral stability of the vehicle at high speed.

It can be seen from Fig. 2f–h that compared with the unintegrated control, the integrated control improves the lateral displacement, longitudinal vehicle speed and yaw velocity under the sliding mode and fuzzy composite control. By Fig. 2f, g you can see, not integrated control system, the vehicle lateral displacement and vertical velocity fluctuation appeared in the first ten seconds, especially between the longitudinal velocity in 9–13 s appeared the phenomenon of accelerated the speed down, when the vehicle is under a state of wheel lock skid, make its lateral displacement fluctuate, it is the instability of the vehicle performance, which eventually lead to the vehicle control, and affect driving safety. Under the integrated control, the lateral displacement tends to be stable, and the longitudinal speed can be well maintained, keeping the driver’s original steering feeling. As can be seen from Fig. 2h, the actual yaw velocity under the integrated control is closer to the ideal value, while the error between the actual yaw velocity and the ideal value under the non-integrated control is large. Compared with the non-integrated control, the average error of vehicle yaw velocity is reduced by 65.14%.

5 Conclusions

A coordinated control strategy for integrating AFS system and DYC system of four-wheel hub motor-driven electric vehicles is proposed to address the issue of lateral instability of four-wheel drive electric vehicles at high speed. The AFS system and the DYC system based on SMC are designed. The real-time dynamic load is used to realize the effective distribution of the driving torque of each wheel. The fuzzy PID controller is used to determine the integrated control factor and allocate the controller weight using the slip rate as the control objective. Under sine delay steering condition, the simulation results of integrated control and non-integrated control show that AFS/DYC coordinated integrated control can effectively improve the yaw stability of four-wheel drive vehicles under high-speed driving conditions. When a vehicle loses stability, adding a front wheel angle and yaw moment allows the vehicle to regain stability, and the actual yaw rate can better track its ideal value. The average vehicle yaw rate error is reduced by 65.14%, effectively improving the lateral stability of the vehicle under extreme conditions and ensuring safety.

Acknowledgements This work has been supported by the National key R&D program (No. 2022YFE0101000).

References

1. Yu Z, Feng Y, Xiong L (2013) Review of the development status of distributed drive electric vehicle dynamics control. *Chin J Mech Eng* 49(08):105–114
2. Xu K, Luo Y, Yang J (2019) Review of state perception and control of distributed electric vehicles. *Chin J Mech Eng* 55(22):60–79
3. Luo J, Luo Y, Chu W (2014) Distributed electric vehicle system/drive force coordinated driving stability control. *J Tsinghua Univ (Nat Sci Ed)* 54(06):729–733
4. Zhang X, Zheng L (2018) Distributed drive electric vehicle stability control based on torque coordination distribution. *China Mech Eng* 29(15):1780–1787
5. Xiong L, Li Y, Leng B (2017) Handling and stability control of distributed drive electric vehicle based on motion tracking. *J Tongji Univ* 45(Suppl. 1):53–57
6. Wanner D, Wallmark O, Jonasson M (2015) Control allocation strategies for an electric vehicle with a wheel hub motor failure. *Int J Veh Syst Model Test* 10(3):263–287
7. Park J, Park Y (2016) Optimal input design for fault identification of over actuated electric ground vehicles. *IEEE Trans Veh Technol* 65(4):1912–1923
8. Metzler M, Tavernini D, Sorniotti A (2019) Explicit nonlinear model predictive control for vehicle stability control. In: *Proceedings of the 9th international Munich chassis symposium*. Springer Nature, Munich, pp 733–752
9. Yan Y, Zhang Z, Xu X (2019) Neural network PID stability control of independent drive electric vehicles. *Mech Sci Technol* 38(10):1598–1605
10. Zhang L, Yu W, Wang Z (2020) Fault-tolerant control strategy of four-wheel in-wheel motor-driven electric vehicle based on multi-method switching. *Chin J Mech Eng* 56(16):227–239
11. Peng B, Zhang H, Xuan F (2018) Torque distribution strategy of electric vehicle with in-wheel motors based on the identification of driving intention. *Automot Innov* 1:140–146
12. Qin Y, Zhao Z, Wang Z (2021) Study of longitudinal-vertical dynamics for in-wheel motor-driven electric vehicles. *Automot Innov* 4:227–237

13. Li Z, Ran T, Sun C (2021) Multi degree of freedom vehicle modeling and simulation analysis of electric vehicles. *Automot Pract Technol* 46(01):4–7
14. Bhatt AH, Shirodkar PS (2020) Lateral force modelling using magic formula tire model. In: Vasudevan H, Kottur V, Raina A (eds) *Proceedings of international conference on intelligent manufacturing and automation*. Lecture notes in mechanical engineering. Springer, Singapore. https://doi.org/10.1007/978-981-15-4485-9_74
15. Jeong S, Chwa D (2021) Sliding mode disturbance observer based robust tracking control for omnidirectional mobile robots with kinematic and dynamic uncertainties. *IEEE/ASME Trans Mechatron* 26(2):741–752

A Study on the Construction of Crime Prevention System Using the Hadoop Model of Big Data- with Focus on Utilizing Big Data and CPTED Technology in Gimcheon-Si -



Bong-Soo Kim , Dong-wook Kang , Eun-Joo Lee , Dae-bum Lee , and Ji-Woong Ryu 

Abstract The crime rate in Gimcheon through the analysis of big data in 2020 was 3933 cases in Gimcheon City in 2019, up 5.2% from 3740 cases in the previous year. If big data managed by the police is used as information sharing and cooperation among agencies through a common network to reflect it in various policies, it can prevent duplicate benefits if a crime prevention system can be established. Therefore, it is expected to help cover more policies to create a safe city, which could effectively be reflected in budget allocation for supporting safe city. And there is a need to conduct continuous management of and violence education with public institutions or private institutions. In this paper, the prediction model of big data collects basic crime data and Hadoop-based big data and applies the model with the best prediction accuracy by applying data preprocessing and characteristic engineering. The prediction algorithm constituting the prediction engine applies the deep learning RNN algorithm.

Keywords Crime prevention education · Safety Village School · The school violence prevention education · Big data · CPTED

B.-S. Kim (✉) · E.-J. Lee
Department of Police Administration, Gimcheon University, Daehak-Ro 214,
Gyeongsangbuk-Do, 39528 Gimcheon-Si, South Korea
e-mail: sbong78@naver.com

E.-J. Lee
e-mail: amy91012@hanmail.net

D. Kang · J.-W. Ryu
Department of Law Dongguk University, Dongguk University, 30, Pildong-Ro 1-Gil, 04620
Jung-Gu, South Korea
e-mail: dwkang@dongguk.edu

J.-W. Ryu
e-mail: dayou1111@dongguk.edu

D. Lee
Department of Business Administration, Mokwon University, 35349 Doanbuk-Ro 88 Seo-Gu,
Daejeon Metropolitan City, South Korea
e-mail: dblee@mokwon.ac.kr

Table 1 State of crimes by big data

	Overall crimes	Violent criminal	Burglar	Intellectual
2019	3933	67	416	730
2018	3740	65	376	691

1 First Section

In 2019, the five major crimes directly related to public welfare in Gimcheon City increased by 15.4% in 1121 cases a year earlier, 1 murder, 2 cases a year earlier, 100% reduction in 1 case, 418 cases of river and theft, 380 cases a year earlier, 10% increase in 38 cases, 825 cases of violence, 692 cases a year earlier, 19% increase in 133 cases of sex crimes, 58 cases a year, 1% a year.

Violence is not a problem limited only to the victims of violence. Obviously, adolescent violence is imparting enormous adverse effects not only to the victims who incurred physical and mental damages but also to their families.

Therefore, there is a need to conduct continuous management of and violence education with public institutions or private institutions.

2 The State of Crimes Between 2018 and 2019 by Big Data in Gimcheon

In 2020, the number of crimes committed through the analysis of big data in Gimcheon increased 5.2% to 3933 in Gimcheon compared to 3740 in the previous year. Except for special criminals, violent criminals (3%), burglars (11%), intelligent criminals (6%), increased (Table 1).

Based on the results of regional diagnosis and the number of local police stations, two types of crime characteristics (five major crimes in 2019 and the number of reported 112 cases), seven types of population and sociological characteristics (such as criminals, police agencies, foreigners, population density, basic recipients), and the results of local police stations, the risk is calculated and determined to be reversed, central, western, northern, and Yulgok (Table 2).

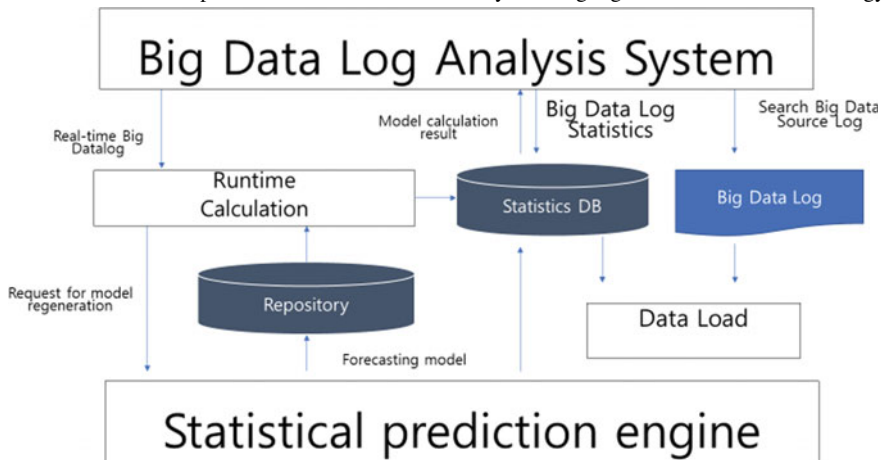
3 The Prevention of Adolescent Crime by Utilizing Big Data and CPTED Technology

Table 3 using Hadoop, a comprehensive crime prevention system using big data such as log data, unstandardized text, sensor data, geographic space information data, and social data will be established.

Table 2 Results of regional diagnosis through big data analysis of CPO (Crime Prevention and Diagnosis Team) in Gimcheon police station in 2020

	Average	Gimcheon station	Center	Western	North	Yulgok
Five major crimes	92.4	145	194	165	226	206
112 report	1855	3124	4062	3628	5172	3162
Second offender	15.7	29	25	37	29	25
Crime-targeted establishment	32.3	142	78	87	53	32
Foreigner	164	114	128	154	898	102
Population density	10,083	7940	11,269	20,529	24,736	28,139

Table 3 Structure of prevention of adolescent crime by utilizing big data and CPTED technology



If big data managed by the police is used as information sharing and cooperation among agencies through a common network to reflect it in various policies, it can prevent duplicate benefits if a crime prevention system can be established.

Therefore, it is expected to help cover more policies to create a safe city, which could effectively be reflected in budget allocation for supporting safe city.

The prediction model of big data collects basic crime data and Hadoop-based big data and applies the model with the best prediction accuracy by applying data preprocessing and characteristic engineering. The prediction algorithm constituting the prediction engine applies the deep learning RNN algorithm.

4 Conclusion

The development of crime prevention programs through big data and CPTED technology can be effectively applied to juvenile crime prevention. All types of data including standardized data, semi-standardized data, and unstandardized data must first go through a collection and integration process. Securing original data can be done by using various tools. Collection tools such as Flume, Sqoop, Scribe, and Chukwa can convert and link various data dispersed in the server to then save it on the Hadoop file system or database safely and without any overlap. In particular, the various unstandardized data from reports and consultation logs, messages, and social networks on smart phone applications currently used by public institutions can be collected. Data, after it is collected, is processed and saved. It must be determined where collected data is saved according to its urgency and nature. A bulk storage system, the Hadoop distribution system, is used to save unstandardized data. Semi-standardized data in the form of transactions are saved on NoSQL databases such as HBase, MongoDB, and Casandra which allows for real-time searches. The MapReduce15 framework is utilized to process placement data and give structure to the whole or part of the data. It is not easy to independently program with MapReduce provided by Hadoop. In the case of standardized data, Hiv which is similar to the SQL type must be used, while Pig is used for unstandardized data. Data analysis is a data mining step which is done through data analysis modeling.

In order to do so, a statistical analysis algorithm, similar and length analysis algorithm for data relativeness analysis, data classification, machine learning algorithm for cluster analysis, content recommendation algorithm based on a cooperative filter, etc., are used. Recently, the open source *R* for big data statistical analysis on big data is receiving the spotlight and the Mahout data analysis tool which uses machine learning based on MapReduce is commonly used. This step provides data that has been mined in various visualized forms such as charts, bars, and maps.

Therefore, there is a need to conduct continuous management of and violence education with public institutions or private institutions.

In this paper, the prediction model of big data collects basic crime data and Hadoop-based big data and applies the model with the best prediction accuracy by applying data preprocessing and characteristic engineering. The prediction algorithm constituting the prediction engine applies the deep learning RNN algorithm.

References

1. Korea National Police, White Paper of Police (2018)
2. Korea National Police, White Paper of Police (2019)
3. Kim HS, Joo HJ, Kim HR (2016) Big data base. Infinite

4. Lee JY (2012) A study on the development of cooperation system between police and victim support center for criminal-victim protection and support. Soc Sci Res Dongguk Univ Instit Soc Sci 19(3):34
5. Kang DW, Song GC (2015) Problems and improvement of the crime victim support system. Victimol Study 23(2):76

Application of Multi-variable Double-Layer MPC in Energy-Saving Operation of Central Chiller



Xiang Li and Kaibi Zhang

Abstract This paper proposes an application of the multivariate double-layered MPC approach in the energy-saving operation of central chiller. The model does not need to know the steady-state operating point of the system. The supply water, return water, and indoor temperature in the chiller are considered. The model predictive control method cannot only meet the cooling capacity required by the building but also improve the energy-saving efficiency of the chiller. The open-loop prediction model, the steady-state target calculation (SSTC) module, and the dynamic control module are the same as in double-layered dynamic matrix control, but its details obey the state-space method. The offset-free control is realized when the system has modeling error and unmeasurable interference by adding interference. In the case of different disturbances, the stability of the system can be guaranteed. Simulations were carried out under different solar radiation intensity, room temperature, and outdoor temperature to verify the effectiveness of the algorithm.

Keywords Model predictive control (MPC) · Central chiller · Energy efficiency

1 Introduction

In large buildings, central chillers are usually used as refrigeration equipment. The research object of this case is a 5-story library equipped with a central chiller. The central chiller uses an evaporator to exchange heat between water and refrigerant. The refrigeration system absorbs the heat load in the water, cools the water, and provides cooling energy for the room. The power consumption of water-cooled chillers accounts for a large part of the total power consumption of the library. Therefore, under the premise of ensuring the demand for refrigeration, reducing energy consumption through energy management is a mainstream issue today.

X. Li (✉) · K. Zhang
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: 741324614@qq.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_22

179

In [1], the authors found the optimal chilled water flow rate and water supply temperature. Compared with the traditional control, the optimal control rule can reduce the total energy consumption of the chiller by 17.55%. The authors in [2] studied the fuzzy PID method of frozen pump frequency control and changed the frequency of frozen pump according to the change of user load to achieve the energy-saving goal. The authors in [3] propose an optimal scheduling scheme that can save system energy consumption under the condition of satisfying cooling load and system operation constraints and applies an improved genetic algorithm to this scheme. The results show that the system energy consumption can be reduced by about 7.4%. In [4], the author considers the relationship between temperature, chiller energy consumption, and room comfort and uses the MOPSO algorithm to provide a series of optimal control setting solutions for the central air-conditioning system and provides a trade-off between indoor comfort and energy consumption.

Model predictive control is a computer control strategy, which has been widely used in industrial processes [5]. The advantage of MPC lies in its ability to handle complex systems with multiple variables and constraints [6–8]. Under the condition of satisfying multiple constraints, it runs smoothly at the boundary of various constraints to achieve the maximum economic benefit and energy saving [9, 10]. This paper presents an application of a double-layer MPC without offset in the central chiller. This algorithm can improve the energy-saving effect of the central chiller in the case library. The contribution of this paper is to propose a control strategy, which can deal with (1) the unknowable situation of the steady-state operating point of the system, (2) and consider the MVs and CVs constraints of the process, so as to improve the thermal comfort and energy-saving potential of the building.

2 The Overall System Architecture

2.1 *Building and System Description*

The case study in this article is a 5-story library equipped with a central chiller. The screw unit of the central chiller can perform heat recovery in refrigeration conditions. The refrigeration host can meet the maximum demand for simultaneous use of the building. The water pump sends the chilled water that reaches the cooling temperature in the water tank to the place where it is needed and returns the chilled water to the evaporator. The cooling tower on the roof exchanges the cooling water carrying waste heat with the air in the tower. The library is equipped with monitoring equipment to record the inflow and outflow temperature of chilled water, room temperature, electricity consumption, etc.

2.2 Modeling of the System

In this case, the room temperature T_{room} and the refrigerated water return temperature $T_{w,r}$, respectively, represent the thermal comfort of the building and the heat exchange between the chilled water and the air. In the case of ensuring thermal comfort and energy efficiency, T_{room} and $T_{w,r}$ need to be maintained at a certain set value and controlled within a certain accuracy range.

For the research case, manipulation variable MV ($T_{w,s}$) and control variables CVs (T_{room} and $T_{w,r}$) need to be adjusted at the same time to make the system reach the desired state. In order to make this control strategy meet the comfort of the most severe area (the longest exposure time), disturbance is added to the model. Another reason for adding interference is that if there is unmeasurable interference or modeling error in the system, it is difficult to achieve offset-free control. Estimate the parameters of the model using the prediction error method [11]. The model is described as [12].

$$\begin{aligned}
 \begin{bmatrix} T_{w,r}(k+1) \\ T_{\text{room}}(k+1) \\ T_{\text{extra}}(k+1) \end{bmatrix} &= \begin{bmatrix} 0.305 & 0.045 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_{w,r}(k) \\ T_{\text{room}}(k) \\ T_{\text{extra}}(k) \end{bmatrix} + \begin{bmatrix} 0.585 \\ 0.822 \\ 0 \end{bmatrix} [T_{w,s}(k)] \\
 &+ \begin{bmatrix} 0.0503 & 0 & 0.0167 \\ 1 & 0.0149 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} T_{\text{amb}}(k) \\ I_{\text{global}}(k) \\ O_{cc}(k) \end{bmatrix}, \begin{bmatrix} T_{w,r}(k) \\ T_{\text{room}}(k) \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} T_{w,r}(k) \\ T_{\text{room}}(k) \\ T_{\text{extra}}(k) \end{bmatrix}, \tag{1}
 \end{aligned}$$

where $T_{w,s}$ is the temperature of supply water, $T_{w,r}$ is the temperature of return water, T_{room} is the temperature of the room, T_{extra} is the additional state, T is the temperature of the ambient, I_{global} is the intensity of solar radiation, and O_{cc} is the occupancy of the room.

3 Double-Layered MPC for State-Space Model

In this section, we follow the standard MPC to design the controller, including the augmented model and control move calculation and implementation.

The state equation of the system can be expressed as follows:

$$\begin{cases} \tilde{x}_{k+1} = \tilde{A}\tilde{x}_k + \tilde{B}\Delta u_k + \tilde{F}\Delta f_k \\ y_k = \tilde{C}\tilde{x}_k \end{cases}, \tag{2}$$

where \tilde{A} , \tilde{B} , \tilde{C} , and \tilde{F} are the constant matrices. x , f , u , and y are system states, measurable disturbance variables, manipulation variables, and control variables, respectively.

3.1 Open-Loop Prediction

The open-loop prediction considers the prediction when the control action is unchanged and uses the output value at the current moment to predict the future output. For the augmented model described by (2), adopting the Kalman filter obtains the following estimation of augmented state:

$$\begin{aligned} \begin{bmatrix} \hat{x}'_{k|k+1} \\ \hat{y}'_{k|k+1} \end{bmatrix} &= \tilde{A} \begin{bmatrix} \hat{x}'_{k-1|k-1} \\ \hat{y}'_{k-1|k-1} \end{bmatrix} + \tilde{B} \Delta u_{k-1} + \tilde{F} \Delta f_{k-1}, \\ \begin{pmatrix} \hat{x}_{k|k} \\ \hat{y}_{k|k} \end{pmatrix} &= \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_{k|k-1} \end{bmatrix} + \begin{bmatrix} L_x \\ L_y \end{bmatrix} \left(y_k - \tilde{C} \begin{bmatrix} \hat{x}'_{k|k-1} \\ \hat{y}'_{k|k-1} \end{bmatrix} \right), \end{pmatrix} \end{aligned} \quad (3)$$

where the state filtering gain L is divided into process model state filtering gain L_x and state output filtering gain L_y . The closed-loop control move is described by

$$\Delta u(k+i|k) = K \Delta \hat{x}(k+i|k) + \Delta v(k+i|k), \quad i \geq 0, \quad (4)$$

where K is the feedback gain matrix such that $A_c = A + BK$ is asymptotically stable, and Δv is the perturbation item. Define $\Delta u^{ol}(k+i|k) = K \Delta \hat{x}^{ol}(k+i|k)$ as an open-loop control move. The so-called open-loop forecast value is an estimate of the future value under the action of open-loop control and $\Delta f(k+j) = 0(j > 0)$. Based on the $i \geq 1$ step ahead Kalman prediction, open-loop steady-state equations of input and output can be obtained.

3.2 Steady-State Target Calculation

Since the disturbances or manual intervention of the operator in any control cycle may change the optimal target of the process, SSTC calculates the steady-state target value in each control cycle. SSTC is an open-loop optimization method for calculating the local optimal steady-state target of the process. This method combines the steady-state mathematical model of the process, considers the input and output constraints of the process, sets the objective function according to the specific requirements, and finally forms the LP problem or QP problem.

The following magnitude and rate constraints will be considered for SSTC:

$$\begin{aligned} \underline{u} \leq u_{ss}(k) \leq \bar{u}, |\delta u_{ss}(k)| \leq N_c \Delta \bar{u}, |\delta u_{ss}(k)| \leq \delta \bar{u}_{ss}, \\ y_{o,h} \leq y_{ss}(k) \leq \bar{y}_{o,h}, y_o \leq y_{ss}(k) \leq \bar{y}_o, |\delta y_{ss}(k)| \leq \delta \bar{y}_{ss} \end{aligned} \quad (5)$$

where $\underline{y}_{o,h} \leq \underline{y}_o$, $\bar{y}_{o,h} \geq \bar{y}_o$, and $\bar{v}(v)$ are the upper and lower limit of the variable, v_{ss} is the steady-state value of the variable, and $\delta v_{ss}(k) = v_{ss}(k) - v(k-1)$.

In order to calculate the steady-state targets of MVs and CVs, the feasibility analysis of the SSTC module is required. We keep the hard constraints unchanged, and relax some soft constraints to make the feasible region solvable. In each priority rank, the soft constraints in this rank may be relaxed so as to be consistent with all the hard constraints.

3.3 Dynamic Calculation Module

Since the control horizon is N_c , $\Delta v(k+i|k) = 0$ holds for all $i \geq N_c$. Based on the Kalman prediction method, the output predictions influenced by the current and future control moves are determined by

$$\Delta \hat{x}(k+1|k) = A_c \Delta \hat{x}(k|k) + B \Delta v(k|k) + F \Delta f(k), \quad (6)$$

$$y(k+i|k) = y(k) + \sum_{j=1}^i C \Delta \hat{x}(k+j|k), \quad i = 1, \dots, N. \quad (7)$$

In the dynamic calculation module, usually the following constraints (i.e., MV rate constraint, MV magnitude constraint, CV magnitude constraint, constraint on slack variable) are considered:

$$|\Delta u(k+i|k)| \leq \Delta \bar{u}, \quad 0 \leq i \leq N_c - 1, \quad (8)$$

$$u \leq u(k+i|k) \leq \bar{u}, \quad 0 \leq i \leq N_c - 1, \quad (9)$$

$$y'_o(k) \leq y(k+i|k) \leq \bar{y}'_o(k), \quad 1 \leq i \leq N. \quad (10)$$

In summary, the following optimization problem, at each time K , is solved:

$$\min_{\Delta \tilde{v}(k|k)} \sum_{i=1}^N \|y(k+i|k) - y_{ss}(k)\|^2 + \sum_{j=0}^{N_c-1} \|\Delta v(k+j|k)\|^2, \quad s.t(9) - (10). \quad (11)$$

With the optimal $\Delta \tilde{v}(k|k)$ being obtained, $\Delta u(k|k) = K \Delta \tilde{x}(k|k) + \Delta v(k|k)$ will be sent to the actual controlled system.

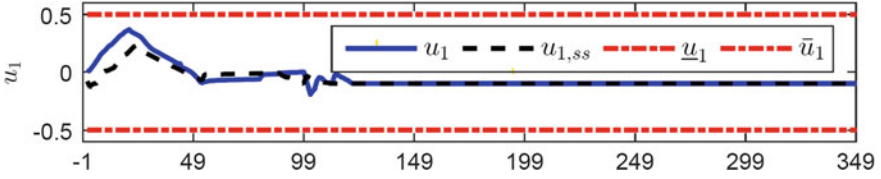


Fig. 1 Control input of closed-loop system

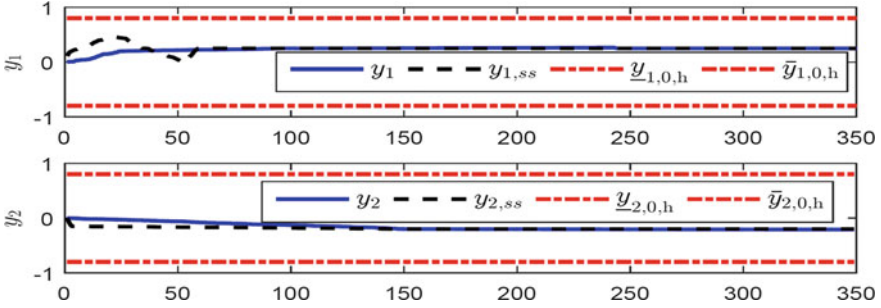


Fig. 2 Output responses in the presence of disturbance

4 Application Example

In this section, we verify the effectiveness of the MPC algorithm in the energy saving of the central chiller in this case.

The weighting coefficients in economic optimization is set as $h = [-2, -1]$. For intervals $k \in [62, 76]$ and $k \geq 120$, the disturbances with a magnitude of $f_{ea} + [-0.6; 0.6]$ and $f_{ea} + [-0.8; 0.8]$ enter the system, respectively.

The parameters of dynamic control are selected as $N=20$, $N_c = 15$, and $\Lambda = \text{diag}\{3, 5, 3\}$. The real plant is picked as $A_r = 0.9A$, $B_r = 0.9B$, $C_r = 0.9C$, and $F_r = 0.9F$. The real system output is multiplied by 0.9 to represent the mismatch between the model and the actual system.

The calculation result of the steady-state target is shown in the black dotted line in the figure. The output and the input results are shown in Figs. 1 and 2. From Figs. 1 and 2, we can see that the steady-state values of CV and MV given by SSTC can be fully tracked by dynamic control.

5 Conclusions

This paper proposes a double-layer MPC algorithm with offset-free and applies it to the energy-saving operation of the central chiller. The open-loop steady-state predicted values of MV and CV are obtained under the open-loop prediction module.

The dynamic control module applies the calculated optimal control move to the real plant. Under the steady-state target calculation module, various constraints are considered, and the cooling capacity of the central chiller is balanced with the actual cooling load under the condition of meeting the thermal comfort of the building, so as to minimize the operating cost.

References

1. Li A, Wang F, Li R (2017) The optimal control strategy for chilled water system in central air-conditioning systems. In: IECON 2017—43rd annual conference of the IEEE industrial electronics society, vol 10, pp 8150–8155
2. Li Z, Lv C, Mei J (2016) Research on energy saving control of chilled water system of central air conditioning system. In: Chinese Control and Decision Conference (CCDC), pp 2875–2878
3. Li M, Qing C (2010) Optimization for the chilled water system of hvac systems in an intelligent building. In: 2010 International conference on computational and information sciences, pp 889–891
4. Yao L, Jaiteh K (2017) Multi-objective control of central air conditioningsystem. *EEEIC/ICPS Europe*, pp 1–6
5. Qin SJ, Badgwell TA (2003) A survey of industrial model predictive control technology. *Control Eng Pract* 733–764
6. Li Y, Shen J, Zhang J (2019) Research on pressure and water level control of the pressurizer for marine nuclear power plant based on multivariable mpc. *ICPRE*, pp 284–289
7. Zhu H, Zhang X, Li M, Liu X, Zhang L (2020) Multi objective optimal control method for t-type three-level inverter based on model prediction, (ICIEA). *IEEE*, pp 753–760
8. Mayer B, Killian M, Kozek M (2017) Hierarchical model predictive control for sustainable building automation. *Sustainability* 264
9. Hu C, Zhao L (2021) Overtaking control strategy based on model predictive control with varying horizon for unmanned ground vehicle. *Proc Instit Mech Eng Part D: J Automobile Eng* 78–92
10. Pawitan GAH, Kim J-S (2020) Mpc-based power management of renewable generation using multiness guaranteeing soc constraints and balancing. *IEEE Access* 12897–12906
11. Lara BGV, Molina LMC, Yanes JPM (2015) Modeling and identification of the cooling dynamics of a tropical island hotel. *Energy Build* 19–28
12. Lara BGV, Molina LMC, Yanes JPM, Borroto MAR (2016) Offset-free model predictive control for an energy efficient tropical island hotel. *Energy Build* 283–292

Study on AEB Control Strategy Considering Vehicle Load Variation



Changhao Piao, Xu Han, Junren Shi, and Tai Liu

Abstract In order to improve the active safety performance of vehicles with frequent load changes such as new energy logistics vans, a hierarchical control strategy is proposed to fully consider the vehicle load changes while ensuring that wheel locking is avoided in the active braking process. In this paper, the upper layer controller firstly determines the current maximum braking torque of the vehicle based on fuzzy control rules, uses the least recursive squares method with the forgetting factor to estimate the load mass of the vehicle, and sets the collision time threshold in the TTC model as a dynamic value to adapt to the different load conditions of the vehicle according to the maximum deceleration value that can be achieved by the maximum braking torque under different load conditions. The lower-level controller adjusts the target braking torque through slip rate feedback and actual deceleration to achieve closed-loop torque control. It is proved by the joint simulation of CarSim and Simulink and the real vehicle test that the control strategy can effectively avoid the vehicle collision under the CCRs condition and the braking on time is more reasonable.

Keywords Automatic emergency braking · Load variation · Slip rate

1 Introduction

Automatic emergency braking system is a hot research topic in the field of active vehicle safety. It can monitor the driving environment in front of the vehicle in real time and automatically start the braking system to slow down the vehicle when there is possible collision danger, so as to achieve the purpose of avoiding or mitigating the collision. According to the research results of relevant scholars, AEB function can reduce traffic accidents by 27% [1].

C. Piao · X. Han · J. Shi (✉)

Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: shijr@cqupt.edu.cn

T. Liu

China Academy of Information and Communications Technology, Beijing 100083, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_23

187

Due to the differences in vehicle load, braking mode, and other factors between passenger cars and commercial vehicles, the AEB function of passenger cars cannot be directly applied to commercial vehicles. In order to establish an automatic emergency braking system suitable for commercial vehicles, Chen et al. [2] determined the pre-collision time by combining the deceleration distribution of commercial vehicles and the driver's response time to the alarm and realized the AEB system suitable for commercial vehicles. However, the determination of deceleration was obtained from a small number of driver samples, so it was not universal. Guo et al. [3] designed the AEB system control strategy based on BP neural network to predict the collision time according to the crash time data of different drivers under different emergency braking scenarios for semitrailer vehicles, which improved the accuracy of collision risk assessment. Bai et al. [4] studied the influence of liquid load on automatic emergency braking, analyzed the attenuation degree of braking efficiency when the load is liquid, and added the average deceleration and braking time into the control algorithm, so as to calculate the reasonable braking time. Other researchers [5–7] also conducted corresponding studies on the determination of braking timing and other issues of commercial vehicles.

The above researchers have proposed many different solutions to the problem of determining the braking timing, extending the idea of designing the AEB function for passenger vehicles to varying degrees. However, they did not take into account the fact that the change in gross vehicle weight between no load and full load is concentrated at 300 kg due to the small load fluctuation of passenger vehicles, and the impact on the braking deceleration under the same braking force is smaller compared to that of commercial vehicles. The insignificant change in vehicle load in the design of the AEB function for passenger vehicles becomes a key factor affecting the braking timing in commercial vehicles.

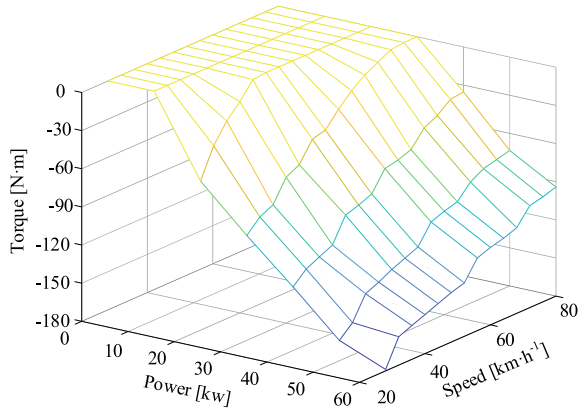
Taking the new energy logistics van as the research object, this paper considers the influence of vehicle load variation on braking timing determination, adjusts the key parameters in the second-order TTC model to dynamic values, and prevents wheel locking by limiting slip rate when AEB enters braking. The upper controller designed the vehicle mass estimation and the maximum braking torque limiting module, the lower controller used PID to control the braking torque, and the variable structure control was used to adjust the wheel slip rate. Finally, CarSim and Simulink are used to simulate typical AEB test scenarios and carry out real vehicle tests to verify the control strategy proposed in this paper.

2 System Architecture

2.1 Maximum Braking Torque Limiting Module

In order to improve the range of new energy vehicles, the battery pack can be charged by switching the working mode of the driving motor to the generator mode when

Fig. 1 Motor torque values based on fuzzy control



the vehicle is braking. In this paper, the braking force is generated by the way that the driving motor works backward to generate the drag torque which prevents the vehicle from moving forward, and the battery pack of the vehicle will be charged during the braking process. In order to ensure that the energy return during braking is within the maximum allowed by the battery management system, the braking torque should be limited. This paper uses fuzzy control theory to calculate the maximum braking torque. The vehicle speed and the maximum allowable charging power are taken as the input of the fuzzy control system to obtain the maximum braking torque in the current state. Mamdani is used for fuzzy relation rule, and centroid is used for fuzzy resolution. The surface diagram of input and output variables is shown in Fig. 1.

2.2 Vehicle Mass Estimation Module

In this paper, the recursive least square method with genetic factors is used to calculate the results of the measured data in a recursive way, so as to reduce the amount of data storage on the basis of ensuring the real-time performance of the data. At the same time, genetic factors are introduced to reduce the proportion weight of historical information to ensure the ability of new data to optimize vehicle quality. In order to make the estimated value of the vehicle converge as fast as possible, the recursive least square method with genetic factors is used when the vehicle starts every time, and the estimated mass of the vehicle can be obtained until the vehicle keeps constant speed.

According to the force analysis of the bicycle model, the longitudinal dynamic equation of the vehicle can be obtained, and the dynamic equation of the whole vehicle can be sorted into the form of RLS algorithm.

$$y_k = \frac{T_{tq} i_g i_0 \eta_t}{r_d} - \frac{C_D A}{21.15} u_a^2 \quad (1)$$

$$H_k = g f_R \cos i + g \sin i + \delta a \quad (2)$$

$$\hat{\theta}_k = m_k \quad (3)$$

In the above formula, the subscript k indicates the current sampling period, y_k is the system input quantity, T_{tq} is the drive shaft output torque, i_g is transmission ratio, i_0 is main reducer transmission ratio, η_t is integral power transmission efficiency, r_d is the wheel radius, and C_D is the air resistance coefficient; A is windward area, u_a is the speed of self-propelled vehicle, and m is the total mass of the vehicle; g is gravitational acceleration, f_R is the rolling resistance coefficient, i is the angle of ramp, δ is the conversion factor of automobile rotation mass, a is the longitudinal acceleration of the vehicle, and H_k is the observable data; $\hat{\theta}_k$ is the parameter to be identified at k time.

The whole vehicle weight of the new energy logistics van is 1510 kg, namely the vehicle no-load weight. The maximum allowable total mass is 2300 kg, which is the vehicle full-load mass. The vehicle mass 2000 kg is selected as the vehicle half-load mass. By collecting the real vehicle information, the motor torque and the speed of the real vehicle are taken as inputs, and the simulation module is built by Simulink to obtain the estimation results of the whole vehicle as shown in Fig. 2.

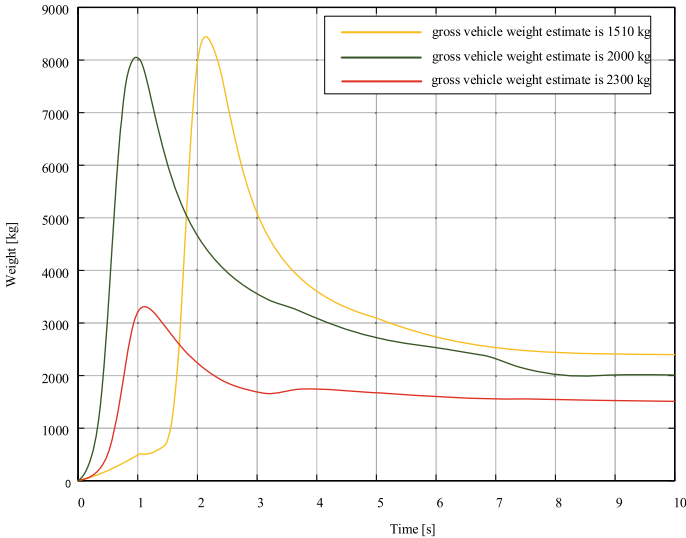


Fig. 2 Estimated overall vehicle mass using RLS when the vehicle load is no load, half load, and full load, respectively

The results show that the minimum recursive square method with forgetting factor is more accurate for vehicle mass estimation, the maximum error is 100 kg, and the estimated mass convergence can be completed in a short time. In the starting stage of the vehicle, a large torque is required to make the vehicle which has a small speed, resulting in an overestimation of the vehicle mass in the beginning stage. Compared with the method Zang et al. [8] used to determine the vehicle weight based on the engine output torque obtained by Kalman filter, the mass estimation error percentage of the method in this paper is 2.5%, 2.4%, and 4.1%, respectively, under different loads, and the estimation error of the vehicle mass is smaller.

2.3 Time Safety Model Considering Vehicle Mass

In this paper, the collision risk of two vehicles is evaluated by the time-based second-order TTC safety model [9], and the specific formula is as follows:

$$\text{TTC} = \begin{cases} -\frac{\Delta s}{\Delta v}, & \Delta v < 0, \Delta a = 0 \\ -\frac{\Delta s}{\Delta v} + \frac{\sqrt{\Delta v^2 - 2 \cdot \Delta s \cdot \Delta a}}{\Delta a}, & \Delta v < 0, \Delta a \neq 0 \\ -\frac{\Delta v}{\Delta a} + \frac{\sqrt{\Delta v^2 - 2 \cdot \Delta s \cdot \Delta a}}{\Delta a}, & \Delta v \geq 0, \Delta a < 0 \end{cases} \quad (4)$$

$$\text{FBTime} = \frac{\Delta v}{a_{\text{ego}}} \quad (5)$$

where Δs , Δv , Δa , and a_{ego} are, respectively, the longitudinal relative distance, longitudinal relative velocity, longitudinal relative acceleration, and self-deceleration of the two vehicles. In the traditional FBTime formula, the auto-vehicle deceleration is set to a fixed value, which has little impact when the body mass does not fluctuate greatly. However, for the van logistics vehicle, the load changes frequently and the fluctuation range is large, and the braking deceleration generated by the maximum braking torque is inconsistent at different speeds. Therefore, by calculating the ratio of the total force exerted by the vehicle to the estimated mass of the vehicle, the achievable deceleration rate of the vehicle was dynamically determined. Figure 3 depicts the maximum deceleration value achieved by different torques under three load conditions. As we have seen, at the same speed, the heavier the load, the less deceleration. Under the condition of the same load, the higher the speed, the smaller the deceleration.

2.4 Brake Control Module

In order to further improve the control accuracy, reduce the error and system interference, PID algorithm is used to control the vehicle brake. In order to ensure that the

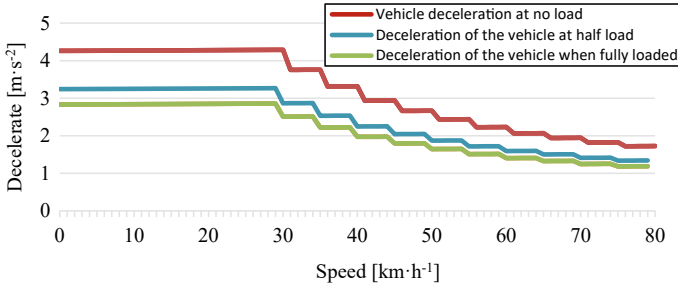


Fig. 3 Maximum deceleration of the vehicle under three different loads, considering the maximum braking torque

ground braking force is kept at the maximum value in the braking process without sideslip, it is necessary to adjust the braking torque to ensure the optimal slip rate in the braking process and avoid wheel lock. The Burckhardt model was adopted in the tire model, and the difference between the expected slip rate and the actual slip rate of the wheel was taken as the control variable. In order to eliminate steady-state errors, the sliding mode surface of proportional integral is designed, and the variable structure control of synovial film is carried out by exponential reaching law:

$$\frac{d(e + k_i \int e dt)}{dt} = -\varepsilon \operatorname{sgn}(s) - ks \quad (6)$$

At the same time, in order to eliminate chattering phenomenon, the sign function is replaced by a relatively smooth arctangent function, and the braking torque value is finally obtained as follows:

$$T_b = J \left[\mu g r \left(\frac{\omega}{u_a} + \frac{m}{J} \right) - u_a (\varepsilon \arctan(s) + ks) - k_i u_a e \right] \quad (7)$$

3 Simulation and Real Vehicle Test Results

In order to verify the effectiveness of the control strategy designed in this paper considering the load variation of the vehicle the test scenario proposed by JT/T 1242–2019 national standard. The AEB test scenario of logistics van was built in CarSim software, co-simulated with Simulink, and the real vehicle test was carried out. Test vehicles and props are shown in Fig. 4.

In the simulation, the initial speed of the front vehicle is set at 0 km/h and maintained, the initial speed of the self-vehicle is set at 40 km/h, and the initial position of the self-vehicle is 100 m in the same lane behind the front vehicle.



Fig. 4 Test vehicles and props

Figure 5 shows the simulation experiment results, the safe time model considering the load quality of the vehicle designed in this paper can make the vehicles under different loads' brake at different times, and the unloaded vehicles enter the brake at 2.8 s. The half-loaded vehicle enters the brake at 2.3 s. The fully loaded vehicle enters the brake at 2.1 s, which is consistent with the driving behavior that the heavier the load, the earlier the braking time. Moreover, since the braking force of new energy vehicles is provided by the motor and limited by the allowed charging power, in order to maximize the safety of drivers, without considering the changes in the vehicle load, the braking timing of the fixed fully loaded vehicle is compared with the control strategy considering the changes in the vehicle load.

The data obtained through the real vehicle experiment are plotted in Fig. 6. According to the analysis, although no-load and half-load vehicles can complete braking under the control strategy of fixed braking time threshold, the premature

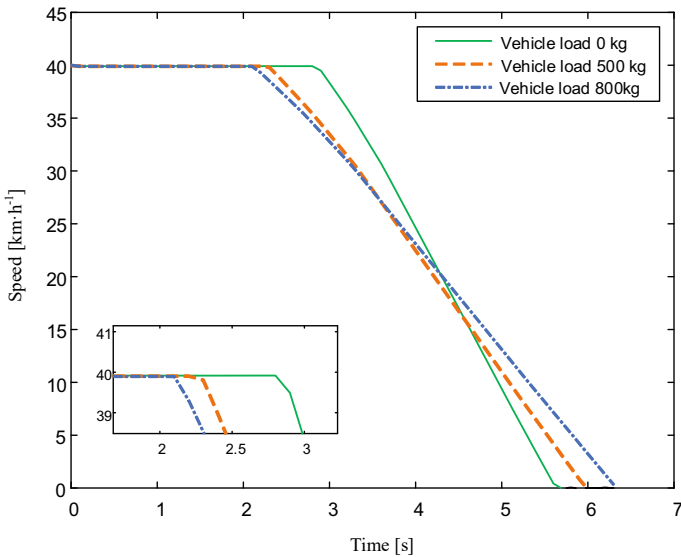


Fig. 5 Vehicle braking simulation experiments under different load conditions

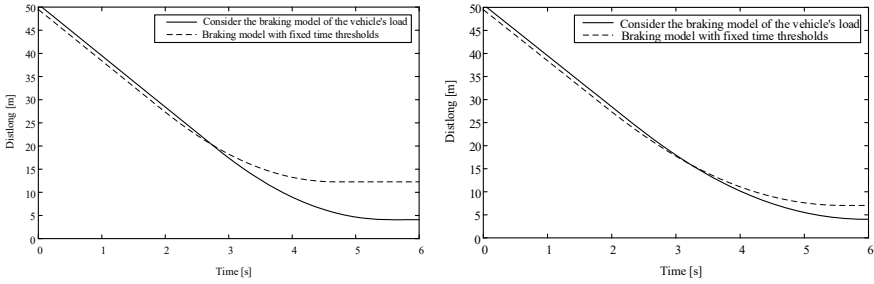


Fig. 6 Comparison of braking process under real vehicle test

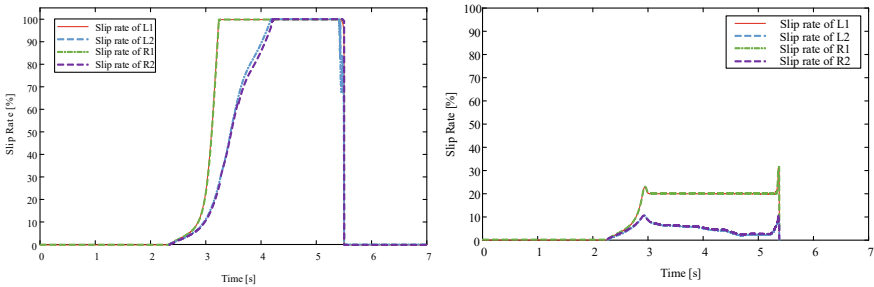


Fig. 7 Comparison of vehicle slip rate before and after variable synovial structure control

braking time leads to the distance from the front vehicle at the end of braking is 12.2 m and 7 m, respectively. On the premise of avoiding collision, the braking model considering the load of the whole vehicle makes the vehicle stop at 4 m away from the vehicle in front, which improves the acceptability of the driver. It proves that the control strategy designed in this paper can effectively avoid collision.

It can be seen from Fig. 7 that the slip rate of the vehicle will soar to 100% during the braking process. The control algorithm proposed in this paper can effectively reduce the slip rate and stabilize it at the target value.

4 Conclusion

This paper proposes an AEB control strategy that considers the variation of the whole vehicle load. Compared with the braking warning timing determined by considering the hazard factor and based on neural networks, etc., this paper takes into account the characteristics of the van's fluctuating body mass during transportation. A scheme is provided to estimate the vehicle load using recursive least squares with forgetting factors, determine the maximum deceleration rate that can be generated by the motor working in reverse for different weights of the vehicle, and determine the braking

timing in reverse as long as the maximum allowable charging power allows. A typical scenario of an automatic emergency braking control system test is built using CarSim and Simulink joint simulation, and the simulation test is carried out at different loads with new energy logistics vans as the test object to verify the effectiveness of the AEB control strategy considering the whole vehicle load variation. The results show that the control strategy designed in this paper can accurately estimate the overall vehicle mass and determine the appropriate braking time threshold based on it.

Acknowledgements This work has been supported by the National key R&D program (No.2022YFE0101000).

References

1. Xia L, Zhu XC, Ma ZX (2017) AEB test scenarios under cut-in dangerous conditions. In: The 14th international forum of automotive traffic safety 2017, pp 142–155
2. Chen T, Kaimin Z, Jiang ZF, Zhang Y (2022) Commercial truck AEBS test and research based on pedestrian crossing scenario. In: 2022 IEEE international conference on artificial intelligence and computer applications (ICAICA) 2022, pp 773–779
3. Guo XJ, Sun P, Deng J, Liu Y, Liu Z, Liu SP (2021) Research on AEB control strategy of heavy semi truck train based on BP neural network algorithm prediction. *J Automot Eng* 43(09):1350–1359+1366
4. Bai J, Tong J, Huang LB, Bi X, Jiang HT (2017) Influence of load variation on automatic emergency braking system of dangerous freight vehicle. *J Tongji Univ (Nat Sci Ed)* 45(S1):175–181+186
5. Ning MX, Wang SX, Ba TY, Tang XL (2022) AEB control algorithm and simulation of unmanned heavy vehicle. *J Chongqing Univ Technol (Nat Sci)* 36(06):72–80
6. E WJ, Ding, YC, Zhao P, Ding NG, Deng SF (2017) Study on autonomous emergency braking control strategy of dangerous goods transport vehicle. *Highway Traffic Technol* 34(S2):44–50
7. Gu ZQ, Hu SY (2019) AEB-oriented braking distance prediction and control method for commercial vehicles in downhill conditions. *Digit Manuf Sci* 17(04):276–280
8. Zang Z, Huo W, Wang YH, Li XK, Deng XG, Li YY (2020) Research on heavy commercial vehicle mass estimation algorithm. *Chin Mech Eng* 31(11):1360–1367
9. Han I, Luan B, Hsieh F (2014) Development of autonomous emergency braking control system based on road friction. In: 2014 IEEE International conference on automation science and engineering (CASE), Taipei, Taiwan, pp 18–22

Enhancing System Utilization by Dynamic Reallocation of Computing Nodes



Seungmin Lee, Hee Jin Jang, and Min Ah Kim

Abstract The field has expanded as supercomputers deal with different characteristics of workloads such as traditional scientific computing and data intensive computing. A batch queue-based Parallel Batch System (PBS) scheduler that manages high performance computing (HPC) tasks and a Kubernetes platform for managing data intensive applications are applied. Computing nodes are currently divided into static partitions for each workload. However, it provides better overall resource utilization of supercomputer as dynamically reallocating computing nodes to partitions according to the number of waiting jobs. In this work, we propose an approach to dynamic resource reallocation of computing nodes. We distinguish our approach from previous works in that our approach provides isolation of software stack and reallocates resources with a bare-metal environment suitable for preventing conflicts between two heterogeneous platforms. We considered node level reallocation which means that allocation is done to the partition in computing node level rather than sharing components of computing resources. A test scenario is used to demonstrate the process and feasibility of this approach and the result shows that it can improve system utilization.

Keywords Dynamic reallocation · Heterogeneous platform · Supercomputer

S. Lee (✉) · H. J. Jang · M. A. Kim

Korea Institute of Science and Technology Information, Daejeon 34141, Republic of Korea

e-mail: smlee76@kisti.re.kr

H. J. Jang

e-mail: jhj@kisti.re.kr

M. A. Kim

e-mail: petimina@kisti.re.kr

1 Introduction

The field has expanded as supercomputers deal with different characteristics of workloads such as traditional scientific computing and data intensive computing. Data intensive computing uses a data parallel approach to process huge data sets with a great diversity of types so that it requires deferent data processing approaches. A batch queue-based Parallel Batch System (PBS) scheduler that manages high performance computing (HPC) tasks for traditional scientific computing and a Kubernetes platform for managing data intensive applications, especially big data analysis, are applied in one of the largest supercomputer, NURION in KISTI [1].

Computing nodes are currently divided into static partitions to process HPC workloads and big data analytics workloads. However, computing nodes become idle when jobs are not enough for one partition, as a consequence waste resources if there are jobs waiting computing resources in another partition. It provides better overall resource utilization of supercomputer, therefore, as dynamically reallocating computing nodes to partitions according to the number of waiting jobs as shown in Fig. 1.

There is a study that applied dynamic resource partitioning to the Athena development system using Mesos [2] and reported experiences on mixed workload [3]. In our case, problems arise when frameworks using different software stacks are installed and used on the local disk of computing nodes. A machine oriented mini-server (MOM) daemon of PBS scheduler not only runs and manages jobs and monitors resource usage on each computing node but also prohibit user daemon from listening port. However, user daemon is inevitable to enforce the principle of least privilege in Jupyter notebook service that leverage big data tools [4].

In this paper, we propose an approach to dynamic resource reallocation of computing nodes. We distinguish our approach from previous solutions in that it provides isolation of software stack and reallocates resources with a bare-metal environment suitable for preventing problems mentioned above. The remainder of this

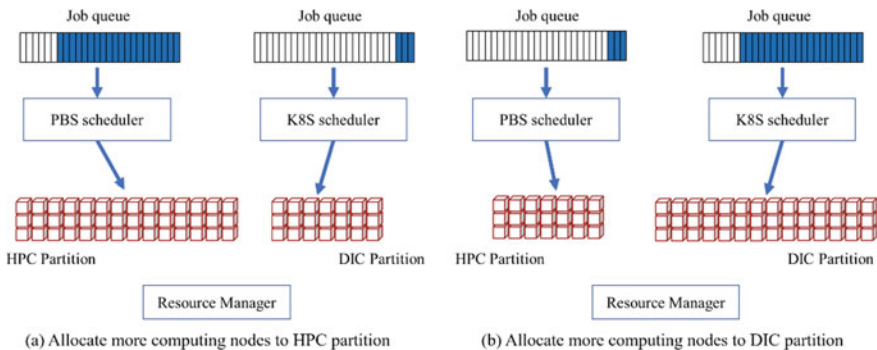


Fig. 1 Dynamically redistributing computing nodes to HPC and DIC partition according to tasks waiting in the job queue

paper is organized as follows: Sect. 2 introduces the proposed approach. Section 3 shows the experimental results and analysis. And Sect. 4 concludes this paper.

2 Methodology

In this section, we describe the key components and the reallocation process of computing nodes. Since HPC and DIC workloads have different characteristics, we first considered node level reallocation which means that allocation is done to the partition in computing node level rather than sharing components of computing resources (i.e., memory and network) so that each application can be executed according to its own characteristics without interruption. Second, the matching between tasks that request resources and available resources is determined by the number of CPUs. Last but not the least, the minimum resource of the partition managed by each scheduler is set as default (i.e., non-swappable and non-reallocatable) to prevent starvation of tasks in one partition. We briefly describe main components of dynamic resource reallocation as follows:

- **Task monitoring** to gather information about waiting tasks periodically
- **A computing node manager** to add and delete computing nodes to and from partitions managed by the scheduler
- **A database (DB)** that stores information and status about computing nodes
- **A dynamic node manager** that determines the need for reallocation through the overall system situation and by combining decision method according to information about available resources

Figure 2 the main components for dynamic resource reallocation and a sequence in the process shows a relation among main components and a sequence in the process. First, the dynamic node manager is a component that performs decision making, and collects information on waiting tasks and available resources through the local scheduler periodically. Then, it calculates the number of nodes required in each partition based on the collected information and decides whether to reallocate currently allocated computing nodes.

When a situation meets the resource reallocation policy, the suitable nodes are selected from the database, and the nodes are rebooted to the new disk image so that the selected computing nodes are allocated to a new partition. After assigned to a new partition, finally the database information is updated.

3 Results and Discussion

Figure 3 shows a testbed to demonstrate the feasibility of our approach. In order to manage heterogeneous platform-based task execution environments and services, bare-metal environment is advantageous in providing stable services. Though the

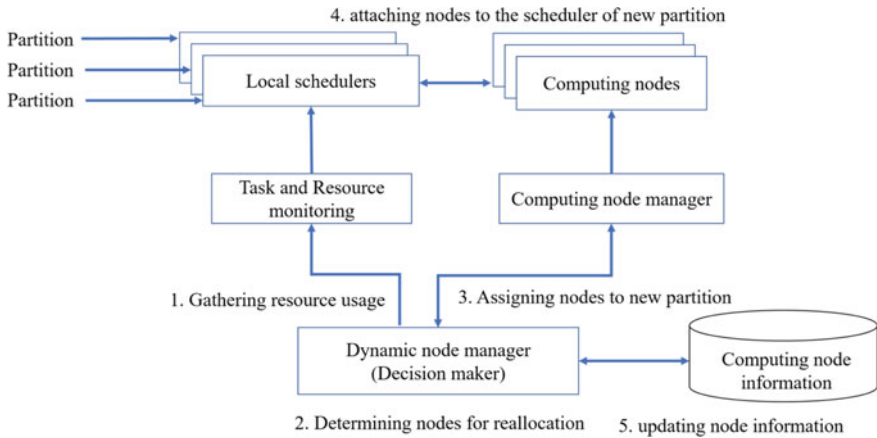


Fig. 2 The main components for dynamic resource reallocation and a sequence in the process

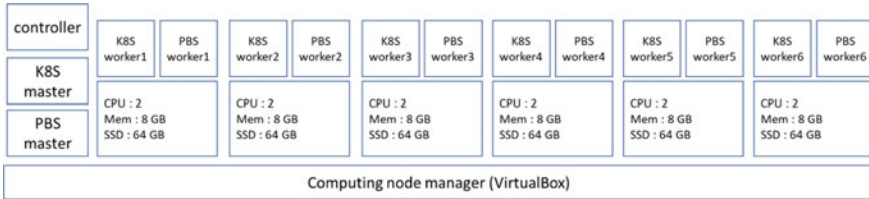


Fig. 3 A testbed environment to demonstrate the feasibility of dynamic resource reallocation

testbed was configured in a virtual machine, we can easily apply this environment to the real computing node of supercomputer by changing computing node manager from a hypervisor of VirtualBox to a MaaS (Metal as a Service) management tool.

Figure 4 confirms that overall throughput is improved by dynamically reallocating resources for the test scenario. In the initial environment, the node managed by the PBS scheduler and the node managed by Kubernetes are 3 nodes respectively, and 2 cpu processors per node. The High Performance LINPACK (HPL) benchmark requests 4 cpu resources per job, and the pbs-worker2 and pbs-worker3 nodes are assigned as shown in Fig. 4a. Among all idle resources, the resources allocated to the HPC partition are not sufficient to process the HPL job requiring 4 cores, so the jobs are waiting in the queue. At this time, based on the resource reallocation policy, the node of the DIC partition is reallocated (pbs-worker4) to the HPC partition. As in Fig. 4b, another job is allocated resources and executed, indicating that it is changed to the running state.

Two jobs using spark framework that request resources of DIC partition through Kubernetes are submitted in Fig. 4c. Each creates one driver and one work process, and each process requires 2 cores. Since there are 4 available cpu cores, 2 nodes with 2 cores are available, so one job can be executed, but the driver process preempts 2

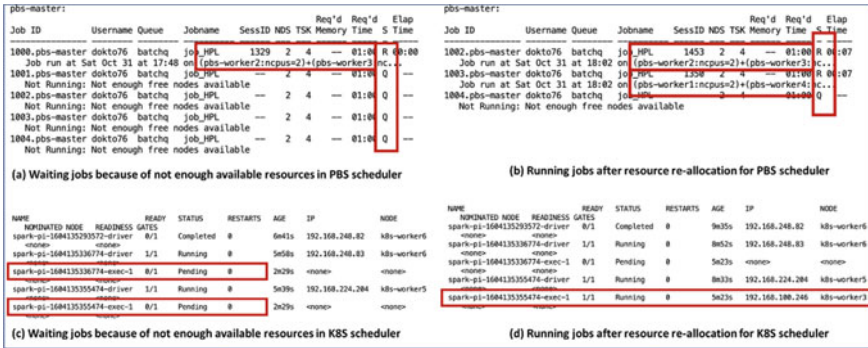


Fig. 4 A sample scenario to show the improved throughput by applying dynamic reallocation of computing nodes to each partition

cores in each job with the Kubernetes scheduler as non-preemption, so the worker process cannot be allocated resources result in pending of the task. After making a decision about the resource reallocation, the pbs-worker3 node in the HPC partition is reassigned to the DIC partition and switched as k8s-worker3. And then the resource is allocated to another worker of the previous pending task in DIC partition and all task changed to running state (Fig. 4d).

4 Conclusion

This work enhances overall system utilization of supercomputer as dynamic reallocating resources and solves the conflict problems between heterogeneous platforms executed on the same computing node by applying node level reallocation with a bare-metal environment. We implement a testbed and simulate a test scenario to demonstrate the feasibility of dynamic resource reallocation and the result shows that our approach can improve system utilization.

Acknowledgements This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIT) (No. CRC21011).

References

1. Lee S, Park JW, Jeong K, Hahn J (2021) Implementation of a container-based interactive environment for big-data analysis on supercomputer. In: Park JJ, Fong SJ, Pan Y, Sung Y (eds) Advances in computer science and ubiquitous computing (Lecture notes in electrical engineering), vol 715. Springer, Singapore. https://doi.org/10.1007/978-981-15-9343-7_58

2. Hindman B, Konwinski A, Zaharia M, Ghodsi A, Joseph AD, Katz RH et al. (2011) Mesos: a platform for fine-grained resource sharing in the data center. In: USENIX Symposium on Networked Systems Design and Implementation (NSDI)
3. Ayyalasomayajula, West K (2017) Experiences running different work load managers across cray platforms. In: Cray User Group conference (CUG'17)
4. Jupyter Notebook. [Online] Available: <https://jupyter.org>

Efficient Detection of Skin Cancer Using Deep Learning Techniques and a Comparative Analysis Study



Mehtab Hashim, Asad Masood Khattak, and Imran Taj

Abstract Many skin lesions may result in the wrong diagnosis of skin cancer, leading to delays and ultimately making the cure impossible. Framed within this statement, this article proposes an efficient skin cancer detection model and compares the six pre-trained models, used for transfer learning in ISIC 2019 dataset. Three most common types of skin cancer—melanoma, nevus, and basal cell carcinoma—are classified by using the transfer learning on the pre-trained models of the ISIC 2019 dataset, to conclude the most accurate detection results with training and test accuracy of 99.73% and 93.79%, respectively.

Keywords Skin cancer detection · Deep convolutional neural network · Pre-trained models · Transfer learning

1 Introduction

According to a WHO report [1], cancer is a leading cause of death worldwide, accounting for nearly one in six deaths. Skin cancers such as melanoma are the deadliest of all the cancers [2], and an early diagnosis of the cancer type can make a huge difference in saving lives. Machine learning (ML) has been used extensively across the healthcare domain [3], e.g., classification of skin cancer for accurate diagnosis is accomplished by various computer vision algorithms [4], comprising of four main steps: image acquisition, segmentation of the images, feature extraction, and feature classification [5]. A well-developed CNN model to process these steps can be

M. Hashim

Pakistan Institute of Engineering and Applied Science, Islamabad, Pakistan

A. M. Khattak (✉)

College of Technological Innovation, Zayed University, Dubai, UAE

e-mail: asad.khattak@zu.ac.ae

I. Taj

College of Interdisciplinary Studies, Zayed University, Dubai, UAE

e-mail: muhammadimran.taj@zu.ac.ae

very expensive in terms of processing and hardware requirements. We address this challenge by employing transfer learning in the ML/DL pre-trained models. We use weights of pre-trained models: MobileNetV2 [6], EffcientNetB2 [7], Inceptionv3 [8], NasNetMobile [9], Xception [10], and InceptionResNetV2 [11], thereby extracting features with minimum computational cost. The extracted features are used to detect one of the three types of skin cancer—melanoma, nevus, and basal cell carcinoma. Out of 12,447 images, we randomly select 829 for our experiments, and finally, we apply data augmentation on the images. The recent advancements in machine learning techniques have encouraged researchers to apply different ML models for the prediction of skin cancer. For instance, Esteva et al. [12] used a pre-trained Google InceptionV3 CNN model for categorization of skin cancer and achieved 0.9 categorization accuracy. Ratul et al. [13] created an automated detection system for the early detection of malignant skin lesions by employing four pre-trained models—including InceptionV3 to obtain a maximum classification accuracy of 89.81%. Brinker et al. [14] employed CNN deep learning model to classify skin lesion photos and trained on 12,378 dermoscopic images and compared the architecture's performance with 157 dermatologists from 12 German university hospitals. There has been a lot of work on using DL/ML techniques for classification. It has been demonstrated that the use of CNN architectures outperforms the use of conventional machine learning approaches in several domains. For example, Cheng [15] creates a training dictionary for facial recognition by employing a two-layer deep CNN for identification, sparse representation classification, and feature extraction.

The rest of this paper is organized as follows. We describe our proposed methodology in Sect. 2 and present our implementation results in Sect. 3. We conclude in Sect. 4.

2 Methodology

We provide dataset to each pre-trained model using transfer learning (TL) and then compare the results. Our DCNN model starts with training images on the collected data using pre-trained weights of the respected model; then, the steps of preprocessing, data normalization, data augmentation, and of the respected model follow, as shown in Fig. 1. We describe each of these steps:

Data Collection: The International Skin Imaging Collaboration (ISIC 2019) [16] challenge dataset is used in the training and testing of the proposed models. This dataset contains 25,331 images of seven types of skin cancer. But for our research, we have selected three most common cancer types: melanoma, nevus, and basal cell carcinoma. We are using 12,447 images corresponding to the three cancer types. The complete distribution for each skin cancer class used from the dataset titled *ISIC 2019 challenge* contains 3323, 4522, and 4550 images.

Data Preprocessing: This step is meant to improve the image data required for classification, by using the geometric transformations such as image rotation, image

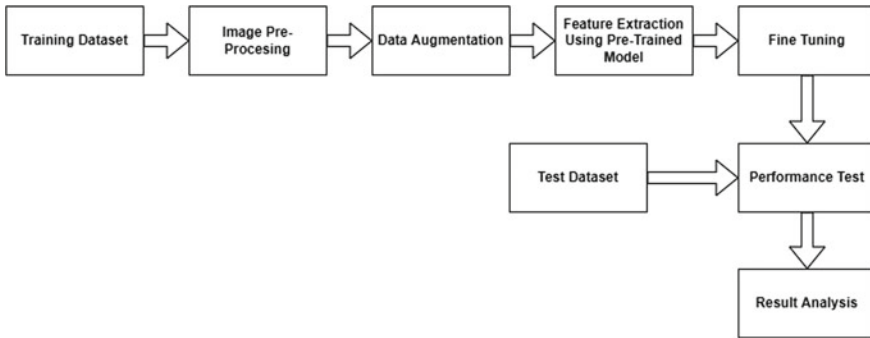


Fig. 1 Proposed system architecture

scaling, and image translation. We scale down all the photos' resolution to 224×224 pixels. We label and group the images according to a keyword search to facilitate picture searching. We took out all the transcribed photos from the dataset and then divide the images into a ratio of 80:20 for training and testing, respectively.

Data Normalization: During data normalization, we convert the RGB images into grayscale. We perform normalization by dividing our data with 225. This step ensures that we have the prominent attributes for the training of the model with no noise, and this is accomplished by removing the unwanted colors from the images.

Data Augmentation: We expand the training dataset by adding more photos, in this step, to prevent overfitting during the training process. We apply modifications to the photographs, such as rotation, width shift, height shift, shear range, and horizontal flip to facilitate the image augmentations. We rotate all images of different classes randomly between 0° and 180° in the range of 10° , while zoom range width, shaft range, and height shift range are set to 0.10.

3 Experiments and Results

The experiments are performed on Google Colab which provides 12.68 Gb free RAM and GPU which enables us to train a large number of images with minimized error rate. We use Python to implement our proposed method. We keep the values of the learning rate, batch size, and epochs the same for all the proposed models. This ensures better compression and unbiased experiments. The minimum learning rate is kept at 0.00000000001, while the batch size is 64 and the number of epochs is 30 for all the models. The data is randomly distributed with an 80% for training and 20% for testing. The learning rate is selected after a careful analysis to favor the global optimal learning at the cost of training time.

3.1 Model Evaluation

To evaluate the performance of the proposed models, we calculate the four metrics—precision, recall, *F1*-score, and weighted *F1*-score—using the following equations:

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{FalseNegative})} \tag{2}$$

$$F1 - \text{score} = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{3}$$

$$\text{Weighted } F1 - \text{score} = \frac{1}{K} \sum_{i=0}^K \beta_i \cdot F1_i \tag{4}$$

where *K* is number of classes used in the model and β_i is the proportion of the number of samples in class *i* to the total number of samples

3.2 Experimental Results

We start this sub-section by the justifying the six pre-trained models employed. Then, we provide a summary of the classification report obtained from each model from Figs. 2, 3, 4, 5, 6 and 7 and the final results and comparisons of all the models are summarized in Table 1.

MobileNetV2: MobileNetV2 functions well on mobile devices. It is built on an inverted residual structure where the bottleneck layers are connected by residual connections. Lightweight depth-wise convolutions are used in the intermediate

	precision	recall	f1-score	support
BCC	0.89	0.79	0.84	209
Melanoma	0.85	0.54	0.66	309
Nevus	0.68	0.97	0.80	311
accuracy			0.77	829
macro avg	0.81	0.77	0.77	829
weighted avg	0.80	0.77	0.76	829

Fig. 2 Classification report screenshot using MobileNetV2

	precision	recall	f1-score	support
BCC	0.94	0.98	0.96	200
Melanoma	0.94	0.90	0.92	295
Nevus	0.94	0.95	0.94	334
accuracy			0.94	829
macro avg	0.94	0.95	0.94	829
weighted avg	0.94	0.94	0.94	829

Fig. 3 Classification report screenshot using InceptionV3

	precision	recall	f1-score	support
BCC	0.91	0.90	0.90	218
Melanoma	0.84	0.81	0.82	305
Nevus	0.85	0.90	0.87	306
accuracy			0.86	829
macro avg	0.87	0.87	0.87	829
weighted avg	0.86	0.86	0.86	829

Fig. 4 Classification report screenshot using EfficientNetB2

	precision	recall	f1-score	support
BCC	0.95	0.95	0.95	254
Melanoma	0.91	0.84	0.87	292
Nevus	0.87	0.95	0.91	283
accuracy			0.91	829
macro avg	0.91	0.91	0.91	829
weighted avg	0.91	0.91	0.91	829

Fig. 5 Classification report screenshot using Xception

	precision	recall	f1-score	support
BCC	0.92	0.21	0.34	223
Melanoma	0.47	0.31	0.38	294
Nevus	0.53	0.99	0.69	312
accuracy			0.54	829
macro avg	0.64	0.50	0.47	829
weighted avg	0.62	0.54	0.49	829

Fig. 6 Classification report screenshot using NasNetMobile

	precision	recall	f1-score	support
BCC	0.96	0.94	0.95	193
Melanoma	0.93	0.90	0.91	308
Nevus	0.92	0.96	0.94	328
accuracy			0.93	829
macro avg	0.94	0.93	0.93	829
weighted avg	0.93	0.93	0.93	829

Fig. 7 Classification report screenshot using InceptionResNetV2

Table 1 Results and comparison of all models

Model name	Training accuracy (%)	Test accuracy (%)	Validation loss	Validation accuracy (%)
MobileNetV2	99.14	76.59	1.726	71.42
InceptionV3	99.38	92.34	0.3755	88.31
EfficientNet B2	99.60	86.36	0.4576	90.26
NasNetMobile	98.04	54.06	1.03	77.65
Xception	99.36	92.16	0.39	88.77
InceptionResnetV2	99.73	93.79	0.3138	89.87

expansion layer as a source of nonlinearity to filter features. A total of 32 filters makes up the first fully convolutional layer in MobileNetV2’s architecture, which is followed by 19 remaining bottleneck layers [6].

InceptionV3: The Inception family of convolutional neural networks includes Inceptionv3. Auxiliary classifiers are used to convey label information lower down the network. Moreover, label smoothing, factorize 7×7 convolutions, and other improvements are made [7].

EfficientNetB2: EfficientNet is a convolutional neural network design and scaling technique that uses a compound coefficient to consistently scale all depth, breadth, and resolution dimensions. The EfficientNet scaling method, in contrast to standard practice, adjusts network breadth, depth, and resolution uniformly using a set of predefined scaling coefficients [8].

Xception: Xception “extreme inception” pushes the fundamental ideas of inception to the limit. In inception, the initial input used is compressed using 1×1 convolutions, and from each of those input spaces, we have applied various kinds of filters to each of the depth spaces. Just the opposite occurs with Xception. Instead, it initially applies the filters to each depth map individually before using 1×1 convolution to finally condense the input space. This approach resembles a depth-wise separable convolution [10].

NasNetMobile: A method for searching a space of neural network configurations called NAS or neural architecture search, was created at Google Brain. The goal of the NASNet research is to identify the best CNN architecture by using reinforcement learning. CNNs are optimized for various sizes using NAS and common datasets like CIFAR10 and ImageNet. NASNetMobile refers to the scaled-down version [9].

InceptionResnetV2: A convolutional neural architecture called Inception-ResNet-v2 expands on the Inception family of architectures while incorporating residual connection information (replacing the filter concatenation stage of the Inception architecture) [11].

We observe in Table 1 that InceptionResNetV2 gave highest training accuracy, testing accuracy, validation accuracy, and lowest validation loss with values 99.73%, 93.79%, 89.87%, and 0.3138, respectively. InceptionV3 produced the second-best results. The least accurate training, testing, and validation results are provided by NasNetMobile, which had respective values of 98.04%, 54.06%, 77.65%, and 1.03 as well as the lowest validation loss. The main factor affecting the accuracy of the models is their respective number of parameters. The InceptionResNetV2 has the greatest number of parameters, i.e., 55.9 M, thereby outperforming the other models. Similarly, the InceptionV3 has the second highest number of parameters, i.e., 23.9 M, which explains the second highest accuracy of the model. The parameters of the other models: Xception, EfficientNetB2, and MobileNetV2 are 22.9 M, 9.2 M, and 3.5 M, respectively, explaining the performance deterioration accordingly. Although there are some exceptions to this observation, e.g., NasNetMobile has more parameters than MobileNetV2; despite that the results show a poor performance by NasNetMobile, which can be attributed to the weight transfer issues [17].

4 Conclusion

This research work used pre-trained deep neural network models and transfer learning for the ISIC 2019 dataset, to classify the three cancer types effectively. The pre-trained models of InceptionV3 and InceptionResnetV2 outperformed the other models by giving better training and test accuracy, even on imbalanced data. We also realized that the mentioned pre-trained models (compared to VGG19 model) are very light and can train a large amount of data with average computational resources. Altogether, our proposed research and models have shown promising results. Our future work will build on the imbalanced data analysis using InceptionV3 and InceptionResnetV2.

Acknowledgements This research work was supported by Zayed University Research Incentive Fund #R21095.

References

1. Cancer (Feb. 2022) Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. Bichakjian CK, Halpern AC, Johnson TM, Hood AF, Grichnik JM, Swetter SM et al (2019) Guidelines of care for the management of primary cutaneous melanoma. *J Amer Acad Dermatol* 80(1):208–250
3. Taj I, Zaman N (2022) Towards industrial revolution 5.0 and explainable artificial intelligence: challenges and opportunities. *Int J Comput Digit Syst* 12(1):295–320
4. Vocaturo E, Ester Z, Pierangelo V (2018) Image pre-processing in computer vision systems for melanoma detection. In: 2018 IEEE International conference on bioinformatics and biomedicine (BIBM), IEEE
5. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: 2015 International conference on computing communication control and automation, IEEE
6. Sandler M et al. (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
7. Szegedy C et al. (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition
8. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning, PMLR
9. Saxen F et al. (2019) Face attribute detection with mobilenetv2 and nasnet-mobile. In: 11th International symposium on image and signal processing and analysis (ISPA), IEEE
10. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition
11. Szegedy C et al. (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
12. Esteva A et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542.7639:115–118
13. Ratul AR, Hamed MM, Lee W-S, Parimbelli E (2019) Skin lesions classification using deep learning based on dilated convolution. *bioRxiv* 860700
14. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Schilling B et al (2019) Deep learning outperformed 136–157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 113:47–54
15. Cheng E-J, Chou K-P, Rajora S, Bo-Hao Jin M, Tanveer C-TL, Young K-Y, Lin W-C, Prasad M (2019) Deep sparse representation classifier for facial recognition and detection system. *Pattern Recogn Lett* 125:71–77
16. Combalia M et al. (2019) BCN20000: Dermoscopic lesions in the wild
17. The detail of the models is available online at: <https://keras.io/api/applications/#available-models>

Autonomous Cooperative Driving Protocol Based on the Security Status of IVN



Daekyo Shin and Soohyun Jang

Abstract Autonomous cooperative driving determines the optimal operating conditions and route based on information acquired by the own vehicle and information provided by surrounding vehicles. The cooperative driving environment is an environment with a high probability of accidents such as lane change, overtaking, avoidance, cutting in, intersection, and merging. In addition, the reliability and integrity of the data transmitted by the other vehicle in the cooperative driving must be ensured. This paper proposes a method to proceed/abandon/ignore/warning of autonomous cooperative driving based on exchanging information on security status of In-Vehicle Network (IVN) and safety status through vehicle communication for successful cooperative driving of autonomous vehicles.

Keywords In-vehicle security · Cooperative driving · V2X · Impact rating · Response driving

1 Overview

Autonomous vehicles perform autonomous driving through their own sensors and given information. However, in order to carry out strategic actions that can affect traffic flow, information must be shared through cooperation with other nearby vehicles. Autonomous driving cooperative driving determines optimal driving conditions and driving routes based on information acquired from its own vehicle sensors and information provided by nearby vehicles.

The purpose of this paper is to determine the integrity and reliability of the internal data, and to provide the security status of my vehicle to the nearby vehicles to enable safe cooperative driving. The security status of IVN data can be found through

D. Shin (✉) · S. Jang

Korea Electronic Technology Institute, Seongnam Gyeonggi 13488, Korea
e-mail: dukeshin@keti.re.kr

S. Jang

e-mail: shjang@keti.re.kr

network intrusion, the location and function of the contaminated ECU, the degree of contamination, etc. The current vehicle’s security status is defined and informed by the impact rating. It also suggests countermeasures when the security status of the other vehicles in cooperative driving is poor.






In this paper, after explaining the existing vehicle cooperative driving method, we propose an autonomous cooperative driving method based on the internal security state of the vehicle.

2 Existing Autonomous Cooperative Driving Method

2.1 Vehicle Classification for Cooperative Driving

In the cooperative driving scenario, various types of vehicles exist depending on whether the autonomous driving system is installed or not and whether there is a communication device, and they are classified as shown in Table 1 by vehicle function.

Table 1 Classification of vehicles subject to cooperative driving

Vehicle type	ADS	V2X	Description
E-CDA 	Yes	Yes	Ego Cooperative Driving Automation, A vehicle that performs a specific driving mission in a cooperative driving scenario
T-CDA 	Yes	Yes	Target Cooperative Driving Automation. Target vehicle performing cooperative driving with E-CDA in cooperative driving scenario
C-VEH 	No	Yes	Connected VEHicle, A vehicle driven by a general driver that transmits and receives basic safety messages
N-VEH 	No	No	Normal VEHicle, A general vehicle driven by an ordinary driver without communication function
CE-VEH 	No	Yes	Connected Emergency VEHicle, A vehicle capable of forcibly ordering a route change request to nearby vehicles

When ADS: Autonomous Driving System, V2X: Vehicle Communication system

2.2 Cooperative Driving Strategy

Cooperative driving of autonomous vehicles is defined according to SAE J3216. The class of cooperative driving is divided into A, B, C, and D as follows, and the class is divided according to the quantity and quality of information shared between vehicles.

Class A cooperative driving, Here I am and What I see, a form of cooperation by sharing recognized traffic environment information to autonomous vehicles. Vehicles equipped with V2X (T-CDA, C-VEH, CE-VEH) broadcast their vehicle information (position, direction, steering angle, etc.) to implement.

Class B cooperative driving, This is what I plan to do, a form of cooperation by sharing driving intentions for planned future actions. T-CDA broadcasts its driving intention (left turn, right turn, reverse, straight forward, U-turn, etc.) to the surrounding area, and E-CDA implements cooperative driving based on the received driving intention.

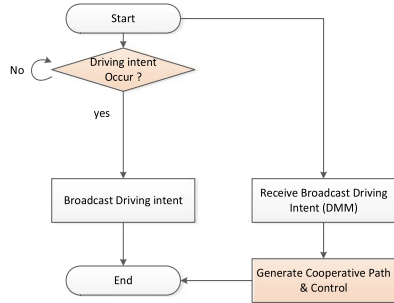
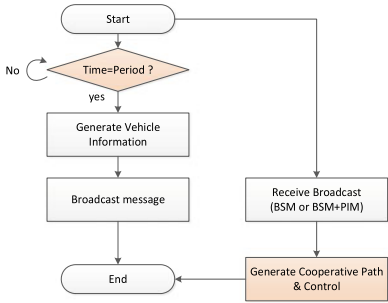
Class C cooperative driving, Lets do this together, a form of agreeing on the driving order through a series of collaborative messages between vehicles. The E-CDA, which has the initiative, recognizes the T-CDA to be cooperated, determines that cooperative driving is necessary for safe driving, proposes driving priority, and transmits it to the T-CDA through a cooperative driving message. After receiving the cooperative driving message, the T-CDA returns a consent message and drives according to the suggested driving order.

Class D cooperative driving, I will do as directed, a form in which a special vehicle instructs nearby vehicles to act. Emergency driving vehicles such as EC-VEH (ex, fire trucks, etc.) broadcast their emergency driving information (EDM) to the surrounding area to induce yielding driving of other vehicles (Fig. 1).

3 Proposed Cooperative Driving Protocol

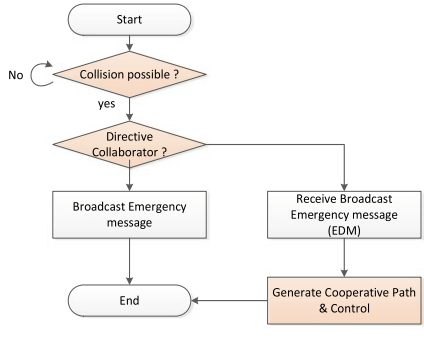
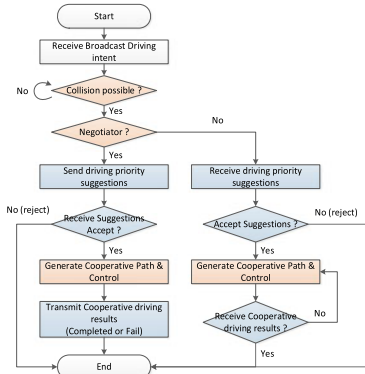
In this paper, T-CDA, a vehicle to be cooperated, is classified according to the security status of IVN as shown in Table 2. The security status of IVN was defined by referring to Annex F.4 Impact rating for operational damage of ISO/SAE 21,434 standard.

Table 2 shows the operational impact ratings for cyber hacking shown in ISO/SAE 21,434. As for the degree of operational damage against invasion, O0 is the lowest and O3 is the highest.



< Existing Class A strategy >

< Existing Class B strategy >



< Existing Class C strategy >

< Existing Class D strategy >





Fig. 1 Existing class cooperative driving strategy

Table 2 Example operational impact rating criteria

Impact rating	Symbol	Criteria for operational impact rating
Severe	O3	The operational damage leads to the loss or impairment of a core vehicle function
Major	O2	The operational damage leads to the loss or impairment of an important vehicle function
Moderate	O1	The operational damage leads to partial degradation of a vehicle function
Negligible	O0	The operational damage leads to no impairment or non-perceivable impairment of a vehicle function

* Refer: ISO/SAE 21,424 Annex F.4 Impact rating for operational damage

Table 3 T-CDA vehicle classification based on security status of IVN

Vehicle type	ADS	V2X	Operational impact rating	Criteria for operational impact rating
T-CDA(O0) 	Yes	Yes	Negligible	Target cooperative driving with impact rating 0
T-CDA(O1) 	Yes	Yes	Moderate	Target cooperative driving with impact rating 1
T-CDA(O2) 	–	Yes	Major	Target cooperative driving with Impact rating 2
T-CDA(O3) 	–	Yes	Severe	Target cooperative driving with impact rating 3

3.1 Vehicle Classification Based on Security Status of IVN

In the proposed cooperative driving scenario, T-CDA vehicles are classified as shown in Table 3 based on the security status of IVN.

3.2 Proposed Cooperative Driving Strategy

We propose a cooperative driving strategy that differentially applies cooperative driving strategies according to security status of IVN, because it informs the surrounding vehicles of the degree of security breach of own vehicle through Operation Impact Rate.

Proposed Class A cooperative driving Strategy, Among the received Basic Safety Message (BSM) of T-CDA, responses such as alert driving and evasive driving, abandonment of Cooperative driving, and cooperative driving are carried out according to the Impact Rating.

Proposed Class B cooperative driving Strategy, Among the received Driving Maneuver Message (DMM) of the T-CDA, responses according to the Impact Rating.

Proposed Class C cooperative driving Strategy, Among the received Driving Maneuver Message (DMM), Request/Response Driving Negotiation Message (DNM_req, DNM_res) of the T-CDA, responses according to the Impact Rating.

Proposed Class D cooperative driving Strategy, The emergency vehicle transmits Emergency Driving Message (EDM) broadcast and receives BSM of the target of cooperative driving, and executes evasive driving and alert driving when Impact Rating > Severe (Fig. 2).

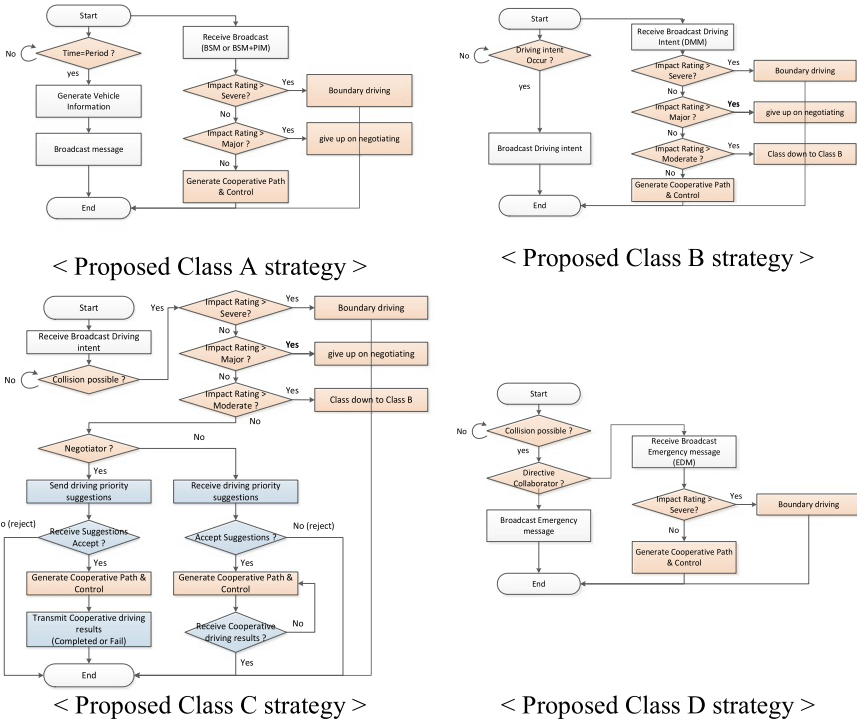


Fig. 2 Proposed class cooperative driving strategy

4 Cooperative Driving Response Plan

This chapter presents countermeasures according to the implementation of cooperative operation according to the security status of IVN. There are three main countermeasures. First, if the target vehicle is in O1 state, the cooperative driving class is changed to cooperative driving grade A. The contents of the operation message transmitted in this state are ignored. Second, if the target vehicle is O2, it is excluded from cooperative driving. It will no longer attempt to drive cooperatively with the vehicle and will ignore the content of any messages received from the vehicle. Finally, if the target vehicle is O3, the vehicle is considered a dangerous vehicle and the occurrence situation is carefully monitored. In the case of abnormal behavior, dangerous situations such as widening gaps are prevented in advance.

Since Operational Impact Rate broadcasting is determined through IVN intrusion detection module separately from IVN operation and broadcast through V2X, even in the case of a dangerous vehicle, its security status can be informed to surrounding vehicles (Fig. 3).

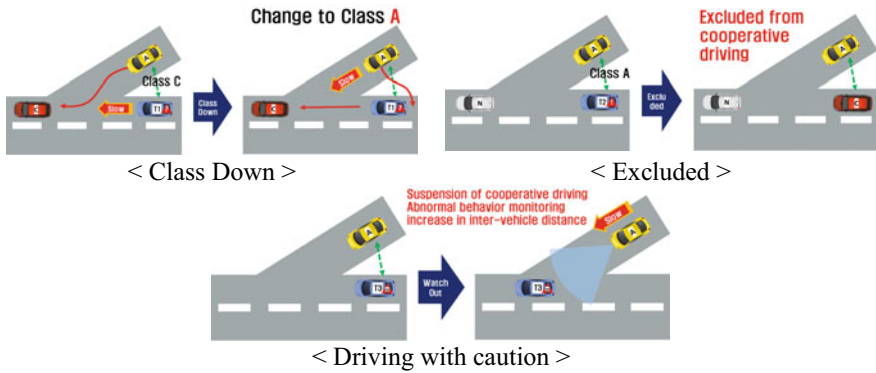


Fig. 3 Proposed cooperative driving response plan

Acknowledgements This research was supported by the 2021 autonomous driving development innovation project of the Ministry of Science and ICT, ‘Development of technology for security and ultra-high-speed integrity of the next-generation internal network of autonomous vehicles’ (No. 2021-0-01348).

References

1. Taxonomy and Definitions for Terms Related to Cooperative Driving Automation for On-Road Motor Vehicles (2020) SAE J3216
2. Road vehicles - Cybersecurity engineering (2021)ISO/SAE 21434
3. Jung H-G et al (2018) Reliability verification procedure of secured V2X communication for Autonomous cooperation driving. Int Conf Inf Commun Technol Conver (ICTC) 2018:1356–1360. <https://doi.org/10.1109/ICTC.2018.8539617>
4. Insu O, Jeong E, Park J, Jeong T, Park J, Yim K (2021) Cyber attack scenarios in cooperative automated driving. Adv Broad-Band Wirel Comput Commun Appl 159:416

Confidence Calibration of Hand Gesture Recognition Networks Based on Ensemble Learning



Zongjing Cao, Yan Li, and Byeong-Seok Shin

Abstract Confidence calibration refers to the degree of matching between predicted confidence and actual correctness probability. The ensemble of neural networks, also known as the ensemble of models, is a common method for estimating predictive uncertainty. However, this method requires a lot of modifications to the training procedure, and the computational cost is also expensive compared with other existing methods. We propose an approximate ensemble method that is simple to implement and results in well-calibrated confidence estimates without introducing various sets of models. More specifically, the historical model weights in the training phase are selected to simulate a new set of model parameters for uncertainty estimation. We show that our proposed method can produce high-quality predictive uncertainty estimates comparable to standard ensemble methods through a series of experiments on the Jester benchmark dataset.

Keywords Confidence Calibration · Ensemble Learning · Gesture Recognition

1 Introduction

Gesture recognition is a type of perceptual computing user interface that allows computers to capture and interpret human gestures as commands. The purpose of gesture recognition is to classify a gesture clip into one of the predefined action categories. Computer vision methods based on deep neural networks have brought unprecedented predictive performance on multiple machine learning tasks. Despite

Z. Cao · Y. Li · B.-S. Shin (✉)

Department of Electrical and Computer Engineering, Inha University, Incheon 22212, South Korea

e-mail: bsshin@inha.ac.kr

Z. Cao

e-mail: zjcao@inha.edu

Y. Li

e-mail: leeyeon@inha.ac.kr

the impressive accuracy achieved, deep neural networks tend to produce overconfident predictions on out-of-distribution data. In other words, deep neural networks perform poorly in quantifying the uncertainty of predictions and do not know what they do not know. Overconfident incorrect predictions pose a significant challenge to deploying models in real-world applications. Especially in real-world decision-making systems, such as autonomous driving and medical care, models need to have a certain degree of confidence in their prediction to make more informed decisions [1–3]. Thus, reasoning about the uncertainty associated with the model predictions is critical to the practical application of deep neural networks. In other words, deep neural networks should not only make predictions but also provide well-calibrated degrees of confidence.

Confidence calibration is defined as the ability of a model to provide an accurate probability for any of its predictions [2, 3]. A well-calibrated classifier is a probabilistic classifier for which the output of the predicted probability method can be directly interpreted as a confidence level. For example, a well-calibrated (binary) classifier classifies samples, and about 80% of the samples whose predicted probability values are close to 0.8 should belong to the positive class. More formally, a perfectly calibrated gesture recognition model can be defined by the following equation:

$$\mathbb{P}(\hat{Y} = Y | \hat{P} = p) = p, p \in [0, 1] \quad (1)$$

where \hat{Y} is gesture prediction and \hat{P} is its associated confidence. The confidence estimate \hat{P} is expected to be calibrated, which intuitively means that \hat{P} represents the true probability. Figure 1 (left) shows the visualization of the calibration curve (also known as the reliability diagram) of the hand gesture recognition model based on the 3D ResNet [4] network on the Jester dataset [5]. The x -axis represents the average predicted probability for each bin, and the y -axis is the fraction of positives, which is the proportion of samples with positive categories in each bin. These diagrams plot the expected sample accuracy as a function of confidence. If the model is perfectly calibrated—i.e., if Eq. (1) holds—then the chart should draw the identity function. And any deviation from the perfect diagonal represents that the model is miscalibrated. Figure 1 (right) is a histogram showing the distribution of the predicted probabilities, which is used to provide further insight into the behavior of the classifier.

Calibrating a classifier consists of fitting a regressor (called a calibrator) that maps the output of the classifier to a calibrated probability in $[0, 1]$. Denoting the outcome of the classifier for a given sample by f_i , the calibrator tries to predict $p(y_i = 1/f_i)$. The ensemble of neural networks, also known as the ensemble of models, is currently a standard method for estimating predictive uncertainty. However, the technique requires significant modifications to the training procedure and is also expensive in terms of both training time and computational consumption. To solve these challenges, we propose an alternate ensemble learning approach. The method

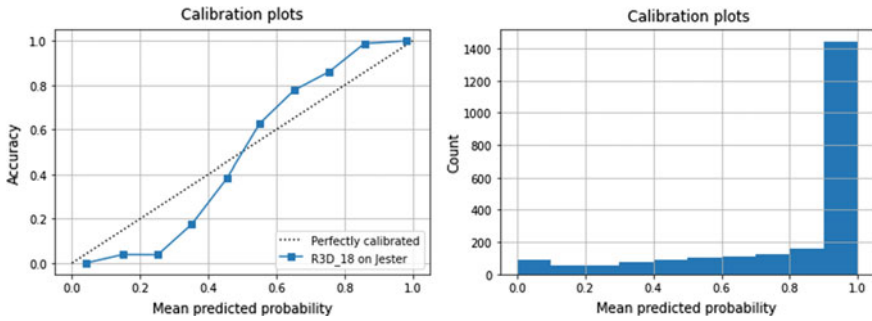


Fig. 1 (left) Calibration curve of 3D ResNet-18 network-based gesture recognition model on the Jester dataset; (right) Histogram of the number of the samples in each predicted bin

is simple to implement, requires very little hyperparameter tuning, and yields well-calibrated confidence estimates. A series of experiments on the Jester benchmark dataset show that our method can produce high-quality estimates of prediction uncertainty. Compared with the standard ensemble of neural networks, our proposed method can achieve comparable performance. The experimental results also further indicate that our method obtains well-calibrated confidence estimates.

2 Calibrating the Classifier

The calibrated confidence estimates obtained from the deep neural networks are critical to applying the model in real industry. In real-world industrial applications, the prediction of gesture classification networks should not only be accurate but also be indicated when its prediction results are uncertain. Specifically, the gesture recognition network should provide a calibrated confidence measure in addition to predicting the class of the gesture. Here, the well-calibrated confidence can be interpreted as the likelihood that the probability associated with the predicted gesture label should reflect its ground truth correctness.

The objective of the ensemble approach is to combine the predictions of multiple underlying estimators built with a specific learning strategy to improve the robustness of the model [6]. The key idea of the average-based ensemble approach is to create several independent estimators and then average their prediction results. It is observed that the combined estimator is usually better than any single base estimator because its variance is reduced. In [6], the authors demonstrated that averaging network predictions within an ensemble could also be used to derive helpful uncertainty estimates. However, ensemble learning of model combinations is often hard to learn, and any wrong choice can lead to lower predictive accuracy than an individual model. Moreover, compared with other methods, such as Bayesian neural

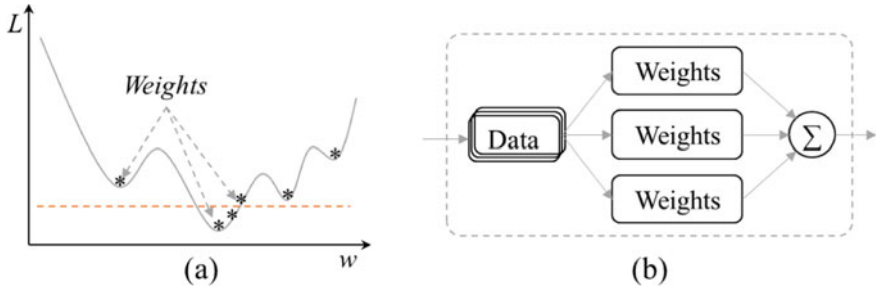


Fig. 2 **a** Training curve of neural network model; **b** the architecture of the proposed method

networks and the post-calibration method, the ensemble method is expensive in terms of training time and computational consumption. To solve this problem, we present an alternative ensemble approach that is simple to implement and produces well-calibrated confidence predictions. The main idea of our proposed approach is shown in Fig. 2. In the training phase, we use the training set to perform normal supervised learning on the model. The learning curve of the learning phase of a neural network is shown in Fig. 2a. In the testing phase (Fig. 2b), for a new given input, we sample multiple new parameter sets from the training history weights of the training phase as approximate weights for the new model and predict numerous outputs.

In this work, we treat the ensemble as a uniformly-weighted mixture model and combine the predictions as:

$$p(y|x) = \frac{1}{M} \sum_{m=1}^M p_{\theta_m}(y|x, \theta_m) \tag{2}$$

For our hand gesture classification task, this corresponds to averaging the predicted probabilities. For ease of computing quantiles and predictive probabilities, we further approximate the ensemble prediction as a Gaussian whose mean and variance are, respectively, the mean and variance of the mixture. The mean and variance of a mixture Gaussian are given by Eqs. (3) and (4):

$$u_*(X) = \frac{1}{M} \sum_m \mu_{\theta_m}(X) \tag{3}$$

$$\sigma_*^2(X) = \frac{1}{M} \sum_m (\sigma_{\theta_m}^2(X) + \mu_{\theta_m}^2(X)) - \mu_*^2(x) \tag{4}$$

3 Experiments and Results

We evaluated the proposed method on the commonly used Jester datasets [5]. It includes 148,092 labeled video clips of humans performing basic, predefined hand gestures in front of a laptop camera or webcam. In this work, we divided the dataset into two groups, closed set, and open set, and then constructed min-datasets by random sampling from the closed set at a ratio of about 1/4. The mini close set cover 20 different class gestures split in the ratio of 8:1:1 for training, validation, and testing. The detailed dataset divisions are shown in Table 1. The numbers of gesture samples in training, validation, and testing splits are 22,000, 2400, and 2240, respectively. All models are trained, validated, and tested on the Jester closed set.

The proposed network was implemented using the PyTorch deep learning framework. The model was trained and validated on a server equipped with four NVIDIA GeForce RTX 3090 GPUs. We used 3D ResNet-18, pre-trained on kinetics-400 as the backbone network, and then fine-tuned it on the Jester training set. We adopted SGD as the optimizer with a momentum of 0.9. We train the model for 25 epochs with a batch size of 16. The initial learning rate is set to 0.001 and then decays by 0.1 times every seven epochs. The training accuracy and loss curves of the model during the training and validation phases are shown in Fig. 3.

Measuring Calibration: Negative log likelihood (NLL) and miscalibration area (MA) are two proper scoring rules and two famous metrics for evaluating predictive uncertainty [7, 8]. In addition, we measured the accuracy of the classification. The

Table 1 Statistics of the dataset splitting

Datasets	Closed set		Open set		Mini closed set	
	Classes	Samples	Classes	Samples	Classes	Samples
Training	20	82,963	7	35,599	20	22,000
Validation	20	10,112	7	4675	20	2400
Testing	20	10,169	7	4574	20	2240

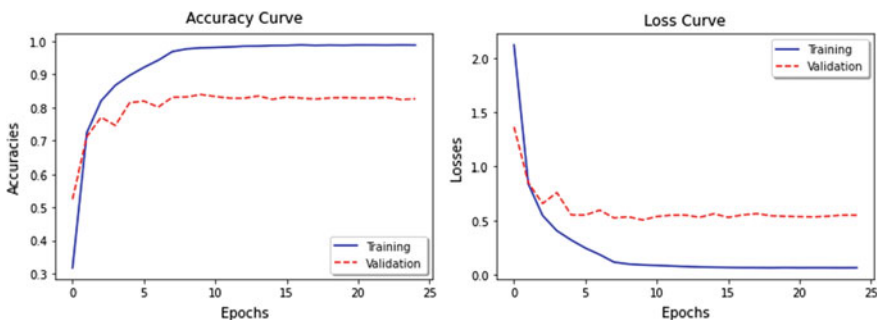


Fig. 3 The accuracy (left) and loss (right) curves of the network during the training and validation phases

results are shown in Table 2. We observe that our calibration method improves both NLL and MA but does not significantly alter the prediction accuracy measures (precision, recall, and $F1$ score). The experiments illustrate that our method produces well-calibrated uncertainty estimates. To compare and evaluate the accuracy of classification, we visualized the confusion matrix, as shown in Fig. 4. The higher the diagonal values of the confusion matrix, the better, indicating many correct predictions.

Visualizations: Calibration curves compare how well the probabilistic predictions of a classifier are calibrated. Well-calibrated classifiers are probabilistic classifiers for which the output of predicted probability can be directly interpreted as a confidence level. When performing gesture classification, one often wants to predict the class label and the associated probability. This probability gives some kind of confidence in the prediction. To intuitively understand the calibration of the predicted probabilities, we visualized the calibration curves, as shown in Fig. 5.

Table 2 Comparison of NLL and MA and classification reports

	NLL	MA	Precision	Recall	$F1$	AUC	Accuracy
Before	3.16	0.35	84.44	83.71	83.81	91.42	83.71
After	1.21	0.10	84.34	83.35	83.44	91.23	83.35

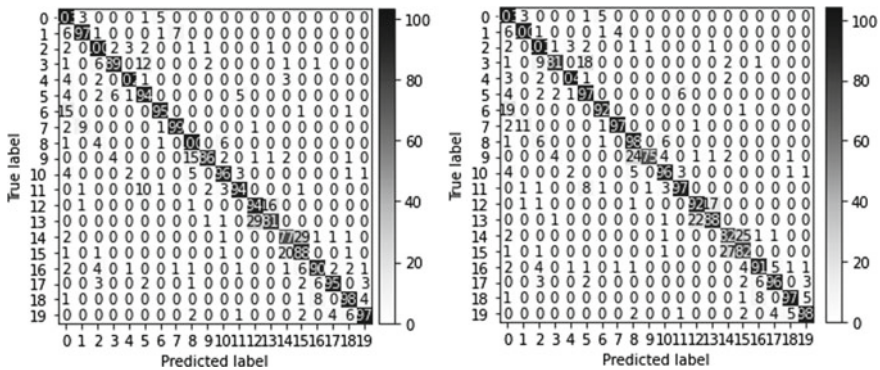


Fig. 4 Comparison confusion matrix of accuracy on Jester dataset. (left) Before calibration; (right) after calibration

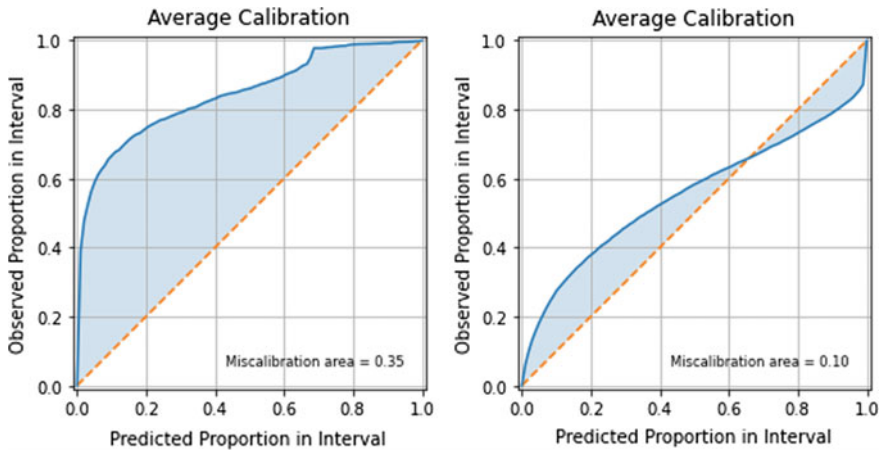


Fig. 5 Comparison results of calibration curves on the Jester dataset. (left) before calibration; (right) after calibration

Acknowledgements This research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0017157-World Class Plus Project). This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (No.NRF-2022R1A2B5B01001553) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)).

References

1. Gawlikowski J et al. (2021) A survey of uncertainty in deep neural networks. arXiv preprint [arXiv:2107.03342](https://arxiv.org/abs/2107.03342)
2. Jospin LV et al. Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Comput Intell Magazine* 17.2:29–48 (2022)
3. Scheirer WJ et al. (2012) Toward open set recognition. In: *IEEE transactions on pattern analysis and machine intelligence*, vol 35.7, pp 1757–1772
4. Tran D et al. (2018) A closer look at spatiotemporal convolutions for action recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*
5. Materzynska J et al. (2019) The jester dataset: a large-scale video dataset of human gestures. In: *Proceedings of the IEEE/CVF international conference on computer vision workshops*
6. Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv Neural Inf Process Syst* 30
7. Zhang H et al. (2017) Mixup: beyond empirical risk minimization. arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412)
8. Guo C et al. (2017) On calibration of modern neural networks. *Int Conf Mach Learn PMLR*

Optimization Strategy for Residential Heating Based on Model Predictive Control



Zhiwei Jia , Yue Gu, and Yanxue Li 

Abstract Heating energy consumption accounts for a large proportion of energy use in residential buildings, and is a major problem in building energy efficiency; in addition, occupant demand for thermal comfort increases heating energy consumption and peak loads, further affecting total energy consumption. In this study, an optimization strategy based on model predictive control is proposed to optimize the occupants' thermal comfort and heating energy consumption as the optimization objectives. Moreover, the study also introduces photovoltaic generation to consider the impact of the presence of photovoltaic on indoor thermal comfort and heating energy consumption. The proposed controller considers the outdoor environmental disturbances, and its control performance is tested by simulation on two independent residences and compared with the existing baseline control methods for dwellings. The results show that the optimization strategy based on model predictive control improves the fluctuation of room temperature, increases thermal comfort, reduces heating energy consumption and increases the PV consumption.

Keywords Residential heating · Model predictive control · Optimization strategy

1 Introduction

The building sector is responsible for approximately 40% total primary energy usage and accounts for 38% energy-related carbon dioxide emissions [1, 2]. Currently, most operational parameters of building heating or cooling supply systems are based on experiences or rule-based control. In existing buildings, the heating, ventilation and air conditioning system shares about 50% of the building energy use. In this respect, developing an effective real-time management strategy for building energy systems

Z. Jia (✉) · Y. Gu · Y. Li

Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology, Fushun Road 11, Qingdao 266033, China
e-mail: 17863977262@163.com

is highlighted to contribute to achieving the carbon neutrality transition in building sector.

To reduce energy consumption and carbon dioxide emission, high thermal insulation standard and on-site renewable energy development are adopted as main energy conservation efforts [3, 4]. The structural thermal mass of buildings is increasingly considered as an important flexibility resource through a trade-off between the energy consumption and thermal comfort [5]. Regarding the thermal mass flexibility, accurately characterizing the demand-side flexibility is becoming an important issue.

With the changes in building energy system with decentralized energy supply and storage, the operational strategy attempts to make effective decisions in response to the dynamic environment received more attention [6, 7]. In contrast to the conventional control methods such as simple on/off or proportional-integral (PI) control, the model predictive control (MPC) is a model-based approach for constrained control considering both current and future states [8, 9]. Ref [10] adopted MPC controller for a retail building HVAC system, bringing 1.7% cost saving while achieving the same thermal comfort.

Building thermal model is a typical nonlinear multivariable system. The low-order gray box model is simple to model and can express the thermal dynamics of the building is popularly adopted for state estimation in building MPC control. The model inaccuracy and uncertain disturbances such as control signals and weather conditions would challenge MPC for optimal operational management of building energy system [11, 12]. The main contributions of this work are summarized as follows:

Design MPC optimization framework to guarantee thermal comfort to the occupants while minimizing the heating energy demand. Assess and compare performances of MPC controller with different goals in terms of heating load, thermal comfort and on-site PV self-consumption.

2 Method

2.1 Building Thermal Model

To capture thermal behaviors of buildings, the thermal dynamic of the building is modeled using a third-order RC network. Including internal thermal mass, indoor air, house envelope and outdoor environment, the heating energy interactions Q among components is described in Fig. 1.

Once all the parameters and boundary conditions are known, the energy balance for each part is described as the following equations:

$$C_n \frac{dT_n}{dt} = \frac{T_m - T_n}{R_{n,m}} + \frac{T_{en} - T_n}{R_{en,n}} + Q_{solar,n} + Q_{ac} + Q_{ven} \quad (1)$$

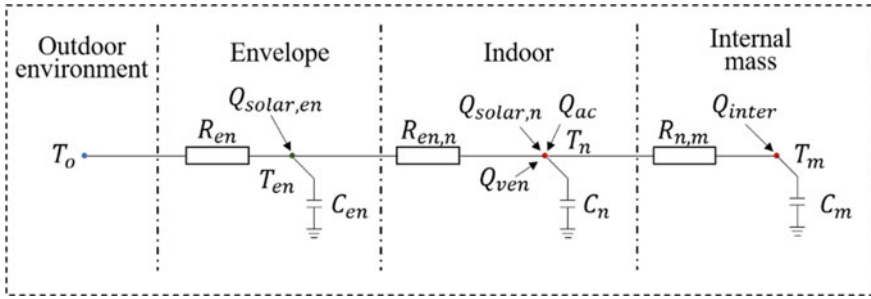


Fig. 1 Schematic of the gray-box RC thermal network (3R3C)

$$C_{en} \frac{dT_{en}}{dt} = \frac{T_n - T_{en}}{R_{en,n}} + \frac{T_o - T_{en}}{R_{en}} + Q_{solar,en} \tag{2}$$

$$C_m \frac{dT_m}{dt} = \frac{T_n - T_m}{R_{n,m}} + Q_{inter} \tag{3}$$

where R and C represent the equivalent thermal resistance and heat capacitance of the nodes, T denotes zone temperatures, subscripts m, n, o, en and win present internal thermal mass, indoor air, outdoor air, envelope and windows, respectively. Q_{ac} refers to air conditioning heating energy load. $Q_{solar,en}$ and $Q_{solar,n}$ are solar heat gains absorbed by external wall and indoor air (through the window), respectively. Q_{inter} indicates internal heat gains by internal thermal mass, Q_{ven} indicates the heat loss due to mechanical ventilation, which is estimated by the ventilation rate.

The system of equations can be expressed with the continuous state-space form can be effectively used to deal with the output feedback problem:

$$dx/dt = A_c x + B_c u + E_c d \tag{4}$$

where x is a state vector (representing the temperatures), y defines the system output (indoor air temperature), u represents control variable (space heating energy input), d is a vector denotes the disturbances. The matrices A_c , B_c and E_c are composed of parameters of the buildings, calculated according to Eqs. (1)–(3). The estimated values of heat resistance and capacitance of investigated houses were determined according to physical characteristics and measured data. Root mean square error (RMSE), mean bias error (MBE) and coefficient of variation of the root mean square error (CVRMSE) are the performance metrics for evaluating the accuracy of the simulation model.

2.2 System Optimization

The purposes of the MPC controller of building energy system mainly focus on minimizing the energy consumption while maintaining indoor thermal comfort. Firstly, the optimization objective of the controller is set to minimize the energy consumption for space heating while maintaining the room temperature in the comfort range, we set slack variables to satisfy the solution of the optimized controller, the optimization problem is expressed by the following equation:

$$J_1 = \sum_{i=1}^N (y_i - y_{\text{ref}})^T V (y_i - y_{\text{ref}}) + \sum_{i=1}^N u_i^T R u_i + \sum_{i=1}^N \sigma_i^T S \sigma_i \quad (5)$$

Constraints of input and states are expressed with the following constraints:

$$\left\{ \begin{array}{l} \forall i, u_{\min} \leq u_i \leq u_{\max}, \\ \forall i, y_{\min} - \sigma_i \leq y_i \leq y_{\max} + \sigma_i. \end{array} \right. \quad (6)$$

where y_i is system output at time i , denotes indoor temperature, y_{ref} represent user-defined temperature, V , R and S are defined as the weight matrix, y_{\min} and y_{\max} present user-defined upper and lower limits for indoor temperatures, u_i denotes the control input at time interval i , u_{\min} and u_{\max} present maximum and minimum input heating capacity energy, respectively. σ_1 and σ_2 define non-negative slack variables that avoid infeasible solution of optimization problem when the constraints are not satisfied.

Secondly, the optimization of the controller is set to balance the needs of reducing energy consumption, increasing the local consumption of rooftop PV generation and maintaining the indoor thermal comfort. The optimization goal can be expressed by the following equation:

$$J_2 = \sum_{i=1}^N e_i^T R e_i - \sum_{i=1}^N v_i^T R v_i + \sum_{i=1}^N \sigma_i^T S \sigma_i \quad (7)$$

The constraints at each time interval i also include the follows:

$$\left\{ \begin{array}{l} \forall i, 0 \leq e_i, \\ \forall i, 0 \leq v_i \leq \text{pv}. \end{array} \right. \quad (8)$$

where e and v represent the grid import and electricity consumption of the photovoltaic generation, respectively, pv is the real-time PV generation.

The selected two ZEHs were built in 2017, located in Kitakyushu, Japan. The rooftop PV system equipped, air-conditioner uses PI controller for heating supply to maintain the indoor temperature, the mechanical ventilation rate is 0.5 ac/h.

3 Model Validation and Simulation Result

Construction of a reliable model to accurately capture thermal behavior of the house is key to developing an effective MPC controller. Firstly, to evaluate the accuracy of the dynamic system model, we calibrate the RC thermal model with corrections and evaluate the model with error metrics. The results of RMSE, MBE and CVRMSE of two ZEHs are summarized in Table 1, CVRMSE < 30% and MBE < 10% which satisfied the requirements of ASHRAE Guideline 14-2014 [13].

The proposed MPC strategy is simulated on the validated two ZEHs, and the simulation results of the achieved room temperatures are given in this section. Figure 2 shows the comparison of the room temperatures of ZEH 1 and ZEH 2 under different controls. By comparison, it can be observed that in the baseline (measurement), the temperature fluctuation range is large and the room temperature with MPC strategies is significantly improved. The MPC 1 strategy maintains the room temperature in the comfortable range with stable variation. The MPC 2 strategy introduces photovoltaic generation to increase PV consumption by heating the room when PV is generated without violating the comfort range.

Figure 3 shows the distribution of electricity consumption between the measured and simulated consumption of the two ZEHs air conditioners during the heating period. In order to maintain a stable and comfortable room temperature, the electricity consumption is stable in MPC 1 strategy. In the MPC 2 strategy, more electricity is consumed during daytime hours for PV consumption, thus increasing the room temperature also reduces the electricity consumption of the grid later.

Table 1 Deviation between the simulated and the measured data

Houses	RMSE	MBE (%)	CVRMSE (%)
ZEH 1	0.96	1.65	4.12
ZEH 2	1.72	0.93	9.41

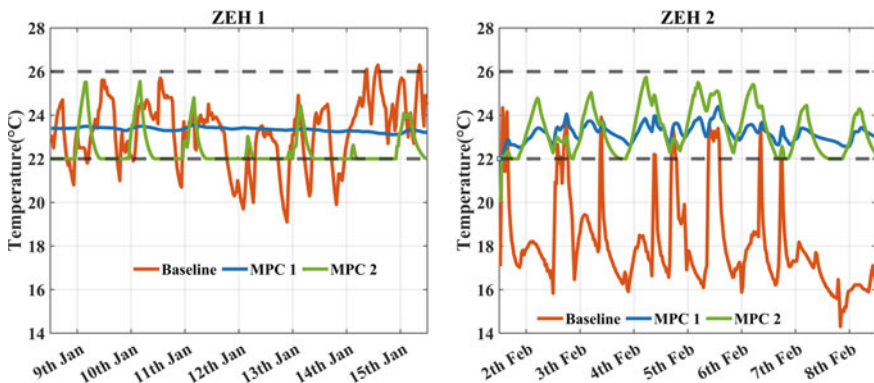


Fig. 2 Comparison between simulated room temperature under different MPC control and measured room temperature

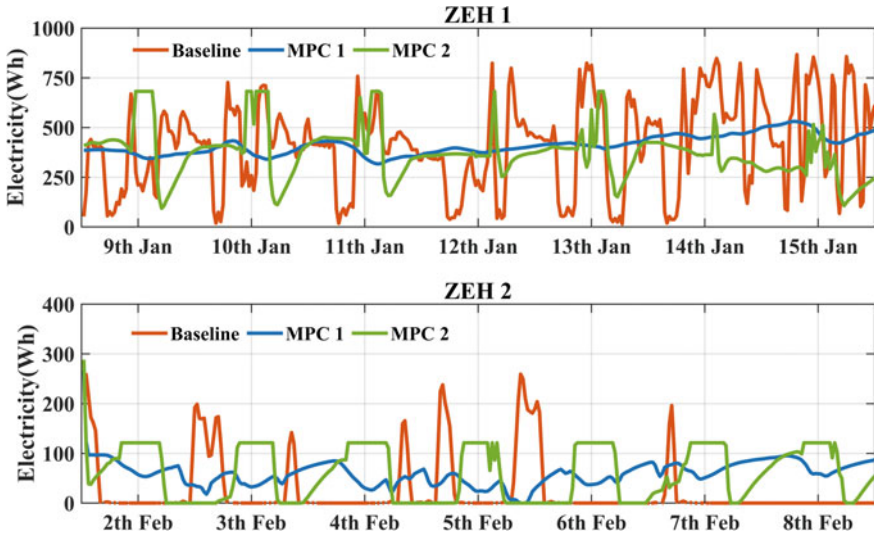


Fig. 3 Comparison between simulated electricity consumption under different MPC control and measurement

4 Conclusion

In this study, we propose a model predictive control (MPC) strategy to realize the optimal control of winter heating systems, aiming to improve the indoor thermal environment, reduce heating energy consumption and increase PV consumption. The proposed approach was tested in two residences by the simulation method. Simulation results show that the proposed MPC controller optimizes the room temperature to keep it in the comfort range and reduces the energy consumption and increases the PV consumption.

Acknowledgements This study was supported by Shandong Natural Science Foundation ‘Research on Flexible District Integrated Energy System under High Penetration Level of Renewable Energy’, grant number ZR2021QE084 and the Xiangjiang Plan ‘Development of Smart Building Management Technologies Towards Carbon Neutrality’, grant number XJ20220028.

References

1. Wang Z, Hong T (2020) Reinforcement learning for building controls: the opportunities and challenges. *Appl Energy* 269
2. Barber KA, Krarti M (2022) A review of optimization based tools for design and control of building energy systems. *Renew Sustain Energy Rev* 160
3. Li Y, Gao W, Zhang X, Ruan Y, Ushifusa Y, Hiroatsu F (2020) Techno-economic performance analysis of zero energy house applications with home energy management system in Japan. *Energy Buildings* 214
4. Zhao X, Gao W, Qian F, Li Y, Ushifusa Y, Yang Z, Yin W, Ge J (2020) Economic performance of multi-energy supply system in a zero-carbon house. *Energy Buildings* 226
5. Xu L, Wang S, Xiao F (2019) An adaptive optimal monthly peak building demand limiting strategy considering load uncertainty. *Appl Energy* 253
6. Thilker CA, Madsen H, Jørgensen JB (2021) Advanced forecasting and disturbance modelling for model predictive control of smart energy systems. *Appl Energy* 292
7. Dongellini M, Naldi C, Morini GL (2021) Influence of sizing strategy and control rules on the energy saving potential of heat pump hybrid systems in a residential building. *Energy Convers Manage* 235
8. Kuboth S, Heberle F, König-Haagen A, Brüggemann D (2019) Economic model predictive control of combined thermal and electric residential building energy systems. *Appl Energy* 240:372–385
9. Hu M, Xiao F, Jørgensen JB, Wang S (2019) Frequency control of air conditioners in response to real-time dynamic electricity prices in smart grids. *Appl Energy* 242:92–106
10. Bird M, Daveau C, O'Dwyer E, Acha S, Shah N (2022) Real-world implementation and cost of a cloud-based MPC retrofit for HVAC control systems in commercial buildings. *Energy Buildings* 270
11. Hou J, Li H, Nord N, Huang G (2022) Model predictive control under weather forecast uncertainty for HVAC systems in university buildings. *Energy Buildings* 257
12. Thieblemont H, Haghghat F, Ooka R, Moreau A (2017) Predictive control strategies based on weather forecast in buildings with energy storage system: a review of the state-of-the art. *Energy Buildings* 153:485–500
13. Ashrae AG (2014) Guideline 14–2014: measurement of energy, demand, and water savings. In: American society of heating, refrigerating, and air conditioning engineers, Atlanta, Georgia

A Study on the Development of a Crop Disease Diagnosis Mobile App Service Based on Deep Learning



Dong In Lee, Ji Hwan Lee, Hoon Seo Jung, Seung Ho Jang, Seung Gi Ryu, Jiann An Yang, Ji Hwan Oh, Jae Won Kim, Se Jong Oh, and Ill Chul Doo

Abstract We introduced a crop disease diagnosis mobile application service that improves the weakness of the previous services. We utilized image captioning model with Inception V3-based encoder and transformer-based decoder for detailed

D. I. Lee

Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: dongeen1@gmail.com

J. H. Lee

Linguistics and Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: yw19845@gmail.com

H. S. Jung

School of Computer Science and Engineering, Soongsil University, Seoul, Korea

e-mail: joel39809@gmail.com

S. H. Jang

English-Korean Interpretation and Translation Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: jason9865@naver.com

S. G. Ryu

Department of Brazilian Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: ysg980423@gmail.com

J. A. Yang

Division of French Language, Hankuk University of Foreign Studies, Seoul, Korea

e-mail: moho191113@gmail.com

J. H. Oh

International Finance, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: kmm8777@naver.com

J. W. Kim

Statistics, University of St Andrews, St Andrews, UK

e-mail: handaijen@gmail.com

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

235

J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,

Lecture Notes in Electrical Engineering 1028,

https://doi.org/10.1007/978-981-99-1252-0_29

explanation. Moreover, object detection model with YOLOv5 was used to display bounding boxes to indicate the damaged region to increase the reliability.

Keywords Crop diseases diagnosis · Image captioning · Object detection

1 Introduction

The number of people participating in urban farming is rapidly growing, however, the technology assisting urban farmers are still limited. There are existing applications that help diagnosing crop diseases, however, the reliability and accuracy are still below expectations. In this research, we propose to create an application that provides precise explanations of the diagnosis and bounding boxes that highlight the affected area of a plant [1]. As a result, the study enables to provide a reliable diagnostic service that boosts the performance compared to the existing services. The study suggests “Dr. Sook Sook” a crop disease diagnosis mobile application service using image captioning and object detection model.

2 Methodology

2.1 Preparing Datasets

For model training, crop images were collected from AI-hub’s “Road Crop Disease Diagnostic Image” and “Facility Crop Disease Diagnostic Image” datasets [2], and non-crop images were collected by crawling on websites (Coupang, Danawa, Lotte Home Shopping, Aerospace and Defense, JoongAng Ilbo, Dong-A Ilbo, Chosun Ilbo). Five crop image data were extracted from the crop image dataset using annotation of metadata in AI-hub to detect normal crops, pepper anthrax, pepper powdery mildew, zucchini powdery mildew, tomato leaf mold, bean leaf blight, and bean leaf spot. Also, non-crop images were extracted from the crawled dataset and to distinguish between crop images and non-crop images.

2.2 Image Captioning Model Construction

Image captioning is a technique to generate sentences that detail the various features of an image. The image captioning model used in this research has an encoder-decoder structure. The features of the image were extracted from the encoder, and the decoder generates sentences based on the features extracted from the encoder. The encoder of the image captioning model in this research used the CNN-based InceptionV3 model. For the decoder model, transformer, a representative language model that generates natural language, was used [3].

For model training, each image was labeled with five sentences which describe the disease of the crop. The sentence was produced by assigning keywords according to the early, mid, and late stages of disease damage. In addition, different sentences with the same meanings were generated by changing the word order and back-translation process so that the model could learn various contexts. Finally, a total of 123,913 images and 619,565 sentences were used to train our model, including images of disease, normal condition, and non-crop objects.

The process of generating sentences for the image is as follows. First, when the image is given in the InceptionV3 model, several features of the image are analyzed. Then, the features are sent to the transformer, which also has an encoder-decoder structure. Initially, the features of the image are sent to the encoder of transformer with its locational information. The information given is analyzed through the self-attention process, and after the process, it is sent to the decoder of transformer. Likewise, the labeled sentences of the image are also input to the decoder. By analyzing the features of images and labeled sentences in the decoder using the self-attention model, the app generates a sentence of prediction [4] (Fig. 1).

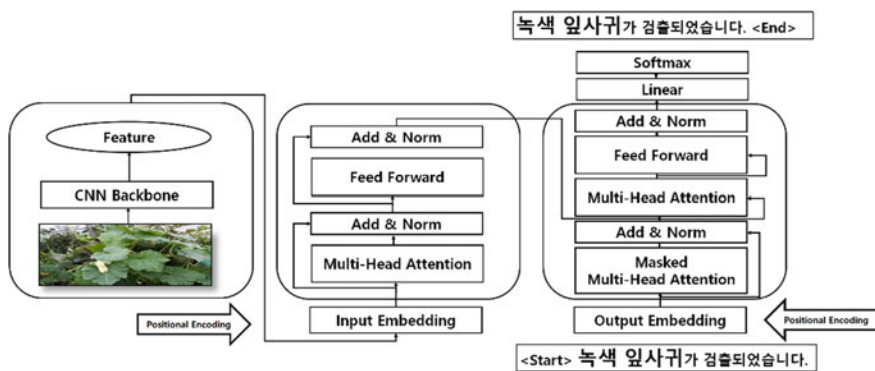


Fig. 1 Structure of image captioning model

2.3 Object Detection Model Construction

Object detection is a computer vision technique that locates instances and bounding boxes of objects in images or videos. We decided to use YOLO algorithm, which was introduced in 2015, as an object detection model. YOLO shows faster and more accurate performance compared to other models such as Fast *R*-CNN [5]. We chose YOLO V5m as a pretrained model from YOLO v5.

Images were labeled after collecting dataset and defining classes for prediction. Since anthracnose and leaf blight were detected in two types of plants, we classified seven predicted classes, which were different from our image captioning model that defines nine predicted classes. Labeling was proceeded with “LabelImg”, an open image label software, and annotations were based on metadata in AI-Hub. Bounding boxes were generated on the area that contains affected areas. There were also a number of cases that box the entire part when the damaged areas were widely spread on the crop.

Meanwhile, since the number of original datasets was about 6000 images, it was not enough to train the model. Therefore, we decided to augment image data on Python environment using “imgaug” package. The total number of image datasets after augmentation process were 31,394 with 25,458 images for training and 5936 images for validation. The images were also resized to 640 * 640 to reduce the training time of the model.

2.4 Designing and Implementing Application

For model porting and function implementation, backend uses Flask and AWS EC2 to build a separate server. If a client sends an image by sending a request, the server runs the object detection model and the image captioning model sequentially, transmitting the detected image with the diagnostic caption to the clients. Also, by allocating some of the capacity of AWS EC2’s hard disk as swap memory, it can prevent overload and maintain the server stable.

Flutter framework and Android Studio were used to build the frontend architecture. Datasets for camera, weather condition, diagnosis, and management suggestions were transmitted through API transaction between the server and the client. The information was displayed on the mobile screen. Overall system structure is illustrated in Fig. 2.

3 Result

Our service has three differences compared to the previous projects. First, our service can diagnose diseases with detailed explanations. Moreover, users do not have to

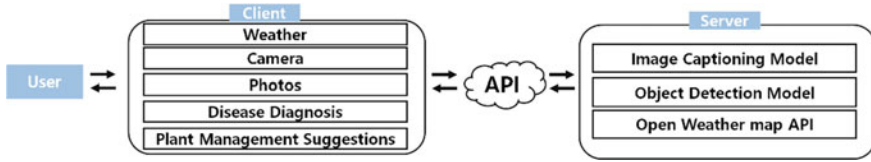


Fig. 2 Application system structure

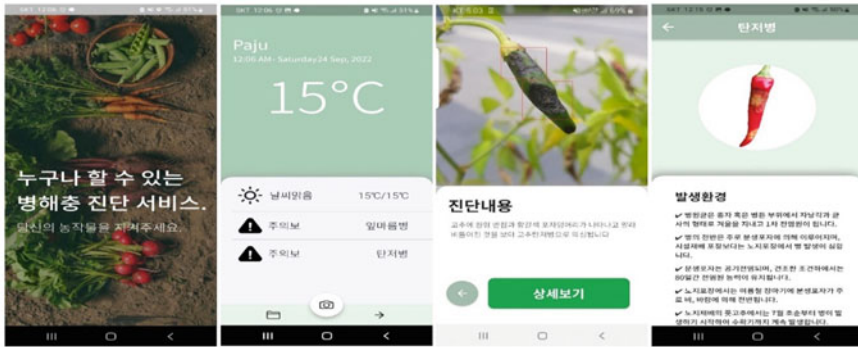


Fig. 3 Main screen of the application

select the crops every time to use the application. This is because our model focuses on the diseases itself rather than the crop classification. Finally, as summarized above, our service detects more crops than existing service, in other words accuracy has improved.

As seen in Fig. 3, users can find the explanations of diseases with bounding boxes on the damaged area after taking photos. “See more” button shows the vulnerable conditions for crops, allowing users to prevent the diseases in advance.

4 Conclusion

The study aims to reinforce the reliability and accuracy of existing crop disease diagnostic services. To do so, the study is designed to present evidence sentences of its diagnosis using image captioning, and to box the damaged area by using object detection. Therefore, compared to the previous projects, the study enables to provide detailed descriptions and boxes of affected areas as its evidence. Consequently, the study expects to satisfy the social needs of the increasing urban farmers and home-farmers. Moreover, it anticipates to contribute to sustainable development and to stimulate the agricultural economy.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the Ministry of Science, ICT (MIST), Korea, under the National Program for Excellence in SW, supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Vinyals O, Toshev A, Bengio S, Erhan D (2015) Show and tell: a neural image caption generator. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3156–3164
2. AI-Hub Homepage. <https://www.aihub.or.kr/>. Last Accessed 19 Sep 2022
3. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst* 30:6000–6010
4. Keras Homepage. https://keras.io/examples/vision/image_captioning/. Last Accessed 19 Sep 2022
5. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 779–788

A Study on the Development of SDGs Implementation Analysis System at Universities Based on NLP



Ha Jin Kwon, Hye Ju Yun, Hyun Bin Jin, Ji Eun Lee, Ju-Young Lee, Sang Hoon Oh, Kwang Wook Seo, Se Jong Oh, and Ill Chul Doo

Abstract This paper presents the SDGs performance status analysis system of universities based on deep learning techniques. Using BERT, the classification of cases corresponding to individual goals was carried out, and TextRank was used to extract keywords for transition cases between overseas universities and domestic

H. J. Kwon

Department of Information Communications Engineering, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: kweon0324@naver.com

H. J. Yun

Linguistics & Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: megan000629@naver.com

H. B. Jin

Political Science and International Studies, Yonsei University, Seoul, Korea

e-mail: rihein@yonsei.ac.kr

J. E. Lee

Department of Thai Studies, Hankuk University of Foreign Studies, Seoul, Korea

e-mail: eunpyo1284@gmail.com

J.-Y. Lee

Artificial Intelligence Convergence(Software&AI), Hankuk University of Foreign Studies, Yongin, Korea

e-mail: summercheriy@gmail.com

S. H. Oh

Department of Statistics, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: eric5050@naver.com

K. W. Seo

Department of Brazilian Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tommypro2130@naver.com

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

universities. Using TF-IDF and cosine similarities, we present examples of overseas universities' SDGs activities that domestic universities can refer to. As a result, we analyzed the current status of SDGs at each university and presented overseas examples that could be further referenced, and created a website that would be helpful for SDGs transition at domestic universities.

Keywords Sustainable Development Goals · BERT · NLP

1 Introduction

Sustainable Development Goals (SDGs) are a universal set of goals that United Nations (UN) member states have pledged to fulfill until 2030. Various organizations including universities, are required to achieve these goals as important stakeholders within society. Accordingly, universities worldwide are publishing annual sustainability reports about their performance status of SDGs, and its implementation is used as one of the indicators of university rating services. However, domestic universities lack SDGs-related performance scores compared to overseas universities, and the number of domestic universities that publish SDGs reports is also limited. In need of systems that can help domestic universities to perform SDGs, this paper focuses on the contents of the service development.

Previous research related to SDGs has mainly discussed ways to implement SDGs at the national or corporate level. Although there exists a paper that deals with SDGs in relation to universities, it also regards universities as an institution in charge of SDGs education, rather than a subject of SDGs implementation [1]. In addition, there also exists a prior study that used natural language processing to create a service that classifies topics in a specific domain [2]. In this paper, we used the text classification model that goes beyond simple subject classification. To be specific, we analyzed the degree of SDGs implementation of universities and established a service to support the implementation of related activities.

2 Methodology

See Fig. 1.

2.1 *Develop SDGs Activity Classification Model*

The purpose of using an automatic classification model is to automatically classify news data from domestic universities into higher classification of SDGs. Considering that among the 17 subsections of SDGs, there are ambiguous subsections to be

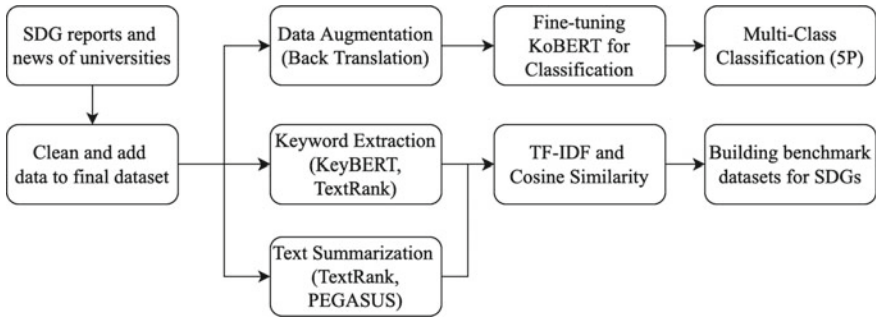


Fig. 1 Flowchart of system

applied to universities, the data to be used for training was reclassified into a higher classification called ‘5P’ presented by the UN. In addition, by providing an analysis of five simplified classifications in consideration of the convenience of service users, we made information to be intuitively delivered compared to 17 detailed classification.

For the model for automatic classification, BERT, a pre-trained language model, was selected [3]. BERT was decided as the best model as it can be applied to various NLP tasks by fine-tuning according to the purpose of the downstream tasks even if the sentence length of the training data is long.

Training data for modeling was collected by web crawling. Universities which publish SDGs reports by category were selected as collection targets, and a total of 7070 training data were crawled from 92 SDGs reports from 60 domestic and overseas universities. Next, we proceeded with the preprocessing of the collected data. Since the model aims to automatically classify Korean text data, the training data was constructed by translating the original data. In order to minimize the deterioration of data quality in the translation process, preprocessing was carried out by diving steps. Unnecessary characters were removed using regular expressions and translated into Korean using the Googletrans library. Finally, by comparing the sentence length of the translated sentences, we set the criteria for checking targets that have not been properly translated. The final dataset was constructed after re-translating the selected target according to the criteria.

In addition, we used data augmentation techniques to obtain additional training data. EDA and back translation were considered with reference to previous studies. In the case of EDA, we selected the RI and RD methods that showed the greatest performance improvement on small datasets in previous studies [4]. The Korean libraries KoEDA and Ktextaug were used to augment Korean text data, and each parameter was applied with 0.05 (alpha) and 1 (repetition). In the case of back translation, Japanese, Chinese, and French were considered based on whether or not the original meaning of the translation was preserved, and finally, the English-Japanese-Korean method was selected [5]. As shown in Table 1, the performance of the model was compared by combining augmented data according to the number of

Table 1 Accuracy of classification model by text data augmentation technique

Aug	Not applied	RI	RD	RI + BT	BT
Accuracy	0.68	0.64	0.71	0.72	0.79

cases, and the model was trained after confirming the training data with the highest performance combination.

2.2 Identifying Similar Cases at Overseas Universities Through Keyword Extraction and Similarity Measurement

Furthermore, keyword extraction and text summarization were conducted (based) on the domestic and international SDGs activity datasets and the individual domestic university news articles datasets that have undergone a classification model. Then, by applying similarity measurement, these summarized datasets were compared to overseas universities' SDGs activity cases to suggest the aspects the domestic universities can consider benchmarking.

Keyword extraction was done by both KeyBert, which leverages BERT embeddings and cosine similarity, and TextRank, a graph-based ranking algorithm [6]. In addition, the text summary method is divided into two categories: an extractive summarization and an abstractive summarization. In this paper, TextRank was used as an extractive summarization, and a comparison was conducted using PEGASUS as an abstractive summarization [7]. Both summarization methods have been examined beforehand to select the more appropriate method for the purpose of this study. Since it was important to present the benchmark case of overseas universities as the original text, TextRank has been selected.

We computed TF-IDF weights and the cosine similarity score to find summary data similar to the main keywords obtained by the process above. For selecting benchmarkable cases of overseas universities, the processing method was applied differently depending on whether overlapping keywords between overseas university keywords and individual domestic university keywords exist. If there are overlapping keywords, we first remove all the overlapping keywords from all overseas university keywords, then combine the remaining keywords into one sentence to measure the similarity with the summarized overseas dataset. And we presented only the top three summary sentences as benchmarkable cases that have the highest similarity scores. If there are no overlapping keywords, we arbitrarily select three summary sentences from each classification category and present them as benchmarkable cases of overseas universities.

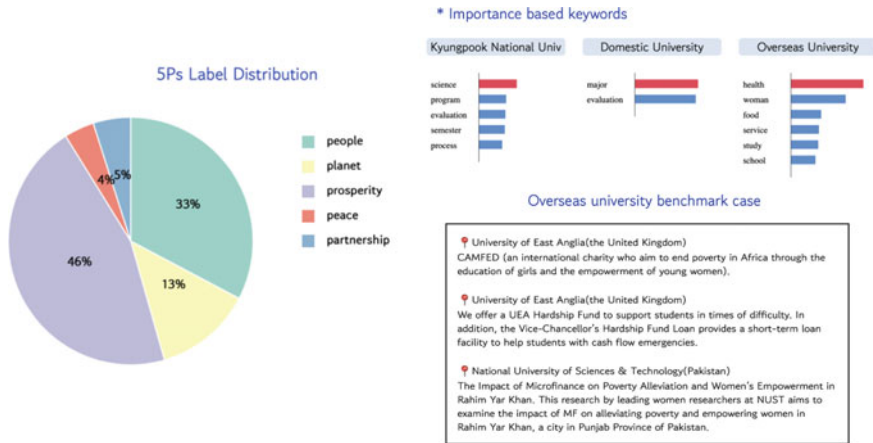


Fig. 2 Analysis dashboards within the sustainED website

3 Result

The web service implemented based on the system is as follows. Figure 2 shows the SDGs analysis dashboard of the university selected on the main page. First of all, the pie chart represents the ratio of news articles for each of the five items derived through an automatic classification model. When each item is clicked, it shows the distribution of keywords based on TextRank by item. Finally, we show examples that can be referred to by the university using similarity-based benchmark algorithms for keywords from other universities that do not exist in the selected university.

4 Conclusion

Recently, as interest in SDGs has increased, the demand for implementation of domestic universities is also increasing, but related activities are still insignificant. Therefore, this paper proposed the design and implementation of an analysis system to support the implementation of SDGs by domestic universities. The system presents the analysis results in the form of a dashboard so that domestic universities selected for analysis can efficiently check activities for SDGs, and provides additional examples for reference when planning related activities. Furthermore, this system is expected to serve as a motivation for raising awareness of SDGs activities of domestic universities that have not been selected for analysis and establishing related plans.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Lee H (2020) Necessity and challenges of sustainable development education (ESD) and SDGs education in universities. *J Liberal Arts* 12:257–284
2. Kim A, Jung Y (2021) Classification of themes of domestic music using KoBERT. In: *Proceedings of the Korean information science society conference*, pp 1738–1740
3. Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. *Google AI Language*, pp 1–16
4. Wei J, Zou K (2019) EDA: easy data augmentation techniques for boosting performance on text classification tasks. In: *Protago labs research*, Tysons Corner, Virginia, USA, Department of Computer Science, Dartmouth College, Department of Mathematics and Statistics, Georgetown University, pp 1–9
5. Edunov S, Ott M, Auli M, Grangier D (2018) Understanding back-translation at scale. In: *Facebook AI research*, Menlo Park, CA & New York, NY, Google Brain, Mountain View, CA, pp 1–12
6. Mihalcea R, Tarau P (2004) TextRank: bringing order into texts. Department of Computer Science University of North Texas, pp 1–8
7. Zhang J, Zhao Y, Saleh M, Liu PJ (2020) PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. In: *Data science institute*, Imperial College London, London, UK, Brain Team, Google Research, Mountain View, pp 1–55

Object Detection Improvements on Skin Burn Image Data via Data Augmentation and Semi-supervised Learning



Dae Hwa Hong, Sin Yeong Kim, Min Kyu Park, Se Jong Oh,
and Ill Chul Doo

Abstract Collecting skin burn image data that is treated as medical data is generally restricted to medical practitioners. Labeling them is even more challenging since it requires professionals to review the whole process of detecting burn area and classifying its degree. A typical machine learning model trained with a small dataset gathered by non-professionals does not guarantee sufficient performance since they are mostly left unlabeled. Data augmentation on images leads to overcoming the limitations of a small dataset. Semi-supervised learning and knowledge distillation enable the application of unlabeled data. This study examines a combination of various augmentation and machine learning methods to maximize performance on detection and classification. Having an $F1$ score of 38.0% in the supervised learning model without any treatment, in comparison, the experimental results showed an $F1$ score of 57.5% in the semi-supervised learning which accumulates distilled unlabeled data and augmented labeled data using rotation, flip, noising, and cutout methods.

Keywords Object detection · Data augmentation · Semi-supervised learning

D. H. Hong

Department of German, Hankuk University of Foreign Studies, Seoul, Korea
e-mail: daehwadavidhong94@gmail.com

S. Y. Kim

Department of Statistics, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: cbrnt1210@gmail.com

M. K. Park · S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

M. K. Park

e-mail: mkpark@hufs.ac.kr

1 Introduction

Skin burn, compared to other external injuries, tends to be less consciously observed and is often treated with a first aid in an inappropriate manner. Moreover, people underestimate the aftermath of a burn so that they do not visit any clinical institute or even neglect any first aid on it.

We intended, therefore, to provide an object detection-based AI service which classifies a degree of a burn from a photo taken to guide casualties to proper treatment. Access to medical data, however, is generally restricted to medical practitioners only, and hence, it is difficult to construct a large dataset. Another following issue regarding medical data is that they require labels reviewed by professionals while there exists few gatherable labeled data. In this paper, we build a machine learning model with a small dataset consist of labeled and unlabeled data and derive the best performance. Here we examine an optimal combination of data augmentation and training methods on burn image data with experiments.

Tran et al. constructed convolutional neural network (CNN) specialized to burn images recognition, while it remains to have a limited accessibility since they built a high-quality dataset in cooperation with hospital faculties at Vietnam and targeted medical practitioners as users [1]. Şevik et al. focused on texture-based feature extraction from 105 images of which dataset constructed jointly by the burn centers in Turkey [2]. They, however, might have achieved better performance by applying techniques to overcome the shortage of data.

2 Method

2.1 Data Collection

A dataset of previously labeled and classified skin burn images was found on public online dataset hub [3]. Among 1227 labeled burn image data, 125 images were sorted out evenly per degrees 1, 2, or 3 and above. Additionally, 1000 distinct unlabeled burn images were collected via web crawling.

2.2 Data Augmentation

Data augmentation, a method of preventing overfitting and enhancing performance of a machine learning model by increasing the diversity of originally labeled data, is executed using Python library “Albumentation” which is economic to memory and single instruction multiple data (SIMD) register usage and is effective at common transform operations on uint8 data [4] (Fig. 1).

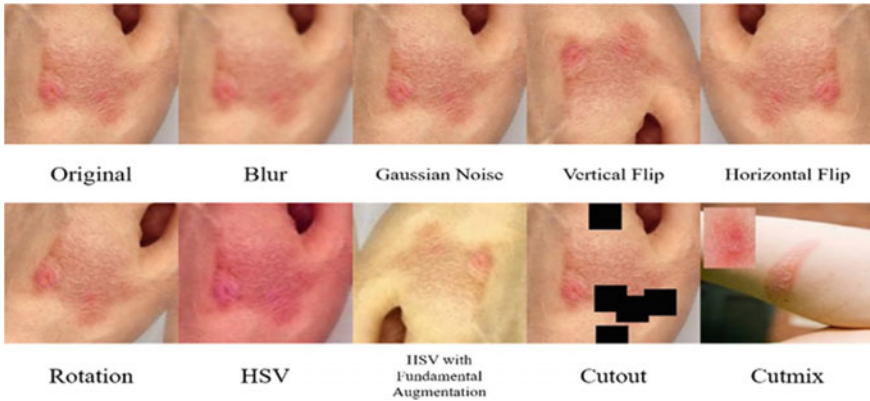


Fig. 1 Examples of data augmentation methods of an image

Data are first resized to 256×256 pixels and then are performed fundamental image augmentation via blur, Gaussian noise, vertical and horizontal flip, rotation, and HSV (Hue, Saturation and Value) adjustment.

Cutout. Cutout generates a new image by eliminating certain area from an original image. This improves performance by letting a model to examine features that were not in focus on the original image [5].

Cutmix. To compensate the loss of information via cutout method, cutmix lets a model to focus also on small image patch attached from one of training data and hence improves performance [6]. Here, saliencymix operation was incorporated with meaningful portion using saliency map [7]. Applying saliencymix via OpenCV library, however, tends to crop image patch bigger than its object's bounding box, thus the cutmix operation was performed based on the bounding box of cropping image.

2.3 Machine Learning

It is generally known that YOLO model outperforms faster *R*-CNN when it comes to object detection in images [8]. YOLO model is advantageous over accessibility as an open-source machine learning engine and expandability on training with other medical image data. We chose to experiment on the YOLOv5 model for these reasons (Fig. 2).

After applying *K*-Fold cross-validation on dataset, which gets more effective as each performance upon partitioning train and validation sets varies more, image augmentations are applied [9]. To utilize both labeled and unlabeled data in training, pseudo-labeling method from semi-supervised learning was implemented [10].

Thereon knowledge distillation was applied, a method of transferring parameters of an initial model to a new model. Here, we name two different approaches: type 1

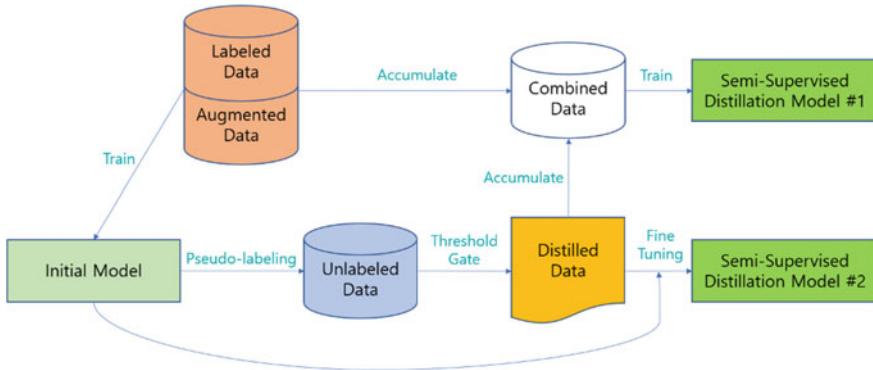


Fig. 2 Machine learning model diagram

generates a model of which knowledge transferred from ensemble of models; type 2 generates a small model sharing parameters with a larger original model [10, 10].

2.4 Measurement

Performance of an object detection model can be quantitatively measured by precision, recall or *F1* score that denotes the former two simultaneously [10]. The experiment compares performances of each model based on *F1* score above.

3 Result

Table 1 represents the improvement of performance on supervised learning between random- and cross-validation-based partitioning of train-validation data subsets. Onward experiments are based on the train data partitioned above. Table 2 delineates results of supervised learning on labeled data with diverse data augmentation methods. Top two results of augmentation combinations were trained under semi-supervised learning with knowledge distillation type 1 and 2. Consequently, as shown in Table 3, the performance is maximized having an *F1* score of 57.5% when labeled data are augmented with fundamental augmentation and cutout and are trained by semi-supervised distillation type 1.

Table 1 Supervised learning with *K*-fold cross-validation

Method	Precision	Recall	<i>F1</i> score
Supervised learning	0.261	0.7	0.380
Cross validation	0.326	0.615	0.426

Table 2 Supervised learning with data augmentation

Method	Precision	Recall	F1 score
Fundamental augmentation (FA)	0.397	0.745	0.518
Fundamental with HSV augmentation (FHA)	0.473	0.61	0.533
Cutout	0.422	0.62	0.502
Cutmix	0.383	0.27	0.317
Cutout with FA	0.467	0.71	0.563
Cutout with FHA	0.466	0.645	0.541
Cutmix with FA	0.492	0.64	0.556
Cutmix with FHA	0.486	0.595	0.535

Table 3 Semi-supervised learning with knowledge distillation (SSD) with augmentations

Method	Precision	Recall	F1 score
SSD type 1 with Cutout-best	0.457	0.775	0.575
SSD type 2 with Cutout-best	0.364	0.8	0.500
SSD type 1 with Cutmix-best	0.449	0.635	0.526
SSD type 2 with Cutmix-best	0.311	0.82	0.451

4 Conclusion

This research is initiated to utilize burn data of which gathering is restricted to medical practitioners and hence can only be constructed small. Especially, we intended to maximize performance from small dataset containing labeled and unlabeled data. To sum up, the experiment shows that the best performance on skin burn degree classification is obtained when labeled data are augmented with rotation, flip, noising, and cutout methods and are trained with semi-supervised learning which accumulates distilled unlabeled data. The research is expected to be applicable and expandable to detection and classification on any other dermatopathy.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Tran H, Le T, Nguyen T (2016) The degree of skin burns images recognition using convolutional neural network. *Indian J Sci Technol* 9(45):1–6
2. Şevik U, Karakullukçu E, Berber T, Akbaş Y, Türkyılmaz S (2019) Automatic classification of skin burn colour images using texture-based feature extraction. *Instit Eng Technol* 13(11):2018–2028
3. Skin Burn Dataset|Kaggle. <https://www.kaggle.com/datasets/shubhambaid/skin-burn-dataset>. Last accessed 03 Sep 2022
4. Buslaev A, Iglovikov V, Khvedchenya E, Parinov A, Druzhinin M, Kalinin A (2020) Albumentations: fast and flexible image augmentations. *Information* 11(2):1–12
5. DeVries T, Taylor G (2017) Improved regularization of convolutional neural networks with cutout. 1–6. [arXiv:1708.04552](https://arxiv.org/abs/1708.04552)
6. Yun S, Han D, Oh S, Chun S, Choe J, Yoo Y (2019) CutMix: regularization strategy to train strong classifiers with localizable features. In: *Proceedings of the IEEE/CVF international conference on computer vision (ICCV) 2019*. IEEE, South Korea, pp 6025–6030
7. Uddin A, Monira S, Shin W, Chung T, Bae S (2021) SaliencyMix: a saliency guided data augmentation strategy for better regularization. In: *International conference on learning representations 2021*. ICLR, Austria, pp 2–9
8. Lee Y, Kim Y (2020) Comparison of CNN and YOLO for object detection. *J Semicond Disp Technol* 19(1):89–90
9. Scikit-Learn Homepage. https://scikit-learn.org/stable/modules/cross_validation.html. Last accessed 15 Aug 2022
10. Murphy K (2022) *Probabilistic machine learning: an introduction*. 1st edn. The MIT Press, Massachusetts, pp 171–173, 634–635, 645–646
11. Abbasi S, Hajabdollahi M, Karimi N, Samavi S (2020) Modeling teacher-student techniques in deep neural networks for knowledge distillation. In: *2020 International conference on machine vision and image processing (MVIP)*. IEEE, Iran, pp 1–2

Establishment of NLP-Based Greenwashing Pattern Detection Service



Ji Sue Kim, Jun Bo Sim, Yoon Ju Kim, Min Kyu Park, Se Jong Oh,
and Ill Chul Doo

Abstract This study aims to establish a search service that provides greenwashing pattern detection and infographic of enterprises through AI-based NLP technology. The purpose of it is to provide decision indicators by checking data on greenwashing trends of companies by users, including investment institutions and general consumers. We used a deep learning model to adopted evaluation factors to score the ‘greenwashing’ and collected the title of media data with active monitoring of companies based on those factors. In addition, construct greenwashing sentence discrimination model and a corporate greenwashing scoring model using the BERT model. As a result, a service provided scores using data collected through its keywords through visualization and the process of calculation when users input the name of the company.

Keywords Greenwashing · BERT · Visualization

J. S. Kim

Linguistics and Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: luxred@naver.com

J. B. Sim

Information Communications Engineering, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: sjbssk0809@gmail.com

Y. J. Kim · M. K. Park

South Slavic Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: kimyoonju518@gmail.com

M. K. Park

e-mail: mkpark@hufs.ac.kr

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

1 Introduction

This study focused on the emerging ‘greenwashing’ problem in the ‘environment’ area while companies consider ‘sustainable management’ using ESG indicators with environmental/social responsibility/correct governance [1]. Greenwashing is a combination of ‘Green’ and ‘WhiteWashing’, which refers to the action of a company using green characteristics of an organization, product, or person on insufficient or unrelated grounds to target economic profits such as stock price increases and sales increases [2]. Cooperates related to greenwashing are weakening brand image, can give a sense of betrayal to losing customer base, and lead to fatal consequences [3]. There are two ways to identify current greenwashing trends in companies. First, it is necessary to find sustainability reports (ESG reports) enterprises issued. Second, sites and papers such as ‘SESAMm’ that externally rate cooperates’ ESG activities and greenwashing trends are required [4]. However, additionally collected data is needed to discriminate the information from the ostensible direction of management because domestic companies have so far been voluntary information disclosure rather than an obligation [5], and it is self-disclosed by the company in the former case. Therefore, media data on corporate trends were collected and analyzed to evaluate greenwashing trends using reporting information outside the company. In the second case, the investment report is less accessible to ordinary consumers because it provides reference data to insiders from an investment perspective on cooperates. This study implemented the service in the web form to provide decision indicators that increase consumers’ access to information so that users can get infographic data on greenwashing trends only by searching for corporate names. At this time the investors, consumers, and corporate users from the inclusion and expanded their range. In another study, there have been attempts to identify greenwashing trends in companies using the BERT model as a natural language processing technique, but as a result, the accuracy level of the model was insufficient, which was the form of a report as well [6]. Therefore, we propose a greenwashing detection search web service that advances existing research to reconstruct the criteria to discriminate greenwashing via the BERT model that has refined the utilization.

2 Methodology

2.1 *Configuration Greenwashing Assessments and Collecting Data*

First, we established standards for determining greenwashing by referring to well-known articles since there is no internationally standardized criteria for ESG. Referred articles were ‘Seven Sins of Green Washing’ by Terra Choice, a Canadian environmental marketing company [2], and ‘the integrated framework article for greenwashing evaluation’ by the Sustainability Journal of MDPI [7]. Directness,

internality, and consistency are three established comprehensive evaluation criteria. Directness evaluates the direct relationship between a company and greenwashing, internality evaluates eco-friendly message behavior, and consistency is a measure to review and evaluate the transparency of information such as production and supply stages and external disclosure in the aspect of ESG management. The greater the direct mention of eco-friendly management, the greater the question of consistency and internality, and the greater the greenwashing trend. Then, we set up a scoring formula so that we could see it objectively. The formula is as follows when x is immediate, y is internal, and z is coherence ($0 < x, y, z < = 1$):

$$x(y + z) * 5$$

Each number can have a maximum of 1 since each x , y , and z value is greater than 0 and less than or equal to 1. The greenwashing score has the maximum value of 10 when the marketing score for eco-friendly message behavior and transparency of corporate trends, the number of internality and consistency is max, and the degree of directness, the degree of greenwashing, is also the maximum. At this time, if the directness score is low, the score for greenwashing is reduced.

Collect natural language data based on the above evaluation criteria. We crawled about 17,000 Naver media headlines to collect external monitoring data rather than internally published information. Next, pre-processed collected data for purification. First, we removed duplicate headings and column names because it may lower the accuracy of greenwashing discrimination. Removed the repeated company names and all the special characters as well because it would also cause an error in the identification of greenwashing throughout the entire sentence. Finally, labeled the data according to the above criteria.

Furthermore, we constructed a BERT model based on the data dictionary obtained above. We compared setting parameters such as Early Stopping and Epoch.

2.2 Constructing Sentiment Analysis Model of Greenwashing

The experiment was conducted to evaluate the context judgment accuracy of the above BERT model. After that, when a user searches for a company name, the titles of articles that can determine the degree of greenwashing are crawled and entered into the model with ‘company name + keyword (the degree of greenwashing of the company is preset for analysis)’. A prediction algorithm was implemented that automatically retrieves press data and scores the degree of greenwashing (Fig. 1).

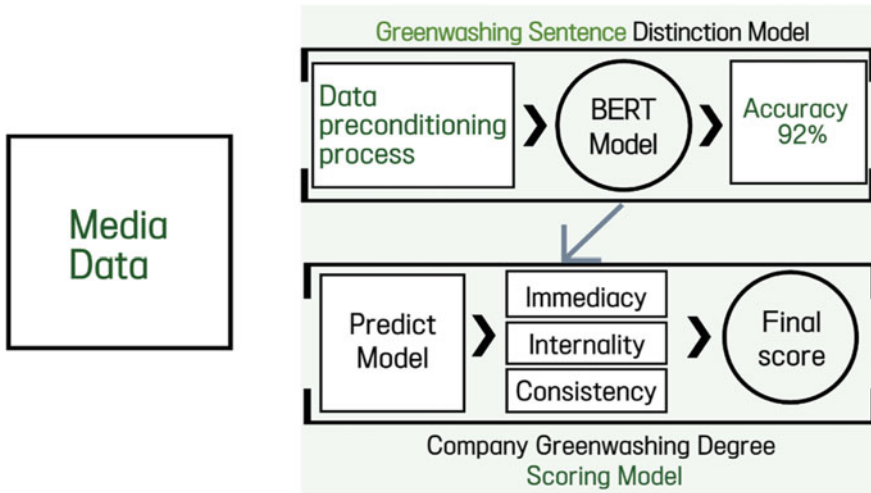


Fig. 1 Describing the structure of the greenwashing sentence discriminant model and the scoring model with BERT

3 Results

The setting parameters of the learned BERT model using the processed data were determined to be 92% accurate, with $max_len = 64$, $batch_size = 64$, $warmup_ratio = 0.1$, $num_epochs = 10$, $max_grad_nom = 1$, $log_interval = 200$, and $learning_rate = 5e-5$. The followings are the experimental results for the greenwashing sentence discriminator. (Fig. 2).

First, it detected greenwashing indifference when the word was input because it simply refers to greenwashing. As a second example, the model detected the general association with greenwashing when the input sentence was ‘Is it truly eco-friendly’, a mixture of eco-friendly and questionable expressions. Finally, it strongly detected greenwashing when the input sentence was ‘Enterprise’s eco-friendly marketing turns out to be consumer deception’ because data clearly shows the behavior of greenwashing. Through the above examples, it is possible to confirm that the model detects greenwashing by grasping the context of the entire sentence. Implemented the web after building the BERT model so users could apply it. On the left side of Fig. 3, directivity, internality, and consistency were visualized and provided using bar diagrams based on the greenwashing score obtained through the predicted model. The right side of the figure is word cloud, which visualizes and shows highly relevant words when searched with keywords that meet three evaluation criteria after the company name in the process of scoring greenwashing, and it seems that meaningful words appear.

하고싶은 말을 입력해주세요 : 그린워싱
 >> 입력하신 내용에서 그린워싱 무관 감지.
 0

Translation:
 Enter words to perceive: "Greenwashing"
 >> Detected Greenwashing indifference from your input.
 0

하고싶은 말을 입력해주세요 : 진짜 친환경인가
 >> 입력하신 내용에서 그린워싱 보통 감지.
 1

Translation:
 Enter words to perceive: "Is it truly eco-friendly"
 >> Detected a general connection to greenwashing from your input.
 1

하고싶은 말을 입력해주세요 : 어느기업 친환경 마케팅하고 소비자 기만
 >> 입력하신 내용에서 그린워싱 강력 감지.
 2

Translation:
 Enter words to perceive: "An enterprise's eco-friendly marketing turns out to be consumer deception."
 >> Detected an intense connection to greenwashing from your input.
 2

Fig. 2 Example of greenwashing sentence learning and evaluation model with BERT model



Fig. 3 Web-implemented visualization screen—word cloud and bar graph

4 Conclusion

This study began paying attention to the emerging ‘greenwashing’ problem in the environment area as companies consider ESG management which refers to the environment (*E*), social responsibility (*S*), and correct governance (*G*). The Greenwashing word is a combination of Green and Whitewashing, which promotes eco-friendliness, but it has a problem that the actual move deceives consumers. Accordingly, a ‘Greenwashing detection AI service’ is intended to help solve the problem by providing decision-making indicators. We implemented a website that offers intuitive and visual information on greenwashing by reading data with relevant keywords even if users search only for the company name. It aims to establish a service that can be freely monitored in the area of citizens’ interest and is being databased through more

professional information such as ESG evaluation and corporate reporting. Furthermore, it is possible to deliver reliable information to related decision-makers if it becomes cross-platform. As a result, we expect this study to be a turning point in practical fields further than NLP research.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Jeongki L, Jay Hyuk R (2020) Current status and future directions of research on sustainable management: focusing on the ESG measurement index. *J Strat Manage* 23(2):65–92
2. TerraChoice (2009) The seven sins of Greenwashing, Summary Report
3. Gae YD (2022) The effect of green marketing and greenwashing on purchase intention. Master’s thesis in Korea Graduate school of KonKuk University, Seoul
4. SESAMm homepage. <https://www.sesamm.com/>. Last accessed 24 Sep 2022
5. Dong Soo K (2022) Recent ESG disclosure trends and greenwashing. *Future Growth Res* 8(1):101–107
6. Min YL, Zheng Q et al. (2022) MGTF 490 capstone project: identify “Greenwashing” funds using NLP in firms’ prospectuses, as you sow
7. Nemes N, Scanlan SJ (2022) An integrated framework to assess greenwashing. *Sustainability* 14(8)

Multiclass Classification of Online Comments Based on Toxicity Scale



Jo eun Kang, Jisun Kim, Hyeon su Do, Sihyun Oh, Si hwa Lee,
Min jee Jeung, Min seo Choi, Se Jong Oh, and Ill Chul Doo

Abstract Our research specifies the criteria for categorizing comments and supplements the limitations of the existing services that filter out only some of the direct malicious comments centered on specific vocabulary and phrases. In comparison to prior research detecting three sentiments as positive, negative, or neutral, we subdivided expressions into six categories. Our model applying classification downstream

J. eun Kang

Department of Interdisciplinary Graduate Program of Linguistics and Informatics, Yonsei University Graduate School, Seoul, Korea
e-mail: j0eun@yonsei.ac.kr

J. Kim

Department of Political Science and International Relations, Hankuk University of Foreign Studies, Seoul, Korea

H. su Do

Department of Japanese Language and Culture, Soochow University, Taipei, Taiwan

S. Oh

Department of English Linguistics & Language Technology, Hankuk University of Foreign Studies, Seoul, Korea
e-mail: oshthepublic@naver.com

S. hwa Lee

Department of Hungarian, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: 129_hwa@naver.com

M. jee Jeung

Department of French Studies, Hankuk University of Foreign Studies, Yongin, Korea

M. seo Choi

Department of Information Communications Engineering, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: yhari214@hufs.ac.kr

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

tasks to a pre-trained KcBERT model draws results reflecting the meaning and social context of the comments. An improved accuracy compared to previous studies shows a meaningful outcome.

Keywords Korean NLP · Multiclass classification · Toxic comments

1 Introduction

Most prior studies related to the classification of web-site comments only judged whether comments are hate speech and merely suggested the necessity of classifying comments according to detailed criteria/standards. In addition, there are some papers that contain some scale of toxicity, but there is a limitation of studying only malicious comments or negative sentences without considering neutral and positive ones. We propose a comment classification model differentiated from these existing studies. First, the concept of ‘Toxicity Scale’ was devised. We set our own standards as several types, referring to previous studies. Second, the scope of analysis was set to all comments. We didn’t classify only malicious comments, but looked at all types of them, including neutral and positive ones. Third, it is classified in 6 categories. In contrast to prior research detecting positive, negative or neutral sentiment, expressions could be subdivided in detail.

Our research specifies the criteria for categorizing comments and proves the possibility of classification. As research focuses on Korean, it should be kept in mind that there might be differences in English-based analysis and interpretation. According to the Korea Press Foundation, 79.2% of respondents watch the news on Internet portals or search engines, and 86.7% of them answered especially on Naver, a major portal site in Korea [1]. Thus, we conducted the study by classifying comments on Naver news which has the largest number of active users.

2 Experimental Setup

See Fig. 1.

2.1 Dataset

Our dataset consists of comments from articles in politics, society, and lifestyle/culture categories which were released between August, 2022 and September, 2022. Besides, the Korean Hate Speech data from Korpora, which consists of ‘Entertainment’ news comment from Naver, was added for a bigger dataset. These crawled text data are respectively labeled depending on 6 criteria we

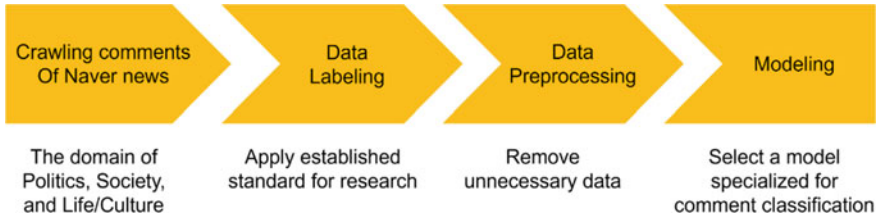


Fig. 1 Project process

established on the basis of previous studies. The dataset was augmented through Easy Data Augmentation techniques, namely synonym replacement and random deletion. For the data preprocessing, we used the same method with the one of KcBERT preprocessing. First, special symbols such as emoji were included as subject of analysis as well as the Korean and English texts. Secondly, repeated character strings were compressed into a couple of strings. Lastly, duplicate comments were removed. After preprocessing, we finally got a dataset made up of 67,434 comments and divided these into three parts: training, validation, and test. The ratio of each portion is 80:15:5.

2.2 Annotation

There are 6 labels from 0 to 5. Level 0 stands for positive comments. Neutral comments were labeled as 1, whereas negative sentiment ranges from 2 to 5. Among negatives, 5 means the highest toxicity scale (Table 1).

Table 1 Establish labeling criteria based on objectionable scale

Sentiment	Label	Criteria
Negative	5	Hate speech against both individuals and groups
	4	Discrimination, degrading attack against particular individuals, violent inflammatory expression for unspecified individuals or Expressions that make its readers feel insulted and humiliated
	3	Degrading attack for unspecified individuals or insults, slurs, and derision using offensive slangs
	2	Opinions or evaluations that simply reflect negative sentiment
Neutral	1	Simple facts and experiences that do not express negative or positive sentiment
Positive	0	Expression with positive sentiment such as encouragement, support, hope, condolence

2.3 Model

We selected KcBERT for the research. Since KcBERT is a NLP model pre-trained with Naver news comments, it was expected to reflect the features of comment data better than other models; slangs, typos, and colloquial expressions. In the training process, the automatic optimization framework adjusted the detailed hyper parameter and the epoch was increased until overfitting doesn't occur. Finally, it was set as 33.

3 Results

Table 2 gives information on the model metric. The model displayed train accuracy of 97% and validation accuracy of 89% after its learning and validation processes. Train loss recorded approximately 0.1 while validation loss reached 0.4. When tested on 3049 comments, the model correctly predicted the toxicity levels of 88.6% of them (Fig. 2).

Table 3 shows *F1* scores of the models trained and used in relevant studies and this study. For ternary detection of hate speech in entertainment news comments, Moon et al. and Lee & Park used KoBERT and ordinal LSTM-Coral, respectively, [2, 2]. Hwang et al. performed binary detection of whether the comment was toxic or not through training a CNN-LSTM model with a corpus containing online news comments, tweets, and so forth [4]. The scores demonstrate that despite being given a 6-class classification task, our model outperformed the previous studies in which only binary and ternary detections were conducted, proving that our model produced a significant result.

Table 2 Model metric

Train accuracy	Train loss	Validation accuracy	Validation loss
0.9742	0.1226	0.8924	0.4067

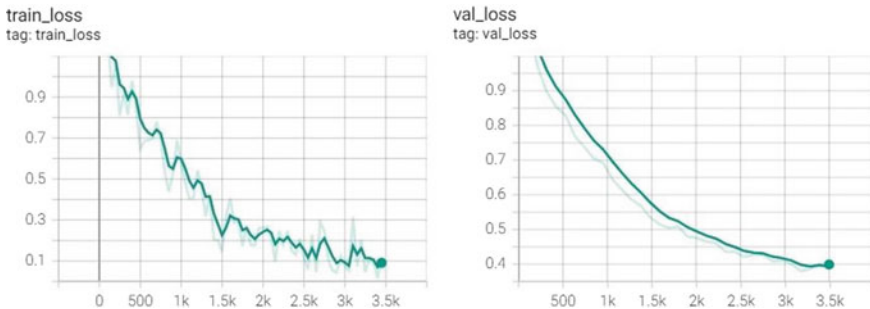


Fig. 2 Train loss value (left) and validation loss value (right)

Table 3 Comparison of *F1* scores of trained models in previous studies and our study

Moon et al. [2]	Lee & Park [3]	Hwang et al. [4]	Our
KoBERT	LSTM-cor	CNN-LSTM	KcBERT
0.57	0.59	0.84	0.89

Table 4 Examples of prediction

Comment	True label	Predicted label
화교녀이엿군 어쩐지 짱개냄새가나더라 So she was a chinese c*nt? I knew it. Smelled like chinks	5	5
31일동안 정부가 방치했다니 정말 경악을 금치 못하겠다 역시 이번5년은 망했다 So shocked that the government neglected them for 31 days. Apparently we are doomed for the next 5 years	2	2
배우들 연기 찰지네요 연출도 좋아요 잘보고있습니다 The actors' performances were lit and I liked the cinematography too. Totally enjoying it	0	0
급여 받아 쳐먹는 인간 N레기들 완전히 많음 다 잡아 처벌해야 There's a whole lot of scums who get paid and do nothing. Must seize and punish them all	3	3

Table 4 shows an example from each level in which the predicted results of the model match the labels manually classified during the data set construction process.

According to the criteria, the first example was labeled as 5 because it contained explicit profanity ‘c*nt’ and hateful expression ‘chinks’. The second example is classified as label 2 because it represents a simple negative opinion and evaluation. It shows that it is classified as 0 for cheering, encouraging, and positive expressions, such as ‘liked’, and 3 for slanderous comments containing negative slang such as ‘scums’.

4 Conclusion

Our prediction results classified into six levels by fine-tuned KcBERT show higher performance than the prediction results of previous studies, and it is also verified that our model reflects the meaning and social context of comment. Our model is meaningful in that it establishes the criteria of classifying comments, which was inadequate before, and supplements the limitations of the classifying technology that filters out only some of obviously offensive comments centered on specific vocabulary and phrases. In the future when we expand our model to an API that can be applied to actual comment boxes, users will be able to actively view comments according to the degree of their acceptable toxicity levels. We expect to contribute

to preventing users from negative comments while guaranteeing the freedom of expression of commenters as much as possible.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Korea Press Foundation: Press Reception Survey (2021). Korea, pp 85–88
2. Moon J, Cho W, Lee J (2020) BEEP! Korean corpus of online news comments for toxic speech detection. In: Proceedings of the eighth international workshop on natural language processing for social media, pp 25–31
3. Lee S, Park S (2021) Hate speech classification using ordinal regression. In: Proceedings of the Korean society of computer information conference, pp 735–736
4. Hwang H, Kim H, Chung J, Chung H, Seo C, Lee S (2021) A chrome plug-in for harmful text filtering based on CNN-LSTM. In: Proceedings of the Korea information processing society conference, pp 543–546

Developing a Speech Recognition Service for Korean Speakers with Dysarthria



Go Woon Choi, Min Hyuk Kim, Ha Young Park, Young Hae Choi,
Se Jong Oh, and Ill Chul Doo

Abstract The purpose of this study is to provide an API for converting the speech of people with dysarthria into a text form by constructing a model that learns the speech characteristics of Korean speakers with dysarthria. A speech recognition model was constructed by using the Korean speech recognition open-source toolkit (Kospeech) which embodies DeepSpeech2 model. Ten thousand voice files which recorded the speech of people with dysarthria and the corresponding transcription files were also collected. Both files were augmented and used for model training. By performing the WER/CER evaluation, it was confirmed that the model constructed in this study recognized the speech of the Korean speakers with dysarthria better than the existing speech recognition model.

Keywords Speech recognition · Dysarthria · Korean STT

G. W. Choi

Department of Thai Studies, Hankuk University of Foreign Studies, Seoul, Korea
e-mail: cku7808@naver.com

M. H. Kim

Department of English Linguistic and Language Technology, Hankuk University of Foreign Studies, Seoul, Korea
e-mail: mhkim0929@naver.com

H. Y. Park

Department of Big Data Science, Korea University, Sejong, Korea
e-mail: ij020554@korea.ac.kr

Y. H. Choi

Linguistics and Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea
e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

1 Introduction

Speech recognition not only provides various conveniences in daily life, but is also rapidly accepted throughout society and replaces various areas. However, existing speech recognition services have limitations in recognizing speech of people with dysarthria which has unusual speech characteristics. Many countries, including the US, have been accumulating a considerable amount of study as their awareness of the problem has gained consensus early on, but the situation in Korea has yet to reach these. So, most Korean speakers with dysarthria cannot use various conveniences using speech recognition [1]. Therefore, in order to solve the social inequality caused by these technical problems, we intend to develop an improved speech recognition service by constructing a model using deep learning so that Korean speakers with dysarthria can use speech recognition service equally with non-disabled people.

2 Methodology

See Fig. 1.

2.1 Data Preparation

The data used to develop the speech recognition service is AI Hub's "dysarthria speech recognition data", and among the three speech classifications according to the type of disability, about 120 GB of speech deafness and cerebral neuropathy data excluding laryngeal disorders were used.

First of all, for speech recognition in sentence units, the voice files and transcription files were divided into sentence units. In the case of voice files, it was divided using the Pydub library, and in the case of transcription files, it was divided based on (period), ! (exclamation mark), and ? (question mark), and the preprocessing of removing special symbols in sentences was additionally performed by using regular expressions. In addition, we performed a data augmentation technique that randomly

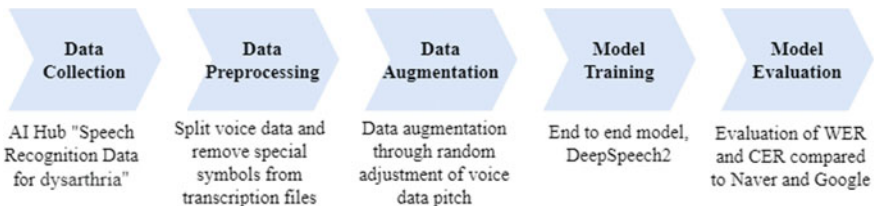


Fig. 1 The process of development of a speech recognition service for people with dysarthria

adjusts the pitch of voice data to increase the number of data and prevent overfitting of the model. Through data augmentation techniques, about 10,000 data that has been initially preprocessed were augmented with about 20,000 training data.

2.2 Building and Training the Model

In speech recognition, an end-to-end (E2E) structure, which is one of the deep learning methods that can output text without going through a separate phoneme or morpheme conversion process, was used, and DeepSpeech2 with good processing efficiency was selected as the base model [2].

In building the model, we used the Korean Speech Recognition open-source toolkit (Kospeech) in which several end-to-end (E2E) models, including the DeepSpeech2 model, were embodied. Then, we modified according to the target model and learning environment. In addition, the accuracy of the speech recognition result was improved by removing the duplicate phonemes caused by the CTC loss function inside the model.

3 Result

To evaluate the performance of the speech recognition model more accurately, it went through a process of comparing with that of other speech recognition APIs. First, we compared the results of speech recognition for the same speech data uttered by people with dysarthria about speech recognition APIs of NAVER and Google, which are supporting Korean and the speech recognition model in this study.

Figure 2 is the result of speech recognition by three speech recognition models we mentioned, for speech data of the sentence “while I took off the baby’s clothes and put them on, I even combed the baby’s hair with a brush. (ə ˌdʒiːər əʊz eul beos ˌdʒiːər ˈ əʊ ˈp ˌhɑɪ myeon ˈsiːəʊ ˈbɪs eu ˈrəʊ ˈmiːəʊ ɪ duː ˈbɪs gyeo ˌdʒuː eoss da)”.

The speech recognition model used in this study showed higher accuracy for speech recognition for the voice of people with dysarthria, compared to NAVER’s speech recognition API, which boasts higher accuracy than Google’s speech recognition API in Korean speech recognition.

Secondarily, we compared the performance evaluation results of three speech recognition APIs to verify the performance of the speech recognition model in this study using standard metrics. As metrics for performance verification, we use Word Error Rate (WER) and Character Error Rate (CER), which are normal metrics for evaluating models in the field of speech recognition [3], and compare performance by randomly choosing 50 speech data of dysarthria patients that were not used for model learning.

Through the two evaluation metrics, it was confirmed that the performance of the speech recognition model in this study was superior to that of other speech

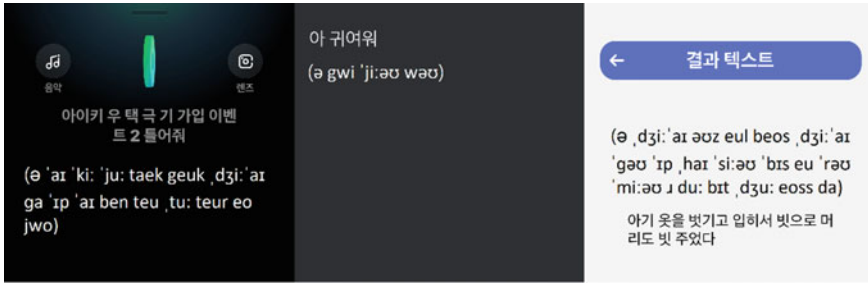


Fig. 2 Speech recognition results from NAVER’s speech recognition API, Google’s speech recognition API and a speech recognition application developed with our speech recognition API

Table 1 The performance evaluation results of the NAVER speech recognition API, Google speech recognition API, and the speech recognition model of this study

Evaluation metric	NAVER API	Google API	Model
Word error rate (WER)	0.857	0.874	0.540
Character error rate (CER)	0.584	0.617	0.114

recognition APIs. Even considering that the speech recognition model of this study is a model specialized in the voice of dysarthria patients, it is significant in that the speech recognition rate for them is very high (Table 1).

These results suggest that providing speech recognition APIs for people with dysarthria can improve speech recognition rates for these people, and further improve overall recognition rates in speech recognition technology.

4 Conclusion

In this study, an improvement measure of the voice recognition model for dysarthria was proposed. As the existing recognition model has been trained with training data made from the non-disabled people [4], it has a restriction to serving proper performance for people with dysarthria. To solve this problem, this work used the speech data of patient with dysarthria, instead of data from non-disabled people. As the result, it can achieve higher recognition performance than the existing one. If it could get more various sentences as training data, it is expected to show a more improved level of speech recognition rate.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation) in 2022” (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Biadsy F, Weiss RJ, Moreno PJ, Kanevsky D, Jia Y (2019) Parrotron: an end-to-end speech-to-speech conversion model and its applications to hearing-impaired speech and speech separation. arXiv
2. Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Zhu Z (2016) Deep speech 2: end-to-end speech recognition in english and mandarin. In: International conference on machine learning, PMLR, pp 173–182
3. Min S, Lee K, Lee D, Ryu D (2020) A study on quantitative evaluation method for STT engine accuracy based on Korean characteristics. *J Korea Acad-Indust cooperation Soc* 21(7):699–707
4. Jefferson M (2019) Usability of automatic speech recognition systems for individuals with speech disorders: past, present, future, and a proposed model. Retrieved from the University of Minnesota Digital Conservancy, pp 11–13

Korean Twitter Bot Detection via Deep Learning



Hyun Seop Yoon, Se Yeon Yang, Dong Hyuk Yoo, Myung Hak Kim,
Seo Yeon Im, Se Jong Oh, and Ill Chul Doo

Abstract We introduce a neural network model detecting Korean malicious bot account in Twitter. We extracted features from Korean malicious bot account in Twitter and implemented a model in the context of classification between human and bot account. This can be seen as an improvement of malicious bot detection, in an area of accuracy and multilingual aspect. The mode also suggests an end-to-end solution framework by offering web service.

Keywords Malicious bot · Twitter · Bot detection

H. S. Yoon

English Linguistics and Language Technology, Hankuk University of Foreign Studies, Seoul,
Korea

e-mail: xianxie31@hufs.ac.kr

S. Y. Yang

Linguistics and Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: eastdh@gmail.com

D. H. Yoo · M. H. Kim · S. Y. Im

Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin,
Korea

e-mail: se912912@gmail.com

S. Y. Im

e-mail: imsylive@gmail.com

S. J. Oh (✉) · I. C. Doo (✉)

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

1 Introduction

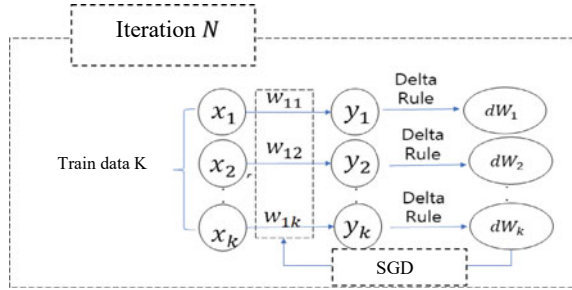
As machine learning technology advances, vicious bot violence is on a rise. The proportion of malicious bots increased from 2.1 to 22.7% in 2022 [1]. The Internet users being threatened by disinformation as malicious bot proportion increases. Filtering malicious bot is crucial for credible Internet environment. Current bot-detecting model: Botometer, lack language variability [2] and has low performance due to few detecting criteria. In this work, we introduce a neural network model detecting Korean malicious bot account in Twitter that attempt to rectify this problem.

Our model modified bot-detecting criteria to enhance bot-detection accuracy. The central idea is to build a Korean malicious bot-detecting criteria with adaptation of current criteria. While few criteria, like Spammer, Fake Follower, were adopted, most of criteria were revised to Tweet Time, Illegal Copy, and Retweet ratio. The model is trained to learn how to detect Korean Twitter malicious bot properly.

2 Methodology

In this study, we use only Korean tweets as datasets and establish new bot discrimination criteria. First, we collect a total of about 200 tweets per account from a total of 1000 accounts, including 500 randomly selected human accounts and 500 bot accounts, including spam accounts, to collect a total of 200,000 tweets. Then, we preprocess each of the tweets according to five new criteria consisting of Retweet Ratio, Tweet Time, Spammer, Fake Follower, and Illegal Copy. First, in the case of Retweet Ratio, we confirm that there is a retweet bot for distributing tweets containing specific topics [3]. Therefore, it is estimated that if the retweet rate of any account shows an extreme number, the probability of being a bot will be high. Retweet Ratio R is obtained by dividing the total number of retweets $n(r)$ of the account by the total number of tweets $n(t)$. In the case of Tweet Time, we confirm that bot accounts in account samples have a constant number of tweets posted per hour in common. Therefore, the Tweet Time T is obtained using the standard deviation of the number of tweets posted per hour. In the case of Spammer S , the average of the spam degree s_r of each tweet in the account is obtained to determine the spam account. To this end, we analyze the tweets of the collected spam accounts with the Korean morpheme analyzer 'kkma' to obtain the frequency f_m of each spam morpheme m , and s_r is the sum of f_m . In the case of Fake Follower, it uses 'followAudit' [4], a service that analyzes Twitter followers, to detect the proportion of fake accounts that are not self-authenticated or are not actually operated by humans. In the case of Illegal Copy, it is to detect an account that illegally copies posts from the community, not directly written articles. We confirm that tweets from these accounts have a feature that when they are searched on a portal site, posts with the same title as the tweets exist on multiple sites. Therefore, using Google search crawling, it is estimated that the higher the percentage of copyright infringement posts, the higher the probability

Fig. 1 DNN weight optimization using SGD method



that an account is a bot.

$$R = \frac{n(r)}{n(t)}, T = 1 - \frac{\sigma(n(t))}{h}, S = E(s_t), s_t = \sum f_m$$

Deep neural network (DNN) based on the universal approximation theorem distinguishes whether the input account is a bot or human. Aggregation of weighted criteria calculates bot-score of each input account. The classification is followed by bot-score computation.

300 training example was labeled. Train/test split is 7:3. Figure 1 gives the overall architecture of model updating weight with SGD method [5].

Adjustment of weight based on delta rules is retrieved with

$$W_{ij} \leftarrow W_{ij} + \alpha \phi(v_i) e_i x_j$$

where ϕ is a function that updates weights using input x_j per iteration.

3 Result

Table 1 gives the ‘Twitter malicious bot account spam-morpheme dataset. The model gives a significant result for detecting malicious bot, with the Spammer criterion.

The last row in Table 2 compares accuracy between the baseline model and weighted DNN model. DNN model outperforms baseline model on accuracy, by resulting 0.9652 accuracy [6]. Spammer criterion, which has 0.9364 weight, is the most influential criterion for detecting malicious bot while Illegal Copy has 0.015 weight.

Our model offers a web service which serves as a solution for detecting malicious bot. The model provides bot-score maximum of 100 by searching user account. A user-interface of the web service is shown in Figs. 2 and 3.

Table 1 Twitter malicious bot account spam-morpheme dataset and frequency

No	Spam words	Frequency	No	Spam words	Frequency
1	바독이 [padugi]	0.9188	11	클릭 [k ^h urɨk ^h ik ^h ʰ]	0.2157
2	성인게임 [s ^h ʌŋiŋgeim]	0.8731	12	토토 [t ^h o ^h t ^h o]	0.2129
3	사이트 [s ^h ait ^h u]	0.8193	13	이동 [idoŋ]	0.1999
4	안전 [andʌn]	0.6826	14	카지노 [k ^h adzino]	0.1955
5	추천 [ts ^h uts ^h ʌn]	0.4492	15	소개 [s ^h oɡe]	0.182
6	온라인 [onrain]	0.3251	16	인터넷 [in ^h ʌnet ^h ʰ]	0.1806
7	홀덤 [hold ^h em]	0.3035	17	바 [pa]	0.1719
8	섹트 [s ^h et ^h u]	0.2871	18	먹튀 [mʌk ^h t ^h wi]	0.1657
9	놀이터 [norit ^h ʌ]	0.2573	19	천안 [ts ^h ʌnan]	0.1381
10	주소 [tsus ^h ʌ]	0.2389	20	스타 [s ^h ʌst ^h a]	0.1291
No	Spam words	Frequency	No	Spam words	Frequency
21	스포츠토토 [s ^h ʌp ^h ʌts ^h ʌt ^h ʌ]	0.9188	26	텍사스 [t ^h ek ^h s ^h ʌs ^h ʌt ^h ʌ]	0.2157
22	강남 [kaŋnam]	0.8731	27	슬롯 [s ^h ʌlʌt ^h ʰ]	0.2129
23	환전보장 [hwandʌnbodʌzʌŋ]	0.8193	28	조기만남 [tsok ^h ʌnmamnam]	0.1999
24	포커 [p ^h ok ^h ʌ]	0.6826	29	바카라 [pak ^h ara]	0.1955
25	릴 [ril]	0.4492	30	네임드사다리 [neimds ^h adari]	0.182

Table 2 Accuracy comparison between baseline and DNN model

Method	Spammer	Illegal copy	Tweet time	Retweet ratio	Fake followers	Accuracy
Baseline	0.2	0.2	0.2	0.2	0.2	0.4153
DNN	0.9364	0.015	0.17	0.1105	0.0714	0.9652



Fig. 2 Human account: bot-score per criteria



Fig. 3 Bot account: bot-score per criteria

4 Conclusion

In this work, we presented a neural network model that classifies bot/human account in a Korea Twitter environment. Compared to current bot-detection model of [2], there is no multilingual aspect so our model can be used in a wider range of settings. Our model approaches the SOTA performance on bot detection and is significantly better than other baselines with the same level of dataset. Our approach proposed a concrete grounds for detecting malicious bot so that can be applied to filtering disinformation in other Internet domain and other task such as personal information security and so on.

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the Ministry of Science, ICT (MIST), Korea, under the National Program for Excellence in SW, supervised by the Institute of Information and communications Technology Planning and Evaluation (IITP) in 2022* (2019-0-01816).

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Munger K (2020) Don't @ ME: experimentally reducing partisan incivility on Twitter. *J Exp Polit Sci* 1–15
2. Yang KC, Ferrara E, Menczer F (2022) Botometer 101: social bot practicum for computational social scientists. *J Comput Soc Sc*. <https://doi.org/10.1007/s42001-022-00177-5>
3. Twitter bot creation. <https://www.labnol.org/internet/retweet-favorite-twitter-bot/28967/>
4. Followeraudit. <https://www.followeraudit.com/>
5. Cho JH, Kim KB (2005) Fuzzy single layer perceptron using dynamic adjustment of threshold. *Korea Open Access J*
6. Maryam H et al. (2016) An empirical study of machine learning algorithms for Social media bot detection, *IEEE*

Development of Literature Short-Form System Using Emotional Keyword



Sang Won Roh, Sung Cheol Kim, Tae Ho Kim, Jeong Woo Bae, Ji In Han, Se Jong Oh, and Ill Chul Doo

Abstract This research aims to develop a system for short-form formatting Korean-English literature through AI-based NLP technology. Using a deep learning model, we extracted the most semantically relevant core text from the full text of literature and constructed an emotional keyword classification model by training the existing pre-trained natural language processing model with a sentence-emotion pair Corpus DataSet. Next, we constructed a model that uses text-to-image technology to create images from the core texts and subjects. By combining the above models, database is built and distributed through the web, and finally, short-form literary content services are provided to users.

Keywords Short form · Emotional classification · Korean literature

S. W. Roh

English Linguistics and Language Technology, Hankuk University of Foreign Studies, Seoul, Korea

e-mail: sangwenyroh@gmail.com

S. C. Kim · J. W. Bae

Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: jude7794@naver.com

J. W. Bae

e-mail: wjddn0308@naver.com

T. H. Kim

Department of Statistics, University of Seoul, Seoul, Korea

e-mail: cosmos2123@naver.com

J. I. Han

Linguistics and Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: gkswldls8671@naver.com

S. J. Oh (✉) · I. C. Doo

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

1 Introduction

The growth of short-form content platforms is accelerating. In the modern society that pursues simplicity, starting with TikTok—a service that provides short videos within a minute—similar platforms such as YouTube Shorts have continued to be released. These platforms and contents have a great influence on the overall modern culture. However, due to their own problem of attracting viewers’ attention in a short time, the short-form platform continues to produce stimulating and meaningless content. In this paper, we try to integrate literature, a collection of cultures, humanities, and sensitivity, into the short-form format. Most of the prior studies used collaborative filtering [1] and recommendation systems based on keyword characteristics [2]. No analysis of Core Text Extraction research has not been conducted. Also, studies focused on sentimental classification: classifying into two categories, positive and negative [3], emotional classification that classifying into more than three emotional keywords is insufficient. Therefore, this paper aims to provide short-form content, consisting of key paragraphs (Core Text) and images of literature that match the user’s emotional keyword, after constructing database by receiving literature data as an input value.

2 Methodology

See Fig. 1.

2.1 Raw Data and Data Preprocessing

In this study, Korean literature raw data including Korean old songs, Korean modern poetry, and Korean modern novels were constructed using [4, 5], and English literature raw data was constructed based on [6]. Each Row of Raw Data consists of five columns: [“Type”, “Title”, “Writer”, “Full Text”, and “Topic”]. Full Text was cut into stanzas and paragraphs. If the subject includes Chinese characters, it was deleted in case the model cannot interpret it. An emotional object such as “Life”, “Self”, and

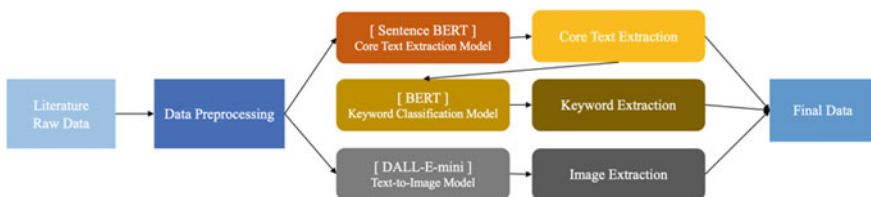


Fig. 1 System configuration and database deployment process

“Nature” was extracted from Topic and stored in the column name “Subject”. Cases in which two or more conflicting emotions are revealed at the same time were selected and classified separately as “Ambivalence” in the column name “Keyword”. Because the text-image model DALL-E Mini interprets only English text, Korean topics are translated into English using the Naver Papago translation model.

2.2 Core Text Extraction Model

An algorithm was established to extract the most important core text from full text of each literature (classical poetry, modern poetry, modern fiction, and English poetry). First, the word “core” was defined as “meaningfully related to the topic”. And sentence BERT (hereinafter referred to as S-BERT) [7] was used to focus on the meaning of the text. Through this, an information search system that focuses on meaning rather than conventional keyword matching is implemented (Semantic Search). In the algorithm, raw data’s column was entered as Query and Full Text as Corpus, and the core text output was stored in the column named “Core Text”.

2.3 Emotion Classification Model

A classification model was designed to extract emotional keywords from literature. First, the emotion keyword was classified into a total of seven emotions by adding six Korean [joy, anger, sadness, anxiety, wound, embarrassment] and English [joy, danger, love, fear, sadness, surprise] and “Ambivalence” in consideration of the characteristics of literature. We designed the model based on BERT and KoBERT [8] and selected sentence-emotional keyword pairs, CARER dataset (English, 416,809 pairs) [9], and AI hub emotional conversation corpus dataset (Korean, about 12,000 pairs) [10], emotional and speech style-specific speech synthesis dataset (Korean, about 15,000 pairs) [10]. To improve model accuracy, Korean language training data was augmented 10 times by utilizing KorEDA’s random deletion (RD) and random swap (RS) [11] used for learning. The topic of the literature RawData was input to the model to receive a keyword, which was stored in the column name “Keyword”.

2.4 Imaging Algorithm

For imaging, the DALL-E Mini model [12], which is a text-toimage model, was used. Among the images produced by inputting Data’s Core Text and Topic, two images were selected for each work, excluding inappropriate images, and stored in database.

Table 1 Example result data

Type	Title	Writer	Full Text	Core Text	Keyword
Englishpoem	La Belle Dame Sans Merci (Original Version)	John Keats	Oh what can ail thee, knight-at-arms, Alone and palely loitering? The sedge has withered from the lake, And no birds sing... (중략)	I made a garland for her head, And bracelets too, and fragrant zone; She looked at me as she did love, And made sweet moan	joy
Korean poem	쉽게 N여진 시	윤동주	창밖에 밤비가 속 살거려 육첩방(六 疊房)은 남의 나 라... (중략)	시인이란 슬픈 천 명(天命)인 줄 알 면서도 한 줄 시 를 적어 볼까	sadness

3 Result

First, in the process of building an emotion classification model, a KoBERT model with better performance between KoBERT (accuracy 65%) and KoGPT2 (accuracy 60%) was adopted. The BERT model with relatively large pre-learning data showed a high accuracy of 92%, while the KoBERT model showed a low accuracy of 65% because the data were relatively small. To compensate for this, we combined [10] datasets and increased the accuracy to 96% by increasing the training data tenfold through data augmentation techniques (Table 1).

In addition, three experts (Two Korean language teachers, an undergraduate majoring in Korean Literature) were consulted to compare the derived dataset and extracted core text results to establish an S-BERT model verification index. The comparison dataset showed a high consistency rate of 81.9%. Therefore, the S-BERT used in this paper can be evaluated as a suitable model.

4 Conclusion

This study began with the idea of recommending literature to users through emotion classification and keyword extraction. To this end, S-BERT was used to extract core texts that were most related to the main theme of literature. In addition, the emotional keywords of literature were extracted through emotional classification. A database was built using extracted emotional keywords, core texts, and images that match the theme of literature, and based on this, a web that can provide customized literature to users was built. In addition, it is expected to provide higher-quality content in the future by solving copyright problems, collaborating with publishers, and expanding emotional keywords (Fig. 2).



Fig. 2 Execution screen

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) in 2022” (2019-0-01816)

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Park JH, Kim MK, Kim MJ, Park (2015) A movie recommendation system based on collaborative filtering and review. In: Proceedings of the Korea information processing society conference, pp 294–296
2. Cho SE, Lim HS (2019) A study on product recommendation system based on user search keyword, J Digital Contents Soc 20(2):315–320
3. Choi DE, Kim HM, Lee HR, Hwang YR (2022) Sentimental analysis of Youtube Korean subscripsts using KoBERT. J Korean Soc Inf Process 29(1):513–516
4. Woorimal net. <http://www.woorimal.net/>. Last accessed 19 Sept 2022
5. Gong-u madang. <https://gongu.copyright.or.kr/gongu/main.do/>. Last accessed 19 Sept 2022
6. Poem analysis. <https://poemanalysis.com/>. Last accessed 19 Sept 2022
7. Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using siamese BERT-networks., In: Proceedings of empirical methods in natural language processing, pp 3980–3990
8. Lee UYJ, Choi HJ (2020) Joint learning-based KoBERT for emotion recognition in Korean. J Korean Inst Inf Sci Eng 47(2):568–570
9. Github. https://github.com/dair-ai/emotion_dataset. Last accessed 19 Sept 2022
10. AI-hub. <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100>. Last accessed 19 Sept 2022
11. Wei J, Zou K (2019) Eda: easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint [arXiv:1901.11196](https://arxiv.org/abs/1901.11196)
12. Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, Chen M, Sutskever I (2021) Zero-Shot text-to-image generation. In: Proceedings of the 38th international conference on machine learning, PMLR 139, pp 8821–8831

English Pronunciation Correction Service for Hearing-Impaired People: BETTer, Focusing on the Personalized Speech Model



Seon Hong Park, Hyun Jin Park, Yoon Jeong Kim, Young Woong Park, Jeong Min Lee, Ye Rin Jung, Seok Ryu Jo, Se Jong Oh, and Ill Chul Doo

Abstract There has been a growing body of research that explores the need for hearing impaired people. However, English Education aids for the hearing-impaired are still lacking. This paper introduces an English pronunciation correction service for the hearing-impaired to help ensure the equal right to education for members of society. The purpose of this research is to allow effective English pronunciation correction by providing user-friendly feedback using visual information such as

S. H. Park

Department of International Economics and Law, Hankuk University of Foreign Studies, Seoul, Korea

e-mail: psh0855@hufs.ac.kr

H. J. Park

Department of Czech and Slovak Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: hyunjinieyo@hufs.ac.kr

Y. J. Kim

Department of Greek Studies and Bulgarian Studies, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: yj9598@hufs.ac.kr

Y. W. Park

TESOL & English Linguistics, Hankuk University of Foreign Studies, Yongin, Korea

J. M. Lee

Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin, Korea

Y. R. Jung

Department of Information Communications Engineering, Hankuk University of Foreign Studies, Yongin, Korea

S. R. Jo

Linguistics & Cognitive Science, Hankuk University of Foreign Studies, Yongin, Korea

S. J. Oh (✉) · I. C. Doo

Artificial Intelligence Education, Hankuk University of Foreign Studies, Yongin, Korea

e-mail: tbells@hufs.ac.kr

I. C. Doo

e-mail: dic@hufs.ac.kr

graphs, images, and personalized corrective voice considering the special needs of hearing-impaired people.

Keywords Computer-Assisted Pronunciation Training (CAPT) · COURSEWARE · Speech generation model

1 Introduction

In Korea, the English pronunciation correction service has been diversified as the need for English skills expands. However, English pronunciation correction service for the hearing-impaired people is not fully developed. As the survey conducted by Gun-min Lee shows, 82.6% of teachers and 46.5% of hearing-impaired replied that English education is necessary for hearing-impaired [1]. An English pronunciation-correcting mobile app for hearing impaired was developed by Young-Joo Lee in 2015, yet there was a limit to giving only phonemic feedback [2]. Therefore, BETTer is designed to give sentence-level feedback. Moreover, it makes learners enthusiastic by providing a personalized voice. We also automatize the creating process of question-answering tasks by using BART [3], and use generalized end-to-end [4], Tacotron [5], and WaveRNN [6] to generate the personalized voice, and use wav2vec 2.0 [7] to produce feedback results.

2 Methodology

2.1 *Preparing Content Data*

We use the fairy tale dataset from the Library of Congress for English speech [8]. By applying BART, pairs of questions and answers are created automatically based on contexts extracted from the fairy tale. BART is a BERT-based text generation model that combines a bidirectional encoder (BERT) and auto-regressive decoder (GPT) [3]. A user speaks the answer to the automatically generated question.

2.2 *Providing Personalized Answer Speech*

This study adopted real-time voice cloning to provide a personalized answer voice [9]. This technology consists of three stages of the pipeline [9]. First, we extract features representing speaker characteristics from the speech through Generalized End-to-End (GE2E) [4]. Then mel spectrogram is generated using Tacotron for these

extracted features [5]. Tacotron is a recurrent sequence-to-sequence feature prediction network with attention that predicts a sequence of mel spectrogram frames from an input character sequence [5]. At the initiatory study, we built a pipeline consisting of Tacotron and WaveGlow [10]. The voice generated by this pipeline has the disadvantage of noise and reverb. Therefore, we choose real-time voice cloning using WaveRNN, which could generate a clear voice.

2.3 Evaluation System and Feedback

In the case of the evaluation system, we use the Wav2Vec2ForCTC model, which is based on wav2vec 2.0, to apply automatic speech recognition (ASR). This is to convert the user's recorded voice into text. In order to give feedback on a phonemized basis regarding spoken sentences, we used an eSpeak-based Python library, "phonemizer" to convert a transcribed text and a correct answer text into phonemes. We scored the correct and converted phoneme strings by measuring the similarity with the correct answer using the longest common subsequence algorithm.

How BETTer works to give feedback is as follows. After the user speaks the answer, BETTer provides the answer text, the user utterances text, and the correct pronunciation symbol. It also provides pronunciation-based feedback using video and personalized answer voice based on real-time voice cloning. At this time, the video is phonemic-unit-based voice videos from BBC English [11], which aims to improve user pronunciation through a native speaker's phoneme voice. It shows the graph using root mean square (RMS) amplitude value to compare the user's voice and the personalized answer voice.

3 Result

The results take the form of web services. Together, these results provide crucial insight into the effective way to correct pronunciation. The sequence of performance is as follows. A user may select levels 1–3. As seen in Fig. 1, the user speaks the answer inferring the blank according to the given text.

Figure 2 can be seen after the user finishes recording. As Fig. 2 shows, there is correct text converted by the ASR model. The similarity of pronunciation compared with the correct voice is expressed as a four-step score (Perfect, Great, Good, Try again). It gives feedback, as shown in Fig. 2. First, stress (^ mark) could help to show where to add stress. A personalized answer voice is played with a sound source play button. The mispronounced phoneme is provided with an activated red square button through a video. A graph comparing the personalized correct answer voice with the user's recorded voice provides intonation feedback.

Lv1. Little Snow-White

Snow-White grew up and became ever more beautiful. When she was seven years old she was as beautiful as the light of day, even more beautiful than the queen herself.

QUESTION

✓ Who grew up and became more beautiful than the queen?

_____ grew up and became ever more beautiful.

Hint Snow-White

Rec Stop

Fig. 1 BETTer service screen (a recording process)

4 Conclusion

This study attempts to establish hearing-impaired people’s equality of educational rights and relieve their learning disadvantage causing individual differences between hearing-impaired and others. The BETTer service is expected to alleviate economic effects on language rehabilitation costs [12]. In addition, it gives social benefits for not only hearing impaired but bilinguals as a language learning tool. We expect the hearing-impaired could demonstrate their potential linguistic talents and become rightful members of the global society.

REPEAT AFTER ME! FEEDBACK

Great
80%

Answer: Snow-White grew up and became ever more beautiful.
Yours: snowwhite grew up an oimam edward wore beautiful

발음기호: sn^ouw^ast gɹ^u: , ʌp ænd bɪk eɪz m^evə m^o:ɹ bj^u: rɪfəl

0:00 / 0:04

Answer
Yours

Fig. 2 BETTer service screen (evaluation and feedback)

Acknowledgements This research was supported by Hankuk University of Foreign Studies Research Fund (of 2020).

This research was supported by the MIST (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) in 2022” (2019-0-01816)

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8065934).

References

1. Lee GM, Kim DO, Lee SY, Park HJ, Choi BG, Heo SY (2015) Comparable study on learning English between students and teachers for the development of English learning smart application for students with hearing impaired. *J Rehabil Res* 19(1):345–365
2. Lee YJ, Lim SM, Choi YJ, Moon BH (2015) A mobile App(See&Speech) of correcting pronunciation for hearing-impaired persons. *J Korean Assoc Comput Educ* 18(4):11–18
3. Lewis M (2020) BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: 58th annual meeting of the association for computational linguistics, pp 7871–7880
4. Wan L, Wang Q, Papir A, Moreno IL (2018) Generalized end-to-end loss for speaker verification. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 4879–4883
5. Wang Y et al (2017) Tacotron: towards end-to-end speech synthesis. arXiv preprint [arXiv:1703.10135](https://arxiv.org/abs/1703.10135)
6. Kalchbrenner N (2018) Efficient neural audio synthesis. In: 2018 International conference on machine learning, pp 2410–2419
7. Baevski A (2020) wav2vec 2.0: a framework for self-supervised learning of speech representations. *Neural Inf Process Syst* 2020:12449–12460
8. Library of Congress. <https://www.loc.gov/free-to-use/classic-childrens-books/>. Last accessed 30 Sept 2022
9. Jemine C (2019) Real-time-voice-cloning. University of Liège, Liège, Belgium
10. Prenger R, Valle R, Catanzaro B (2019) Waveglow: a flow-based generative network for speech synthesis. In: 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 3617–3621
11. BBC Learning English. <https://www.bbc.co.uk/worldservice/learningenglish/grammar/pron/sounds/>. Last accessed 30 Sept 2022
12. Gwon SJ, Lee GM, Cho YT, Kim GY, Gong JY (2009) A study on support plans for improving the quality of life of disabled children and adolescents I: rehabilitation and welfare service improvement plan. In: National Youth Policy Institute 09-R18–5, pp 3–170

Improving the Performance of Judicial Precedent Search by Fine-Tuning S-BERT



Gilsik Park and Juntae Kim

Abstract Legal search has been studied by legal experts who possess specialized knowledge, but recently, various researches are being conducted to allow even nonprofessionals to search for law cases. However, the general public who wants to use the legal search service has difficulty searching for relevant precedents due to a lack of understanding of legal terms and structures. In addition, the existing keyword and text mining-based legal search methods have their limits in yielding quality search results for two reasons: they lack information on the context of the judgment, and they fail to discern homonyms and polysemies. As a result, the accuracy of the legal document search results is often unsatisfactory or skeptical. This paper aims to improve the efficacy of the general public's legal search in the Supreme Court precedent and Legal Aid Counseling case database. To this end, we propose a legal document search method that uses the sentence-BERT model. The sentence-BERT model embeds contextual information on precedents and counseling data, which better preserves the integrity of relevant meaning in phrases or sentences. Our initial research has shown that the Sentence-BERT search method yields higher accuracy than the Doc2Vec or TF-IDF search methods.

Keywords Legal service · Machine learning · Deep learning · BERT · Data mining

G. Park (✉) · J. Kim
Department of Computer Engineering, Dongguk University, Jung-gu, Seoul, South Korea
e-mail: gspark@mail.kw.ac.kr

J. Kim
e-mail: jkim@dongguk.edu

1 Introduction

1.1 A Subsection Sample

Recently, legal-related tasks have been implemented in various forms, including understanding legal documents using artificial intelligence, searching documents, predicting judgments, classifying legal documents, and summarizing [1]. These tasks are expected to help legal professionals in the legal field, such as minimizing complex and repetitive tasks and improving the accuracy of judgments. In addition, there is a growing number of personal litigation in which ordinary people, not law-related experts, directly file lawsuits. It is reported that about 60–70% of civil cases underway these days correspond to personal litigation [2]. However, since there is no help from legal experts in the litigation process alone, it is necessary to accurately understand and interpret legal documents, refer to relevant precedents, laws, and judgments. In this process, difficult judgment terms such as lack of understanding of the exact meaning of legal terms and tricky trial terms make litigation alone more difficult.

Accordingly, deep learning-based legal document search has recently been studied. However, deep learning-based models have the disadvantage of requiring a lot of computation to search for large documents such as legal documents because they must perform computations every time a variety of natural questions are input from the general public. To overcome these limitations, embedding models have been studied, in which a large number of legal documents are represented as vectors in a vector space, then a public legal query is input, vectorized in an independent vector space, and similarity between two vectors is obtained to search for legal documents most similar to legal queries [3–5]. The embedding method performs embedding of legal documents in advance, so the speed is fast, but there is a disadvantage that important information held by documents can be lost during the embedding process. To solve the above problems, the legal search method proposed in this paper is a new legal document search model using S-BERT.

2 Related Work

2.1 Rule-Based Legal Search System

Legal search methods using artificial intelligence technology have been studied since the 1970s. In the early days, automatic calculation systems based on artificial intelligence reasoning became the mainstream [3, 4]. Since the rule-based system was designed based on the knowledge of legal experts in advance, there was no choice but to have considerable limitations in outputting all the appropriate answers to the users' various queries.

2.2 Legal Search System Based on Text Mining

There are methods to use term frequency (TF) and inverse document frequency (IDF) as text mining techniques, which are most commonly used as keyword-based searches when searching legal documents. TF-IDF is a value obtained by multiplying the reciprocals of TF and IDF by each other, and the larger the value, the more important a specific word is used in the document (see Eq. 1).

$$tfidf_{i,j} = tf_{i,j} * \log\left(\frac{N}{df_i}\right) \quad (1)$$

where:

$tf_{i,j}$ = total number of occurrences of i in j

df_i = total number of documents containing i

N = total number of documents

TF-IDF-based search cannot perform a search when a search term does not exist in a document. In addition, keyword-based search can be said to be a method that is vulnerable to synonyms and polytonics.

2.3 Embedding-Based Legal Search System

For text analysis, in the field of natural language processing, words or documents are converted into numbers so that they can be calculated. This work is called embedding. The number of words or documents expressed varies depending on the algorithm being embedded.

2.3.1 Word2Vec and Doc2Vec

Word-based embedding models such as Word2Vec are meaningful in that words are represented as vectors, and then relationship inference and similarity calculations between word vectors are possible [6]. Word-based embedding models are being used in various fields, and in particular, they are being actively conducted in document classification and emotional analysis. Doc2Vec has brought high performance in various tasks related to natural language processing, such as emotion classification and document classification [7]. Doc2Vec is an algorithm that allows Word2Vec to calculate the distance between vectors embedded in sentence or document units to obtain the similarity between documents (Figs. 1 and 2).

Fig. 1 Word2Vec

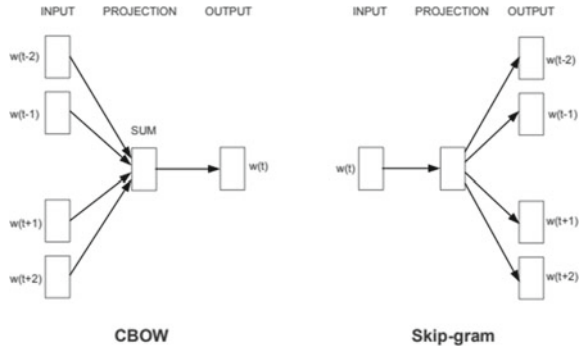
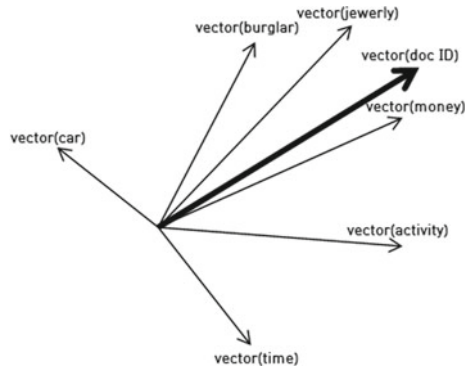


Fig. 2 Doc2Vec



2.3.2 BERT and S-BERT

To overcome the shortcomings of keyword search, the Bidirectional Encoder Presentations from Transformer (BERT) model released by Google is a model generated with large-capacity Wikipedia data for 104 languages around the world [8]. BERT generates a pre-trained model based on unlabeled data, and then modifies the weight as it progresses further on various tasks with labels (see Figs. 3 and 4).

S-BERT is a model that improves the performance of sentence embeddings in the BERT model and is a modeling technique that applies sentence embeddings from BERT to fine-tune BERT [9]. S-BERT inputs Sentence A and Sentence B into each BERT, performs pooling, and then obtains the sentence embedding vectors u and v for each (see Fig. 5). The cosine similarity between the two vectors u and v is obtained and then learned by minimizing the mean squared error with the label similarity.

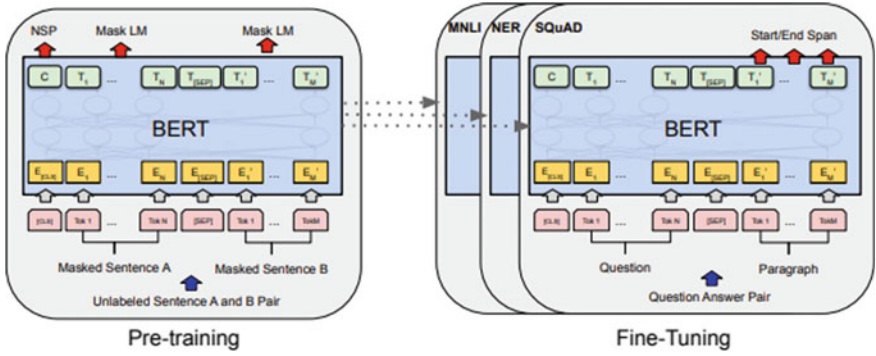


Fig. 3 BERT learning method

Fig. 4 S-BERT

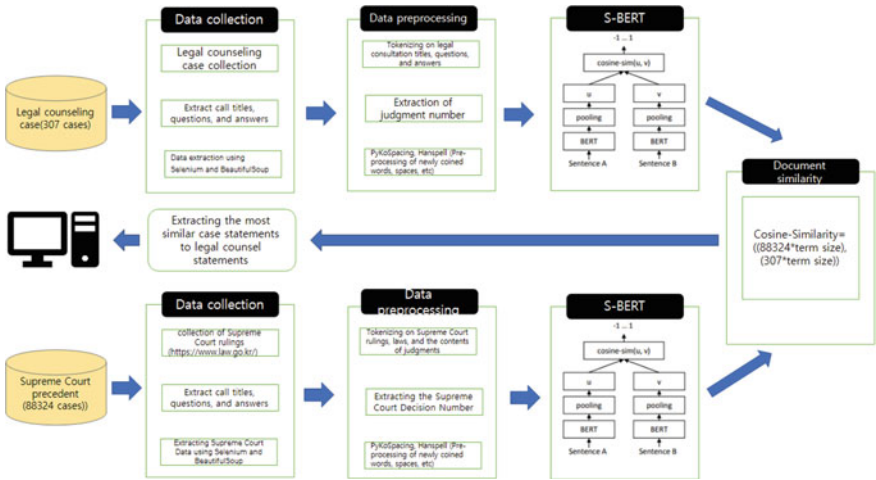
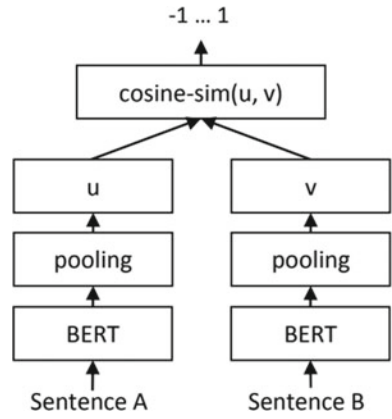


Fig. 5 Legal search system process (S-BERT)

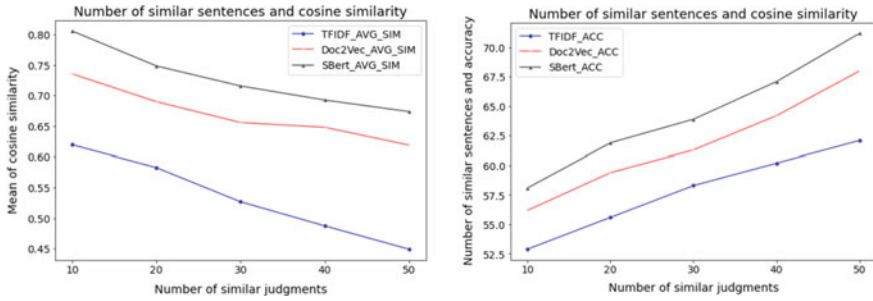


Fig. 6 Average of cosine similarity and accuracy

3 S-BERT-Based Legal Search

3.1 Legal Search System Process

The process of the legal search system to be proposed in this paper is as follows (see Fig. 6).

First, the legal counseling data is scraped on the Korea Legal Aid Corporation site and then documented. In addition, case law data on the Supreme Court's comprehensive legal information site is scraped and made into documents. Second, the spacing is checked and corrected using the PyKoSpacing package, etc., and the typo is corrected using the Hanspell package. Third, for the legal data set that performed data preprocessing, the contents of the judgment, judgment number, legal counsel questions, titles, and answers are extracted, and talk aging is performed. Fourth, we generate TF-IDF matrix (DTM) and embed it into BERT and S-BERT models for legal counseling and Supreme Court rulings that performed torque aging, respectively. Fifth, we obtain cosine similarity for the legal counseling question vector and the Supreme Court judgment vector expressed in the S-BERT embedding vector space for the user legal counsel question and extract the Supreme Court judgment with the highest similarity.

Finally, the judgment number and case law information are provided to the user.

3.2 S-BERT Embedding Vector and Similarity

In this paper, we used a model that fine-tuned the Korean pre-trained model ko-sroberta-multitask with the KorNLU dataset to perform S-BERT embedding [10]. First, embeddings were accepted for the Legal Aid Corporation's legal counsel case data and the Supreme Court rulings collected through web scraping. In this process, all sentences and paragraphs of legal questions and Supreme Court rulings were embedded with 768-dimensional dense vectors, respectively, and cosine similarity

Table 1 Accuracy and cosine similarity by model

Model	N					
TF-IDF	Accuracy	52.9	55.6	58.3	60.2	62.1
	Avg similarity	0.6193	0.5825	0.5268	0.4874	0.4492
Doc2Vec	Accuracy	56.2	59.4	61.3	64.2	68.0
	Avg similarity	0.7352	0.6897	0.6556	0.6478	0.6187
S-BERT	Accuracy	58.1	61.9	63.9	67.1	71.2
	Avg similarity	0.8053	0.7482	0.7154	0.6924	0.6735

between the embedded user question vector and Supreme Court judgment vector was calculated to extract the most similar judgment. From the results searched in this way, the title, question, and answer of the legal counseling with the highest similarity were extracted, and then the reference judgment number was scraped. The reference judgment number can be searched from the Supreme Court case data and provide users with various information related to the judgment, such as “judgment matters” and “summary of judgment” on the case.

4 Experiment Result

Looking at Table 1, it is possible to guess how each embedding model determines the similarity between sentences. First, the TF-IDF model method of calculating similarity based on keywords seems to judge similar sentences by giving high weights for specific words. In the case of search using cosine similarity between legal counseling questionnaire embedded in the S-BERT model and Supreme Court judgment vector, higher accuracy can be derived compared to other algorithms. This can be said to be the result of reflecting the contextual information on the questionnaire and judgment.

The various model-specific searches experimented in this paper were treated as incorrect answers, but it is expected that there will be several Supreme Court rulings similar to the correct answers. This is judged to be a factor that makes the accuracy of search results lower. In this situation, it can be seen that the accuracy of the legal search results using S-BERT is higher than that of other algorithms.

5 Conclusions

In this paper, we propose a method to search for the most similar Supreme Court rulings using S-BERT to search for accurate precedents and judgment information in terms used on a daily basis by ordinary people who do not know legal terms. In addition, even in specialized fields such as legal fields, we were able to build datasets that could be used for learning through preprocessing, and we were able to derive

sufficient performance to be used for real legal services using the S-BERT model. In the future, research will be conducted to derive meaningful results for various cases of legal questions of the general public that have not been verified in this study. In addition, research will be continuously conducted to improve the search accuracy of judgment by making the relationship between sentences clearer.

References

1. Zhong H et al (2020) How does NLP benefit legal system: a summary of legal artificial intelligence. In: Proceedings of the 58th annual meeting of the association for computational linguistics, pp 5218–5230
2. Gangwon Ilbo (2022) <http://www.kwnews.co.kr/page/view/2022031500000000016>. Last accessed 15 Mar 2022
3. Rissland EL et al (2003) AI and law: a fruitful synergy. *Artif Intell* 150:1–15
4. Buchanan BG et al (1970) Some speculation about artificial intelligence and legal reasoning. *Stanford Law Rev* 23:40–62
5. Le Q et al (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp 1188–1196
6. Mikolov T et al (2013) Efficient estimation of word representations in vector space. In: International conference on learning representations
7. Zhang H et al (2019) Similarity judgment of civil aviation regulations based on Doc2Vec deep learning algorithm. In: Proceedings of 12th international congress on image and signal processing, BioMedical engineering and informatics
8. Devlin J et al (2019) Bert: pre-training of deep bidirectional transformers for language understanding. NIPS
9. Reimers N et al (2019) Sentence-BERT sentence embeddings using siamese BERT-Networks. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, pp 3982–3992
10. Ham J et al (2020) Kornli and korsts: new benchmark datasets for korean natural language understanding. In: Proceedings article published 2020 in findings of the association for computational linguistics, pp 422–430

Machine Learning-Enabled Distributed Framework for Attack Detection in Social Networks



Sangthong Yotxay, Abir E. L. Azzaoui, and Jong Hyuk Park

Abstract With rapidly evolving technology, social networks are the most popular medium for communicating information from person to person on the Internet. Nowadays, people of all ages spend most of their time on social networking. As a result, vast amounts of information are being generated and exchanged through social networks worldwide. Moreover, the information shared through social networks and media spreads rapidly, nearly instantly, making it appealing to attackers to damage the transmission. Therefore, the privacy and security of social networks must be investigated from multiple perspectives, including security, privacy, and authenticity risks associated with the user's information shared whenever the user publishes personal data such as images, videos, audio, and more. Therefore, security and privacy are the major issues in social networks. To solve these issues, we propose machine learning-enabled distributed framework for attack detection in social networks. Extreme learning machine (ELM) algorithm is used at the edge layer with the classifier and train model for attack detection and communication latency in the networks. Furthermore, distributed blockchain is leveraged at the fog layer for data verification and validation, and then data is stored at the cloud layer. Moreover, we illustrate a methodological flowchart of the proposed framework.

Keywords Machine learning · Blockchain · Social network · Security and Privacy

S. Yotxay · A. E. L. Azzaoui · J. H. Park (✉)
Department of Computer Science and Engineering, Seoul National University of Science and Technology (SeoulTech), Nowon-gu, Seoul, South Korea
e-mail: jhpark1@seoultech.ac.kr

S. Yotxay
e-mail: yotxaysangthong@seoultech.ac.kr

A. E. L. Azzaoui
e-mail: abir.el@seoultech.ac.kr

1 Introduction

In a few years, social networks are speedily expanding with social media applications, including WhatsApp, Facebook, Instagram, YouTube, Twitter, LinkedIn, Google+, and many more that are utilized for individual and professional users. These applications have various advantages for everyone worldwide because people are easily connected and speak, exchange information, hobbies, and experiences find information resources, and more [1]. Moreover, social networks can be used for access to many targets such as entertainment, education, communication, business, and so on. Nevertheless, there are benefits to visiting these sites and some risks. For example, as a result of the high number of users, private information supplied on these networks is up-to-date, and users are asked to provide personal information such as name, date of birth, activities, gender, residence, education, and other sensitive information on social network sites. Therefore, the malicious user can easily access this data from the website because it is not privately stored. For example, every social media can share pictures, videos, Livestream, and more. Also available are “inbox features,” where users can post messages to and communicate with their friends. YouTube has even contained this function as well. However, it gives the attacker an excellent chance of phishing his targets. For example, the message might include a shortened URL that takes the recipient to a malicious site. Because it is difficult to determine whether a shortened URL is valid, attackers take advantage of this and disguise their harmful data in shorter URLs. Otherwise, there are several challenges in social networks, such as privacy, security, big data, low latency, and many more [2].

Many authors proposed solutions to address attack detection, security, privacy, and communication latency issues. For example, Siddique et al. [3] proposed a spam email detection model based on machine learning and deep learning algorithms. These algorithms generate a dataset with spam and ham emails which are preprocessed for several techniques. Evaluation is based on performance, accuracy, precision, recall, F-measure, and ROC-AUC. Prabhu et al. [4] investigated secure data acquisition methods for fake accounts detection in mobile communication and online social networks, dependent on machine learning technology, and provided detection for spammers, a hybrid gathering of spam communications on social networking sites. SVM, ANN, and RF algorithms are utilized for fake accounts detection and have hybrid features such as user-based and content-based features. The proposed framework performance is better than this method based on recall, accuracy, and F-measure.

Miao et al. [5] presented cyberattacks targeting controlled information methods that are based on machine learning. This method examines the attack depending on three major points: (1) controlled user behaviors, (2) controlled ML model-related information, and (3) controlled authentication information. Zheng et al. [6] proposed an ELM-based spammer detection technique for social network platforms. An ELM-based classification technique discriminates and retrieves content sets and behavior characteristics. Our proposed method has been demonstrated to be viable, efficient, and much more reliable than previous SVM-based models through a series of tests and

evaluations. Patial et al. [7] introduced ELM technique is utilized, which identifies the malicious links on the website and displays them on the page using the HTML CSS attribute. Only one form of harmful link has been found, namely, the phishing link on the site. As a result, the phishing dataset is employed. The ELM model is compared against the SVM model in terms of performance measures. The ELM model requires less training time than the SVM model. Nayak et al. [8] proposed a novel IoMT framework with Bayesian optimization and ELM hybridization to provide security in IoMT devices and privacy for patient data. Compared to similar state-of-the-art approaches, the proposed model achieves promising performances with improved accuracy in decision-making.

However, the existing research studies did not address various problems completely, such as attack detection, storage, identification, centralization, security, privacy, etc. To solve these challenges, we propose a machine learning-enabled distributed framework for attack detection in social networks.

Below mention is the main contribution of this research:

- Propose machine learning-enabled distributed framework for attack detection in social networks.
- ELM algorithm is used at the edge layer to mitigate the impact requirements of high accuracy, stability, processing data, and efficient performance in the social network in which the classifier algorithm is updated training model and then transfers data for the evaluation model.
- Distributed blockchain is used at the fog layer to solve the issues of data verification and validation, and then data is stored at the cloud layer.

2 Proposed Framework

This section discusses the proposed machine learning-enabled distributed framework for attack detection in social networks and addresses attack detection, data identification, centralization, security, privacy, and others.

2.1 Design Overview Proposed Framework

The overview of the proposed machine learning-enabled distributed framework for attack detection in social networks is shown in Fig. 1. It consists of four layers: connection layer, edge layer, fog layer, and cloud layer. In social media networks, we categorized them into four groups, including entertainment, communication, business, and education. These groups are connected to several IoT and sensor devices: smartphones, laptops, personal computers, wearable devices, smart television, and many more.

First, social media devices under the connection layer are used for recorded data and acquisition from IoT and sensor devices. As a result of increasing data, accuracy,

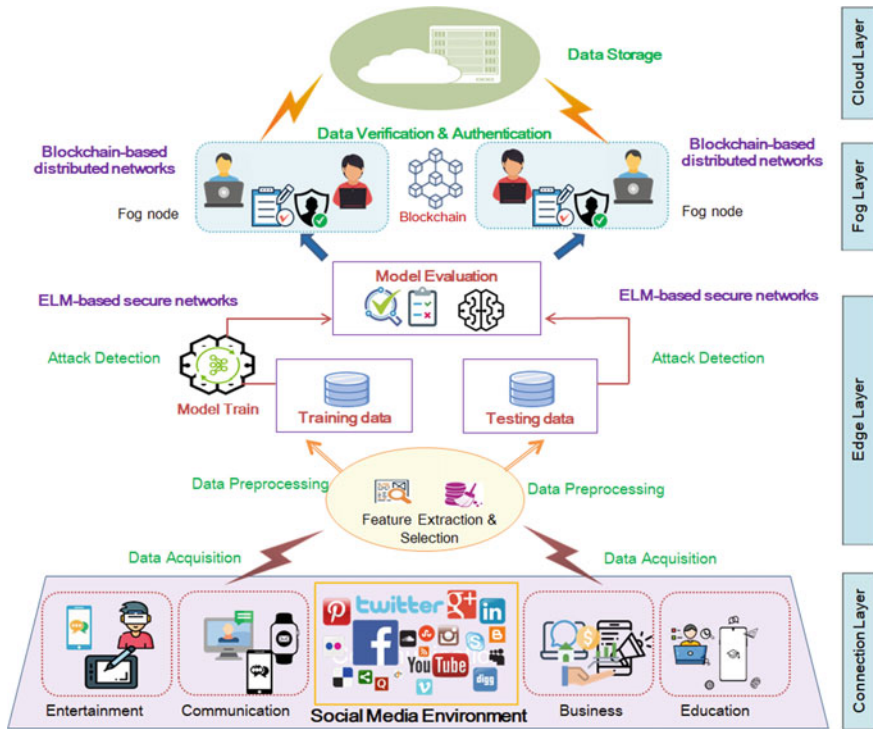


Fig. 1 Overview machine learning-enabled distributed framework for social networks

latency, and confidence become an issue for data processing and analysis. Therefore, to mitigate these challenges, we used ELM-based secure networks at the edge layer, providing high accuracy, stability, processing data, and efficient performance in the network using the classifier algorithm with an updated training model and evaluation model. Second, blockchain networks are utilized at the fog layer for data verification and validation with the Proof of Authority (PoA) consensus algorithm. After the validation and verification, data is transferred and stored in the cloud layer.

2.2 Methodological Flowchart of the Proposed Framework

The methodology flowchart of the proposed framework for social networks is shown in Fig. 2. Firstly, collecting data from devices such as smartphones, laptops, personal computers, wearable devices, smart television, and others to generate data to the upper layer. Secondly, the edge layer consists of two parts: ELM functionality and blockchain functionality. The ELM functionality includes feature extraction and feature selection methodology. Feature selection and extraction are used for processing data with data computation and data balancing. Network traffic and the

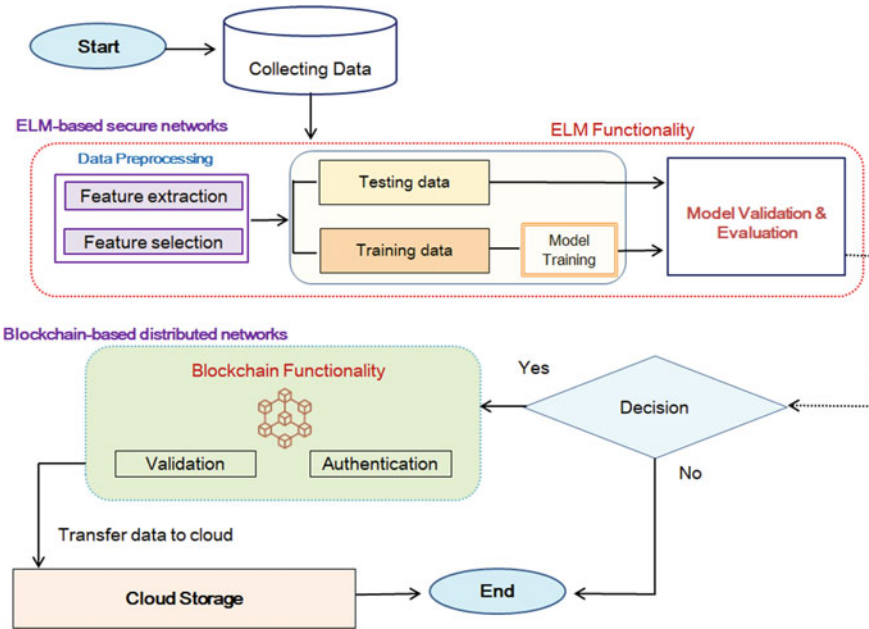


Fig. 2 Methodological flowchart of proposed framework

classification algorithm with the updated training model, testing model, and evaluation model performed high accuracy, stability, processing data, and efficient performance in the network. After that, data is transferred to the fog layer; this layer uses blockchain-based distributed networks consisting of data authentication and verification, which help security, privacy, scalability, and efficient data analysis. Finally, data transfer and stored in the cloud layer.

3 Conclusion

Social networks have become a component of our daily lives. Most online users spend more time on social networks communicating with others by sharing stories, photos, videos, and others. At the same time, social networks have a dark side and can put the user at a severe security risk. In this paper, we have proposed a machine learning-enabled distributed framework for attack detection in social networks. It relies on an extreme learning machine (ELM) algorithm employed at the edge layer to contain the data classification and train model for effective attack detection, high accuracy, stability, processing data, and efficient network performance. Furthermore, blockchain technology is distributed in the proposed framework with decentralized data analysis for security, privacy, scalability, and efficient communication. Finally, we also illustrate a methodical flowchart of the suggested framework. In the future, we

will extend the proposed framework with advanced machine learning algorithms and provide the performance evaluation analysis with a comparison of existing research studies.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1071926).

References

1. Singh SK, Cha J, Kim TW, Park JH (2021) Machine learning based distributed big data analysis framework for next generation web in IoT. *Comput Sci Inf Syst* 18(2):597–618
2. Sharma PK, Rathore S, Park JH (2019) Multilevel learning based modeling for link prediction and users' consumption preference in online social networks. *Futur Gener Comput Syst* 93:952–961
3. Siddique ZB, Khan MA, Din IU, Almogren A, Mohiuddin I, Nazir S (2021) Machine learning-based detection of spam emails. *Sci Program*
4. Prabhu Kavin B, Karki S, Hemalatha S, Singh D, Vijayalakshmi R, Thangamani M, Haleem SL, Jose D, Tirth V, Kshirsagar PR, Adigo AG (2022) Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks. *Wireless Commun Mobile Comput*
5. Miao Y, Chen C, Pan L, Han QL, Zhang J, Xiang Y (2021) Machine learning-based cyber-attacks targeting on controlled information: a survey. *ACM Comput Surv (CSUR)* 54(7):1–36
6. Zheng X, Zhang X, Yu Y, Kechadi T, Rong C (2016) ELM-based spammer detection in social networks. *J Supercomput* 72(8):2991–3005
7. Patil Y (2020) Detection of Clickjacking attacks using the extreme learning Machine algorithm. Doctoral dissertation, Dublin, National College of Ireland
8. Nayak J, Meher SK, Sourli A, Naik B, Vimal S (2022) Extreme learning machine and bayesian optimization-driven intelligent framework for IoMT cyber-attack detection. *J Supercomputing* 1–26

Analysis of Noise Reduction Measures for 220 kV Substation Based on Acoustic Simulation Technology



Haitao Sun, Yu Yang, and Peijie Liu

Abstract This study selects the representative 220 kV substation as the research object and simulates its indoor noise and plant boundary noise by using Odeon and Cadna/A software. Conducting noise reduction control on the substation through sound absorption, sound insulation, and other technical means investigates the noise reduction effect of different sound absorption materials and sound insulation materials in order to propose reasonable noise reduction measures and provide a practical reference for urban substation noise control.

Keywords 220 kV substation · Acoustic simulation · Noise control

1 Introduction

The noise problem generated by substations cannot be ignored and has become the focus of increasing concern [1]. Therefore, the simulation analysis of substation noise and the analysis of the impact of different materials on noise control are of positive significance for reasonable control of substation noise. In this study, a 220 kV substation is selected as the research object. The method of combining indoor and outdoor simulation is used to predict the noise of the unit room in the main transformer room of the substation. Odeon software is used for indoor noise simulation, and Cadna/A software is used for outdoor noise simulation. Firstly, Odeon software calculates the noise spectrum inside the street-facing exterior wall of the main transformer room. Then the calculated noise spectrum is inserted and attenuated according to the sound insulation spectrum of the wall. The calculated spectrum is used as the exterior wall source strength to predict the noise in the Cadna/A software. Finally, the noise value of the sensitive points at the substation boundary is predicted.

H. Sun · Y. Yang · P. Liu (✉)
South China University of Technology, Guangzhou 510641, People's Republic of China
e-mail: pjliu@scut.edu.cn

2 Overview of Substation Noise

Substation noise is mainly divided into electromagnetic noise and mechanical noise. The electromagnetic noise is mainly generated by the vibration of the iron core silicon steel sheet inside the transformer, reactor, and capacitor in the alternating electric field; mechanical noise is mainly generated by the cooling fan and oil pump during operation [2]. The primary noise source affecting the indoor environment and boundary noise of the substation is the transformer, especially the high noise level of the main transformer (i.e., the main transformer) [3]. The running noise of the main transformer belongs to the medium- and low-frequency noise, and the human ear is relatively sensitive to it due to its low tolerance. At the same time, low-frequency noise can quickly cause significant damage to the human body. Therefore, according to the causes of substation noise, the noise control of a substation can be carried out from three aspects: noise source, transmission path, and receiver. In this study, the primary consideration is controlling the noise transmission path. Noise reduction measures such as sound absorption, sound insulation, and noise elimination are adopted to attenuate the noise in the transmission process. The sound absorption technology usually installs the sound absorption material on the indoor surface of the noise source to eliminate the reverberation effect of sound waves in the room. Sound insulation technology mainly uses unique materials and components to isolate indoor and outdoor air and reduce the transmission of sound energy.

3 Influence of Sound Absorbing and Insulating Materials on Noise Control in Substation

3.1 *Sound-Absorbing Materials*

By arranging sound absorption materials on the surface of the indoor enclosure structure and the ceiling of the noise source, the characteristics of the indoor sound field are changed, and the noise transmission is controlled [4]. According to the sound absorption principle, sound-absorbing materials can be roughly divided into two categories: porous sound-absorbing materials and resonance sound-absorbing materials. Porous sound-absorbing materials have good absorption performance for medium- and high-frequency noise and are widely used in substation sound absorption and noise reduction projects. Resonant sound-absorbing materials are commonly used resonance sound absorption materials, including perforated sound-absorption boards and microporous sound-absorption boards. The sound absorption coefficient is high in a specific frequency band [5], but it also has the disadvantage that the sound absorption band is not wide enough.

In actual practice, the combined sound absorption structure of “panel + glass wool + cavity” is usually used to extend the sound absorption band. The spectrum of exterior wall source intensity calculation shows that the 500 Hz noise is the

Table 1 Sound absorption coefficient and structure of sound-absorbing materials

Sound-absorbing material for interior wall	Structure description	Frequency (Hz)						NRC
		125	250	500	1000	2000	4000	
Microporous combined sound-absorbing plate	1	0.49	0.91	0.91	0.81	0.76	0.69	0.85
Microporous honeycomb sound-absorbing plate	2	0.55	0.92	0.93	0.86	0.75	0.66	0.85
Ultra-microporous sound absorption plate	3	0.41	0.89	0.91	0.74	0.61	0.49	0.80
Aluminum strip seam sound-absorbing board	4	0.35	0.75	0.88	0.81	0.7	0.56	0.75
Aluminum plate mesh sound-absorbing plate	5	0.36	0.88	0.86	0.90	0.87	0.70	0.85
Plaster	—	0.02	0.02	0.03	0.03	0.03	0.04	—

1. Full frequency ultra-microporous sound absorption board + 50 mm thick glass wool + 150 mm cavity 2. Microporous honeycomb sound absorption board + 25 mm thick glass wool board (volume weight: 32 kg/m³) + 50 mm thick glass wool (volume weight: 48 kg/m³) + 50 mm cavity 3. Ultra-microporous sound absorption board + 50 mm thick glass wool board (volume weight 48 kg/m³) + 100 mm cavity 4. Aluminum strip sound-absorbing board + 50 mm thick glass wool board (volume weight 32 kg/m³) + 100 mm cavity 5. Aluminum mesh sound-absorbing plate + 100 mm cavity.

octave frequency spectrum with an immense contribution to LAeq [6]. Therefore, in order to reduce the spectrum data of exterior wall source intensity in the substation indoor noise simulation, the following five combined sound absorption structures are used in the unit room of the main transformer room, namely, microporous combined sound absorption plate Microporous honeycomb sound-absorbing plate, ultra-microporous sound-absorbing plate, aluminum strip seam sound-absorbing plate, aluminum mesh sound-absorbing plate. The sound absorption materials of the interior wall are arranged on the other three walls and ceilings except for the street wall of the main transformer room. The corresponding structure description, sound absorption coefficient, and noise reduction index NRC are shown in Table 1.

3.2 Sound-Insulation Materials

Sound insulation is one of the essential means of noise control, which is to use unique materials or components to isolate the air to prevent the spread of noise. Generally, sound insulation’s noise reduction effect is better than sound absorption’s. Standard sound insulation components include sound insulation walls, sound insulation doors, and sound insulation windows. As the external wall of the main transformer room

Table 2 Sound insulation capacity and structure of sound insulation materials

Material of pressure relief wall	Structure description	Frequency (Hz)						Rw
		125	250	500	1000	2000	4000	
Fiber cement board	1	27.5	35.6	44.2	52.2	50.5	56.3	47
Lightweight wallboard	2	34.8	33.2	34.1	36.7	49.2	50.2	39
Fabricated keel wall	3	30.4	36.8	43.7	49.7	56.4	48.5	47
Galvanized sheet	4	16.3	21.9	27.3	34.8	40.1	44.2	32

1. 12 mm fiber cement board + 50 mm cavity + 12 mm inorganic incombustible composite board + 50 mm cavity + 12 mm fiber cement. 2. Ceramic lightweight environment-friendly wallboard. 3. Putty (5 mm thick) + 9 mm thick fiber cement board + 90 mm thick keel (embedded with 90 thick fireproof and sound insulation glass wool) + 9 mm thick fiber cement board + 5 mm thick putty. 4. Aluminum alloy perforated plate (1.2 mm thick, perforation rate 20%) + superfine glass wool felt wrapped with PVF waterproof membrane (glass wool felt thickness 65 mm, volume weight 32 kg/m^3) + 1.0 mm thick galvanized sheet.

may explode under extreme conditions, some walls shall be designed as pressure relief walls with a density of $\leq 60 \text{ kg/m}^3$ [7]. This study will select the street wall of the main transformer room as the pressure relief sound insulation wall and mainly study the relationship between the sound insulation of the pressure relief wall and the external wall source noise.

According to the different sound insulation mechanisms, the pressure relief walls mainly include the following: Fiber cement board, 75 mm lightweight environment-friendly wallboard, fabricated keel wall, and galvanized plate. Table 2 shows the structures of four kinds of pressure relief wall sound insulation materials and their sound insulation quantities at each frequency.

3.3 Simulation Data Analysis

Odeon calculates the noise spectrum of the street-facing exterior wall of the main transformer room. We studied the noise spectrum at 0.3 m position inside the pressure relief wall, inserted and attenuated the simulated noise spectrum according to the sound insulation spectrum of the wall, and finally corrected the A-weighted spectrum to obtain the final A-weighted spectrum of the main transformer room's street-facing exterior wall source strength. According to the standard《Technical Guidelines for Noise Control in Substations》, considering the air attenuation, simulate the real plant boundary noise, and use the Cadna A to predict the environmental noise of the substation. The receiving surface is arranged at the boundary of the substation to predict the boundary sound pressure level. The calculation results are shown in Table 3.

It can be seen from the calculation results that when the interior wall material is plastered without considering the sound insulation wall, the sound pressure level

Table 3 Indoor sound pressure level, external wall source strength A-weighted sound pressure level, and factory boundary sound pressure level of four sound insulation materials under five sound absorption materials

Sound absorbing material	Indoor sound pressure level (0.3 m inside the pressure relief wall)					Sound insulation material	Sound pressure level of exterior wall	Plant boundary sound pressure level
	125	250	500	1000	2000			
1. Microporous combined sound-absorbing plate	60.9	63.9	68.5	56.5	54.1	46.2	24.5	16.0
							32.0	23.0
							24.0	15.0
							39.7	31.0
2. Microporous honeycomb sound-absorbing plate	60.8	63.8	68.2	56.6	54.3	46.4	24.3	15.0
							31.8	23.0
							23.8	15.0
							39.5	31.0
3. Ultra microporous sound absorption plate	61.1	63.9	68.9	57.5	55.6	48.1	24.7	16.0
							32.4	23.0
							24.3	15.0
							40.0	31.0
4. Aluminum strip seam sound-absorbing board	62.1	64.1	68.5	56.9	55	47.3	24.8	16.0
							32.1	23.0
							24.2	15.0
							39.9	31.0
5. Aluminum plate mesh sound-absorbing plate	61.1	64.2	68	56	54	46.2	24.4	15.0
							31.6	23.0
							27.4	18.0

(continued)

Table 3 (continued)

Sound absorbing material	Indoor sound pressure level (0.3 m inside the pressure relief wall)						Sound insulation material	Sound pressure level of exterior wall	Plant boundary sound pressure level
	125	250	500	1000	2000	4000			
6. plaster	71.9	75.4	80.3	67.8	63.7	53.1	d. Galvanized sheet	39.5	31.0
							a. Fiber cement board	36.0	27.0
							b. Lightweight wallboard	43.8	35.0
							c. Fabricated keel wall	35.6	27.0
7. Plaster							d. Galvanized sheet	51.4	42.0
	71.9	75.4	80.3	67.8	63.7	53.1	No consideration of sound insulation	78.1	69.0

of the prominent transformer room noise at the plant boundary is 69 dBA. The noise can be reduced by about 11–12 dB (A) if the sound-absorbing materials are reasonably arranged on the indoor walls of the main transformer. Among them, when the microporous honeycomb sound absorption plate is used as the sound absorption material of the interior wall, the noise sound pressure level at the plant boundary is the lowest, and the sound absorption effect is the most obvious. When plastering and sound insulation walls are used as internal wall materials, the sound pressure level of the prominent transformer room noise at the plant boundary is 27–42 dBA. Among them, fiber cement and fabricated keel wallboard have the most apparent sound insulation effect, while galvanized wallboard has the worst sound insulation effect. When sound-absorbing materials are used for interior wall materials and sound-insulating materials are used for street-facing exterior walls, the noise boundary sound pressure level of the main transformer room is within the range of 15–31 dBA. Proper sound insulation measures can reduce the noise by 8–16 dB (A) based on sound absorption treatment. When sound absorption and insulation materials are used together as noise reduction measures for substations, the sound pressure level of noise at the plant boundary shall meet the specification requirements of less than 45dBA.

4 Conclusion

In this paper, the noise characteristics of each frequency in the main transformer room of the 220 kV substation are simulated and analyzed, and noise reduction is proposed by controlling the noise transmission path. The influence of five kinds of sound-absorbing materials and four kinds of pressure-relief wall insulation materials in the main transformer room was studied. The main noise reduction measures for the 220 kV substation are shown in Fig. 1. The main transformer room adopts passive noise reduction treatment, and the perforated composite sound absorption structure is installed on the three walls and ceilings, except the pressure relief wall. This measure can reduce the radiated noise value outside the pressure relief wall by about 10dBA, which is the preferred measure for noise control in the substation. The five kinds of sound-absorbing materials in this study have little difference in sound absorption effect, so the selection of sound-absorbing materials can be made according to the decoration needs or cost requirements. Indoor substations in urban areas should be equipped with low-noise equipment as far as possible to achieve the purpose of noise control and control and avoid noise interference to residents around the substation.

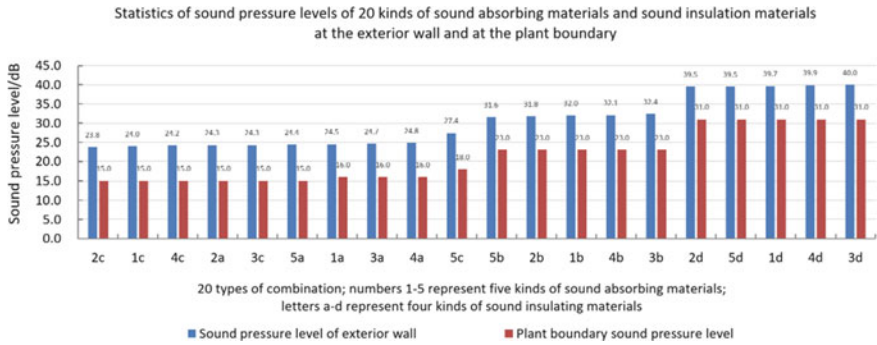


Fig. 1 Statistics of sound pressure levels of 20 kinds of sound absorbing materials and sound insulation materials at the external wall and at the plant boundary

Acknowledgements This research was supported by the National Natural Science Foundation of China (51878281).

References

- Xu L (2013) Noise prediction and optimal control design of outdoor substations. *Noise Vib Control* 01:152–156
- Wang X, Li W, Jin D (2017) Research progress of substation noise control technology. *Power Technol Environ Prot* 06:34–37
- Chen Z, Li Q (2008) Evaluation and prevention of environmental noise in substation. *Power Environ Prot* 01:61–62
- Gao L, Shang F (2007) Research and application of sound absorbing materials. *Chem Ind J* (02):63–65+69
- Fan C, Gong L, Liu Z, Yi Y, Lin L, Chen J (2016) Analysis of noise characteristics and noise reduction measures of typical urban substations. *Smart Grid* (10):983–987
- Zhang L, Chen P, Sun H (2021) Prediction and analysis of 220kV substation based on geometric acoustic simulation. In: *INTER-NOISE and NOISE-CON congress and conference proceedings 2021*, vol 263, no 1, pp 5142–5146. Institute of Noise Control Engineering, Washington DC (2021)
- Rong F, Cheng Z, Liu P (2021) Study on sound insulation performance of pressure relief wall of transformer chamber. In: *INTER-NOISE and NOISE-CON congress and conference proceedings 2021*, vol 263, no 1, pp 5197–5202. Institute of Noise Control Engineering, Washington DC

A Study on Metaverse Realistic Content Education Platform Using Deep Learning



Hyunsook Lee and Sekyoung Youm

Abstract Artificial intelligence technology combined with Metaverse Realistic Content (VR/AR Content) is a powerful combination that provides a diverse educational environment. Recent developments in artificial intelligence technology can enhance the quality of education without the help of assistants in immersive education and it can be seen as a field that can be developed in terms of resource efficiency and time saving by learning repetitive outcomes (success/failure). However, this combination has challenges. Currently, Unity's Machine Learning Agent toolkit exists, but it has limitations for specific languages and specific environments and has additional learning challenges. In addition, it is difficult to realize that it is necessary to support various front-end frameworks and various hardware, and to support the entire combination. In addition, it is necessary to study connections between frameworks to support mobile devices and web environment. This problem has long been recognized in computer science and has been solved by compiler technology. In this paper, we propose a Framework Bridge Model based on LLVM IR (Low-level virtual machine intermediate representation) in VR/AR development environment, so that deep learning developers can choose the optimal combination of frameworks.

1 Introduction

Virtual reality refers to a specific environment or situation similar to reality created by artificial intelligence using a computer, or software itself, but is not real. VR aims to create immersive virtual reality by combining human senses such as hearing, touch, and sight with software and hardware. Recently, artificial intelligence technology has developed rapidly, and artificial intelligence, especially deep learning technology, is

H. Lee · S. Youm (✉)

Department of Industrial and Systems Engineering, Dongguk University, Seoul, Korea

e-mail: sekyoungyoum@gmail.com

H. Lee

e-mail: lhsnjhj@naver.com

being used in the VR/AR environment. This combination has solved many problematic compiler techniques, but it has a long recognized problem. In this paper, we propose a platform integration technology for related technologies and deep learning technologies represented by Unreal and Unity, and a framework model for LLVM IR-based bridge connection in a VR/AR development environment. Finally, we propose an example of using a VR environment in a metaverse environment that implements SkechRNN [1] using Google QuickDraw [2].

A representative effort to combine artificial intelligence technologies in VR/AR environments is the Unity Machine Learning Agent (hereinafter ML Agent) [3, 4].

With the ML Agent SDK, you can transform a game or simulation created with the Unity editor into an environment where you can train intelligent agents through an easy-to-use Python API using machine learning methods such as reinforcement learning. However, such an environment has a problem in that the development environment is limited to a specific language (Tensorflow) and a specific environment (Unity Engine), and new functions must be learned.

Intel nGraphTM [5] is a customized solution that supports standards that can interoperate between solutions and accommodate new architectures as the number of solutions increases as artificial intelligence techniques based on neural networks become more complex. As a result of these efforts, various types of framework integration projects are in progress, and representative projects include NNVM [6], DLVM [7], and ONNX [8].

2 Metaverse Education Platform Configuration

The system is basically developed based on Microsoft.Net Framework. F# selected as VAL is a language supported by Microsoft.Net Framework, and the Windows operating system that.Net Framework is supported is supported by various 3D engines for games, and is the environment in which VR/AR development support is most actively performed. In addition, as the.Net Framework currently supports other operating systems such as Linux/OS-X through the Mono framework, an open source project. It was determined that it would be advantageous to support various computing environments in the future. For efficient execution management of multiple objects in VR space, actor-based AKKA.Net is used as a message infrastructure. Akka.Net is a solution that has already been proven in Java/JVM, and it has also started supporting.Net Framework through Akka.Net. As shown in the platform in Fig. 1, VR/AR App prototypes are created using the VR Editor for beginner and intermediate professionals. Prototypes are retrieved and played back with content created by other users of the platform. In addition, the source code used in the production is shared among users and presents a model that can be collaborated through a web-based sharing platform, and an analysis system is designed and implemented.

In this paper, we propose a framework connection model based on LLVM IR in VR/AR development environment. As introduced in the related research above, creating an artificial intelligence development environment combined with VR/AR is

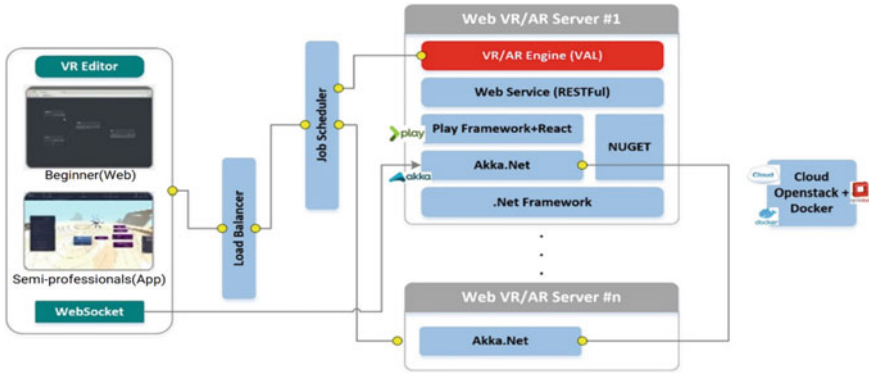


Fig. 1 Metaverse education platform

a VR/AR development environment (Unity, Unreal, etc.) and a deep learning development environment (TensorFlow, NXNet, CNTK, PyTorch, Caffe2, etc.) Various front-end frameworks already exist, and various hardware (NNP, CPU, cuDNN, etc.) must also be supported. In addition, there are many challenges such as the VR/AR environment must support mobile devices or web environments.

The proposed platform applied the IR-to-IR connection method using LLVM technology. Here, the LLVM project [9] is a robust and optimized open source compilation platform that competes with GCC in terms of compilation speed and performance of the generated code. The LLVM-based compiler consists of a part that converts from high-level source language to LLVM IR [10], and a part that provides optimization, program transformation, and static analysis. Here, the intermediate language (IR) refers to a data structure or code used internally to represent source code in a compiler or virtual system.

3 Implementation

Here is a simple application example implemented using the framework connection model based on LLVM IR in the VR/AR development environment. It is very difficult to design a VR environment by selecting a menu in an In-VR environment. We will introduce the process of automatically selecting and arranging the relevant assets in the process of drawing objects through real-time hand tracking and gesture recognition using the SketchRNN technique used in Google's QuickD raw dataset. The implementation example of (Fig. 2) shows how much development potential the application of artificial intelligence deep learning technology in a VR environment is.

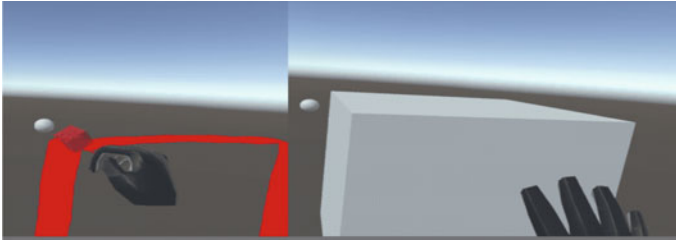


Fig. 2 Automatic asset recognition and selection during the drawing process using deep learning

4 Conclusion

In this paper, we proposed a framework connection model based on the platform and LLVM IR for software education in VR/AR development environment. Although there are many challenges in creating an artificial intelligence development environment combined with VR/Ar, it could be solved by using the intermediate language translation LLVM-IR compiler technology. The core technology proposed in this paper is to link the WebAssembly code generation technology using the LLVM IR technique with the NNVM compiler, a new compiler in the field of deep learning integration.

By using the Metaverse VR/AR content education platform developed through this study, it is expected that the general public, not even professional programmers, will be able to easily obtain the following effects through basic VR/AR experiences.

- (1) Provides an easy VR/AR experience for beginners

It provides convenient screen composition and usage for the general public without specialized knowledge to perform tasks such as creating or programming VR/AR contents.

- (2) Provide deep learning-based in-VR environment

In programming and confirmation of results, it is possible to proceed simultaneously in a deep learning-based in-VR environment. In other words, content production is possible in the in-VR environment, and the operation of the generated content can be checked immediately.

In addition to programming education, which is not a vaguely difficult field, it will be a driving force for future growth as a professional development manpower by inducing understanding and interest in the VR/AR specialized field.

- (3) Arouse interest

It is expected that it will become a better VR/AR education platform if more diverse libraries and curriculum development and usability improvement are carried out in the future.

References

1. Ha D, Eck D (2017) A neural representation of sketch drawings. ArXiv preprint. <https://magenta.tensorflow.org/sketch-rnn-demo>
2. Jongejan J, Rowley H, Kawashima T, Kim J, Fox-Gieg N (2016) The Quick, Draw!—A.I. Experiment. <https://quickdraw.withgoogle.com/>
3. Juliani (2017) Introducing: unity machine learning agents. <https://blogs.unity3d.com/2017/09/19/introducing-unity-machine-learning-agents/>
4. Chen T et al (2018) TVM: end-to-end compilation stack for deep learning. In: SysML conference
5. Bansal AK, Bhiwandiwalla A, Bobba J, Brookhart M, Chakraborty A, Constable W, Convey C, Cook L, Kanawi O, Kimball R, Knight J, Korovaiko N, Kumar V, Lao Y, Lishka CR, Menon J, Myers J, Narayana SA, Procter AM, Webb TJ (2018) Intel@nGraph TM an intermediate representation, compiler, and executor for deep learning
6. Allen PG et al (2017) NNVM compiler: open compiler for AI frameworks. <https://tvm.ai/2017/10/06/nnvm-compiler-announcement.html>
7. Wei R, Adve VS, Schwartz L (2017) DLVM: a modern compiler infrastructure for deep learning systems. CoRR, abs/1711.03016
8. ONNX 2017. Retrieved 4 Jan 2018 from <http://onnx.ai>
9. Lattner C, Adve V (2004) LLVM: a compilation framework for lifelong program analysis & transformation. In: CGO 04: proceedings of the international symposium on code generation and optimization: feedback-directed and runtime optimization
10. Zhao J et al (2012) Formalizing the LLVM intermediate representation for verified program transformations. In: ACM Sigplan notices, vol 47, no 1. ACM

Strategies for Establishing u-health Care Center Through Research on u-health Service Awareness and Intention to Use



Sekyoung Youm

Abstract We established a ubiquitous health (u-Health) service center in the entrance to a Sungji Park hiking trail in Busan that allowed users to directly experience and participate in actual u-Health services. Moreover, a comparative analysis was performed. On a group of participants who personally used the services at a u-Health service center and a group of those who had not used u-Health services but had gained an understanding of it through information. The results suggest that to expand the u-healthcare service, individuals need to be better exposed to it in order to gain easier access. Furthermore, it will also be necessary to increase understanding of the service by sharing success stories and further improving the credibility of the examination results.

1 Introduction

Amid population aging and an increasing prevalence of chronic diseases, issues concerning medical expenditure as well as medical service accessibility and spending are emerging as major social concerns [1]. As such, in addition to establishing a public healthcare system, the government is focusing on utilizing information communications technology to reduce medical expenses and meet the consumer demand for convenient healthcare services. To reflect this trend, u-Health, a concept that combines IT with healthcare services, has recently garnered great interest [2–4]. A development of u-Health-related products and services is progressing rapidly. However, strategies for related service development and distribution via analysis of consumers or service users are somewhat lacking, and these results may distort the service usage intention.

The objective of this study is to establish a u-Health service center in which the users can directly experience the service and to conduct a comparative analysis on service awareness and use intention in a group of individuals who experienced the

S. Youm (✉)

Department of Industrial and Systems Engineering, Dongguk University, Seoul, Korea
e-mail: skyoum@dgu.edu

service firsthand and in a group who did not use the service but have an understanding of it through comprehending the provided information.

2 Materials and Methods

The u-Health service comprises four components: the u-Health service center that includes devices, the database center, a linked system with cooperating medical institutions that can prescribe suitable exercise and healthy routines, and a web service in which the users can check their results. When the users check their health status using measuring devices located inside the u-Health center, they can check their examination results immediately through the internet or via smartphone. And also, to investigate awareness of, and use intentions for, the u-Health service in individuals who had not yet directly experienced the u-Health service, informational materials and a video that explained the services provided (Fig. 1).

A total of 496 participants—270 who visited the u-Health service center and 226 who had never used it. We surveyed for directly user group and provided information user group, and analyzed the data to understand u-Health service awareness and intention use and to analyze the participants' self-perceived health status. The total number of survey questions was 12, and the three categories (awareness, intention and health status) were investigated and analyzed. The two general questions regarding awareness and intention of use have been selected from the Technology Acceptance Model (TAM). Additionally, two questions related to health status have been selected from the Korea National Health and Nutrition Examination Survey (KNHANES).

To investigate awareness of, and use intentions for, the u-Health service in individuals who had not yet directly experienced the u-Health service, informational materials and a video that explained the services provided (Fig. 2).

This study was conducted with a total of 496 participants divided into two groups as shown in Fig. 4. The study subjects who used the u-Health service were 287 individuals who visited the u-Health center between October 18, 2013, and November 3, 2013. Of those u-Health service users who answered a survey, 17 with incomplete responses or data outliers were excluded. As such, a total of 270 surveys was used for the analysis.

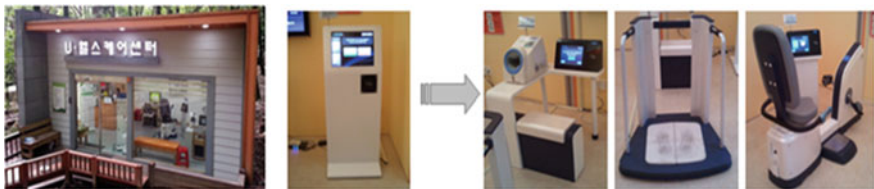


Fig. 1 u-Healthcare center

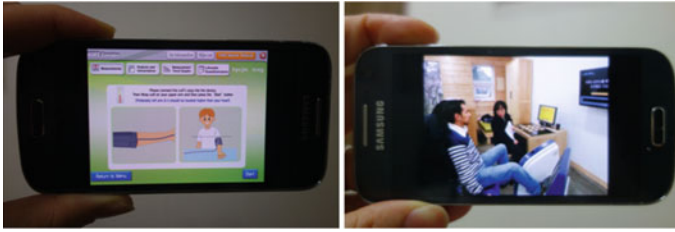


Fig. 2 Instructional materials and the video

During that same period, a survey was conducted with 285 individuals—recruited in front of the Busan City Hall—who had not used the u-Health service. Of those who knew about u-Health but had never used the service directly, 59 with incomplete responses or data outliers were excluded, and a total of 226 surveys were analyzed.

For studying the group of individuals who had no direct experience of the center, those who mentioned they had previously used the u-Health service were prevented from being recruited. For those who were to participate in the study by directly using the service at the park, the additional questions were asked concerning their usage patterns after they came to know about the service.

To investigate the intention of service use, five questions were asked. To analyze the difference for each item between the two groups, a Pearson's chi-square test was conducted.

As for individuals who directly used the u-Health service and who learned about it from the information given, their self-perceptions of health were investigated. For this purpose, a t-test was conducted on the two groups' self-score measurements of their own health status. In addition, we questioned them about their desired life expectancy, and a Pearson's chi-square test was conducted.

3 Results

Only 34.7% of those who directly used the service and 22.8% of those who learned about the service through the provided information said they had ever heard of the u-Health service. The result in Table 1 shows that those who used the service directly are less likely to use the service again or find it credible compared to those who did not experience it directly. This result shows that investigating the service use intentions of participants who only learned about it through provided information can distort the actual intention of use.

In addition, in a question that asked about individuals' self-perceived health status, those who directly utilized the service thought they were less healthy than those who did not use the service, and this difference was statistically significant as described in Table 2. In other words, those who utilized the health care considered things

Table 1 Comparison between directly user group and provided information user group

	Pearson chi ²	p-value
Usefulness	38.960	0.000
Recommend	158.168	0.000
Payment	4.450	0.035
Interests	68.395	0.000

Table 2 Self-health score

t-test				Group average	
t	df	p-value	Average difference	Directly user group	Provided info user group
6.94	494	0.000	-0.514	2.93 ± 0.612	3.45 ± 1.015

more realistically than those who only learned about the service through provided information.

4 Conclusion

This study constructed a u-Healthcare center in which the subjects could experience u-Health services firsthand, and verified that their service awareness, use intention, and service use allowed them to understand their health status while preventing them from overestimating their health. We expect the findings to serve as useful reference data for the construction and promotion of u-healthcare centers in the future. Yet, as this study includes random participants without considering their age and gender, and focuses only on a region where the u-Healthcare center was located, the samples used do not represent the public. Therefore, future research should sample various user groups to represent the whole population and provide results that are more significant. In addition, various management indicators for u-Health should be developed to suggest a more meaningful direction for future u-Health research.

References

1. Kim S (2010) Trend and prospect of medical-IT convergence: u-health service. *Mag IEEK* 37(6):49-59
2. Baek M (2012) Research on recognition for medical quality improvement and Outcomes of U-Healthcare: focusing on employees in major hospitals. Master's thesis in Kyunghee University

3. Hussain M, Al-Haiqi A, Zaidan A et al (2015) The landscape of research on smartphone medical apps: coherent taxonomy, motivations, open challenges and recommendations. *Comput Methods Programs Biomed* 122(3):393–408
4. Kontos E, Blake K, Chou W, Prestin A (2014) Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. *J Med Internet Res* 16(7):e172

Method of Predicting the Danger Zone of Tidal Greening Using Machine Learning



Gayoung Kim 

Abstract A sea forest is a forest made of algae and seaweeds in the sea. It is effective in preventing the melting of the shoreline, increasing the catch, and maintaining the marine ecosystem, and is attracting attention as a key resource for blue carbon. However, due to the sea forest creation project that does not take into account the specificity of each sea area, the number of seaweeds is decreasing in some sea areas compared to before the creation. Therefore, in this study, in order to achieve the goal of creating a sea forest considering the characteristics of each sea area, a model for predicting the rate of tidal greening using ocean data was created. Through the random forest model, which showed 95% explanatory power, the top 20 sites with a high incidence of tidal greening were selected, and the amount of carbon dioxide reduction through sea forest creation was calculated. As a result of the calculation, it is expected that about 1.14 million tons of carbon dioxide will be absorbed when the sea forest is established in the area, which is about 14 times the amount of carbon dioxide reduction that was previously about 80,000 tons.

Keywords Predicting the Danger Zone · Machine learning · Tidal greening · Time series data analysis

1 Introduction

The sea forest, the core resource of blue carbon, is composed of seaweeds and seaweeds and plays a role in maintaining the sea ecosystem.

As the creation of sea forests as a means of 2050 carbon neutrality drew attention, a total area of 26,644 ha was created in 211 locations by 2021.

However, it is known that the effect on this is not very large because the specificity of each sea area is not taken into account [1, 2].

G. Kim (✉)

Faculty of General Education, Kangnam University, Yongin-Si, Republic of Korea

e-mail: dolga2000@kangnam.ac.kr

Among the various causes, sea desertification is considered to be the biggest cause. Mud greening is a phenomenon in which seaweed disappears from the coastal bedrock area and white lime algae sticks to the bedrock area and turns white, which has a great impact on marine ecosystems and fisheries. It is occurring all over Korea, and its area is gradually increasing. Although the exact cause of this phenomenon is not yet known, it is known that it is caused by a combination of physical, chemical, and biological factors. Therefore, this paper identifies marine ecological factors that can affect the occurrence of tidal green, and creates a model that selects the top 20 regions with the largest tidal greening area through time series analysis of the factors, and considers the characteristics of each sea area. After selecting, we want to calculate the expected annual carbon dioxide reduction due to the creation of sea forests [3, 4].

2 Random Forest Model

The random forest model is a machine learning algorithm technique used in classification and regression analysis as one of the ensembles learning methods [5]. The random forest model generates a decision tree by sampling data in a bootstrap method. A plurality of such decision trees are created, and the final value is derived from the results derived from each decision tree. In the case of a classification model, the final result is indicated by a majority vote through voting, and in the case of a regression model, the final result is expressed as the average of the results. In this study, we try to make predictions through random forest. When the number of predictor variables increases, the random forest model has the advantage of increasing predictive power as the number of explanatory variables increases by supplementing the limitations of existing traditional regression analysis or logistic regression analysis models, which have limitations on exogenous predictor variables or distribution of error terms.. This model was selected as it was judged that it is suitable for analyzing the tidal greening phenomenon with many explanatory variables (Fig. 1).

3 Mud Greening Risk Prediction Model

3.1 *Selection of Key Variables for Ocean Data Preprocessing and Analysis*

In this paper, we propose a method for predicting tidal greening risk areas using machine learning.

To do this, first, key variables for ocean data preprocessing and analysis are selected.

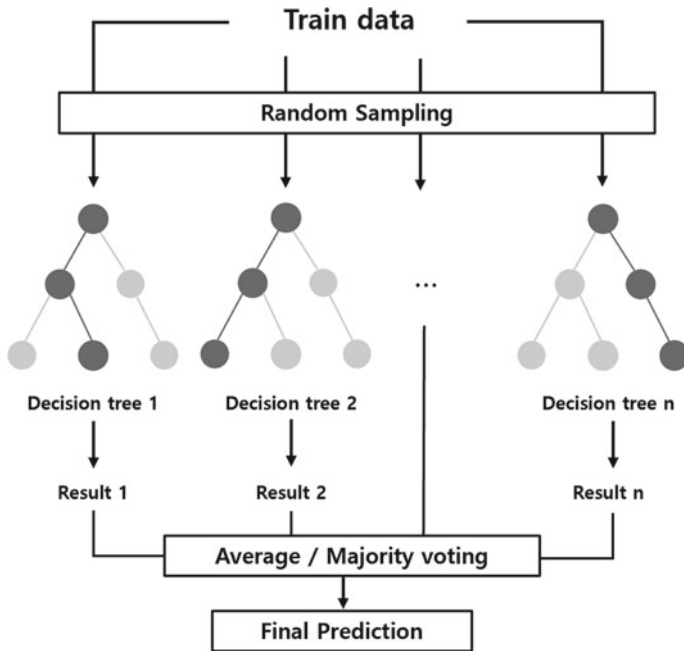


Fig. 1 Random forest simplified

Figure 2 is an overall structural diagram of the process of preprocessing the collected ocean data.

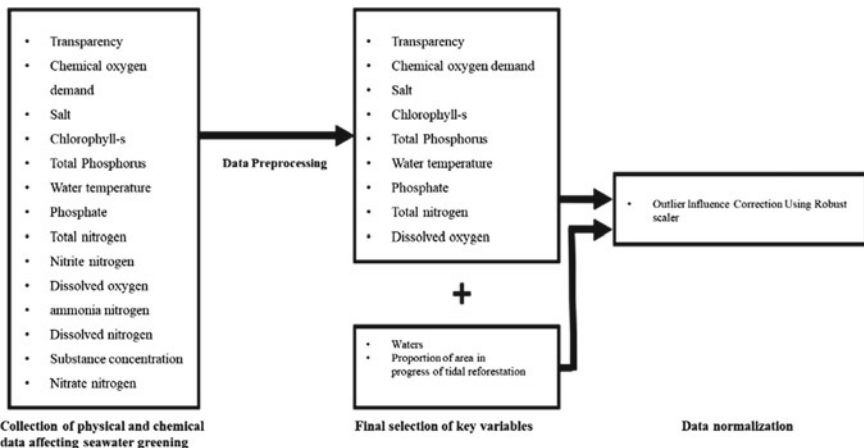


Fig. 2 Structural diagram of the selection of major variables for preprocessing and analysis of ocean data for prediction of tidal whitening event risk areas

Data normalization is performed using the RobustScaler for nine variables, except for the ratio of the area of tidal reforestation, among the finally selected major variables.

Machine learning compares the features of data to find patterns in the data. At this time, all data input to the machine learning model is reflected at the same scale and undergoes a normalization process so that it is not affected by outliers.

For this, the data corresponding to the median (Q2) of each feature is set to 0 using the robustScaler, and then normalized by dividing it by IQR (Q3-Q1), which is the difference between the Q1 and Q3 quartiles.

Through this data preprocessing process, the influence of outliers can be minimized, and the learning accuracy of the machine learning model to be applied later can be increased.

3.2 Predicting Future Ocean Data Through Time Series Data Analysis

Based on the fbprophet, using the marine environmental monitoring network data that has undergone the preprocessing process, the values of major variables that cause tidal greening in the future ocean data are predicted. fbprophet, a time series data analysis library released by Facebook, has the advantage of easy model training with high accuracy, fast and intuitive hyperparameters. In this paper, to predict the data of 2021 using the marine environmental monitoring network data from 2002 to 2020, time series prediction of 8 variables excluding 'sea area' out of 9 main variables to be used for analysis using fbprophet proceed with Predict the model with the smallest Root Mean Square Error (RMSE) value for each feature through hyperparameter tuning. RMSE is a measure used when dealing with the difference between predicting and real value.

3.3 Results of Selecting the Location of the Sea Forest

In this paper, a random forest model is used to predict the expected area ratio of tidal greening using machine learning. In the random forest model, parameters such as `n_estimator`, `max_depth`, and `gamma` exist. `n_estimator` is the number of decision trees in the random forest, and as this value increases, overfitting is prevented, but memory and training time increase, so it is necessary to set it to an appropriate level. `max_depth` means the depth of the tree, and the larger the value, the more complex the tree is created. This parameter can be used to proactively respond to situations where the model becomes complex and overfitting occurs. `Gamma` determines the extent to which a parameter affects one training sample. The larger the `Gamma` value, the

Type	Area (km ²)	Standing stock (g/m ²)	P-B ratio	Carbon cont. (%)	C absorption (× 10 ³ tonC/year)
Seagrass (amamo-ba)	495	200-1,200	4.0	30-40	485
<i>Laminaria</i> (kombu-ba)	357	3,700-5,400	3.5	25-31	1,156
<i>Sargassum</i> (garamo-ba)	857	120-1,800	1.2	33-37	346
<i>Ecklonia</i> (arame-ba)	645	1,000-3,700	1.1	32-34	562
<i>Gelidium</i> (tengusa-ba)	190	220	1.1	36-40	17
Others	615	140-980	1.0	30	103
Total	2,012	-	-	-	2,669

Fig. 3 Seaweed resources as a source of carbon fixation

narrower the influence range, and the more complex the model. In this study, $\gamma = 1$, $\text{max_depth} = 4$, and $\text{n_estimator} = 300$ were set (Fig. 3).

4 Conclusion

Finally, the R_{squared} value of 0.9580 can be confirmed through the random forest. R_{squared} is an evaluation index for the performance of the regression model. It shows how much the independent variable explains the dependent variable. The closer to 1, the higher the explanatory power. Based on this, 20 points with a high increase or decrease in the rate of tidal reforestation are first selected as locations for sea forest creation. At this time, the increase/decrease in the ratio of tidal greening was set as the value obtained by subtracting the ratio of the past tidal greening from the expected area ratio of tidal greening in 2021.

Based on the data provided by the Korea Fisheries Resources Corporation, about 126.275 ha of sea forests will be installed in the above areas for each project sea area. In order to calculate using the expected annual CO_2 reduction (Daisuke Muraoka, 2004), the seaweed with the largest CO_2 reduction for each sea area is selected. If it is assumed that seven kelp in the East Sea area, five seaweed in the West Sea area, and kelp, which has a large amount of carbon dioxide reduction in eight southern sea areas, are planted in the sea forest out of the 20 selected sites, the expected annual carbon dioxide reduction is about 1,146,855 t CO_2 . This absorbs the carbon dioxide generated when traveling 18,024 round trips between Seoul and New York, which is about 45 times the amount of carbon dioxide absorbed by forests with the same area. Accordingly, the creation of sea forests as a means of 2050 carbon neutrality is justified, and additional expected effects such as restoration of marine ecosystems, prevention of the spread of tidal greening, and increase in fishery profits can be expected through sea forests.

Acknowledgements This Research was Supported by Kangnam University Research Grants 2021.

References

1. Kim D-K, Hwang S-J, Choi O-I, Choi I-H, Han M-I, Shin Y-J (2011) The effect of climate change on the spread of sea greening in the Jeju coast. *Fish Resour Manage* 1(1):1–17
2. Kwon HO, Kim JM, Lee SM (2007) Seaweed restoration module in the green sea area. In: *Proceedings of the Korean society of marine environment and energy conference*, vol 72
3. Oh TG, Kim DK, Kim CG, Lee MO, Cho JK (2010) Optimal site selection considering the physical environment of the sea forest creation area. *J Korean Soc Mar Eng* 34(1):183–194
4. Yoo HM, Park JY, Song YS, Lee EJ, Kang SI, Kim JH, Park MJ, Byun YJ, Sung HW, Oh HK, Yoon HY, Lee SH, Jeong SY (2016) A study on the actual condition of gaet greening in Korea. *Appl Geogr* (33):86–89
5. Juliani, *Introducing: unity machine learning agents* (2017). <https://blogs.unity3d.com/2017/09/19/introducing-unity-machine-learning-agents/>

A Study on the Location Selection Model for a Remote Image Dispenser Using Machine Learning



Teachang Ryu, Changhoon Choi, and Ga-young Kim 

Abstract One of the phenomena of regional medical disparity is the time gap during late night hours, holidays, and weekends when pharmacies are not operated, and the concentration of pharmacies or stores selling over-the-counter medicines in the metropolitan area. As a way to solve this problem, an image dispenser appeared. There is a restriction that the image dispenser should be installed only in front of a pharmacy. However, if the restrictions are followed, the temporal gap will be resolved, but the phenomenon of concentration in the metropolitan area still exists. Based on these points, this paper proposes and compares the location selection model for the image dispenser in the case where it is installed only in front of the pharmacy and the case where it is installed without being limited to the pharmacy to the policy proponent. Through thesis analysis, the elderly who use primary medical facilities and pharmacies the most were set as the target. As a result of comparing the two models using the number of the elderly population with access to medicines, it was found that the number of cases where the area where the image dispenser was installed was not restricted to a pharmacy was 573% more than the case where the country was restricted to the previous country. This shows that the case where the installation area is not limited to the pharmacy can solve the preceding problem more efficiently.

Keywords Video dispenser · Machine learning · K-means · Elbow method · Location selection analysis

T. Ryu

Department of Fire and Disaster Protection Engineering, Woosong University, Daejeon, Republic of Korea

C. Choi

Department of Software, College of Science & Technology, Kyungpook National University, Sangju, Republic of Korea

G. Kim (✉)

Faculty of General Education, Kangnam University, Yongin-si, Republic of Korea
e-mail: dolga2000@kangnam.ac.kr

1 Introduction

Unlike a vending machine where a consumer selects the desired product, the remote image dispensing machine [1] is a medical device that has secured various safety features, such as removing the selection function and installing a large screen where the pharmacist guides the medication.

In Korea, pharmaceuticals are divided into over-the-counter drugs and prescription drugs, and the remote imaging device sells limited over-the-counter drugs.

The remote video dispensing machine has net functions such as safety and convenience for consumers, and expansion of access to medicines by consumers, including emergency patients.

It has few side effects through the system, under the remote medication guidance of the pharmacist, and the sale items are limited to over-the-counter drugs, so it is safe for consumers.

In addition, accessibility is improved by filling in the time gap where medicines cannot be purchased, such as late at night, on weekends, and on public holidays.

According to the Pharmaceutical Affairs Act, pharmacy founders must not sell medicines in places other than the pharmacy or store.

However, with the recent emergence of Supreme Court precedents that fully take into account remote work as well as the spread of COVID-19 and the development of advanced technology, amendments to the medical law related to telemedicine are being prepared.

2 Related Research

2.1 *K-means Algorithm*

The k-means algorithm was used because the study was conducted by clustering candidate areas for installing image dispenser.

The k-means clustering algorithm [2] is an algorithm that groups given data into k clusters, and is generally used because of its simple structure and wide application range.

This algorithm minimizes the average Euclidean distance between the data and the center of the cluster to which the data belongs.

Equation 1 is an expression that expresses the goal of this algorithm.

Assuming that μ_i is the center point of the set S_i , we aim to find the set S with the minimum value.

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

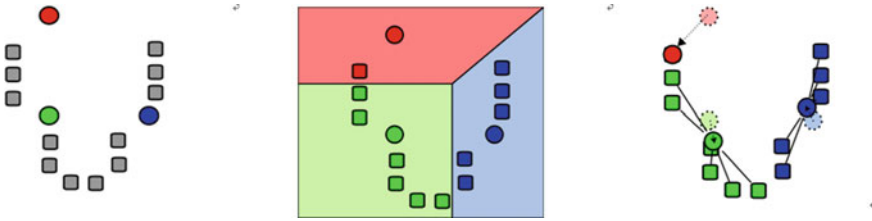


Fig. 1 K-means execution process

Figure 1 shows the execution process of k-means. The execution process of k-means starts by dividing each point into k sets arbitrarily and finding the center of gravity of the set. Each data is grouped based on the nearest average value, and the average value is readjusted based on the center point. By repeating the previous operation, the set of data converges to a state in which the set of data does not change or the center of gravity does not change [3].

2.2 Elbow Method

In the k-means algorithm, the Elbow method was referred to determine the optimal k that does not overfit.

The Elbow method is a heuristic method that selects the appropriate number of clusters in a clustering algorithm that receives the number of clusters k as input [4, 5].

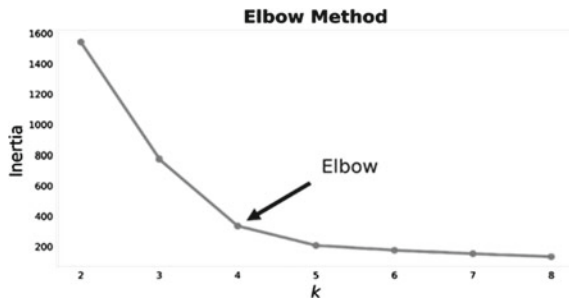
The Elbow method monitors the results by calculating the cluster cost while sequentially increasing the number of clusters.

The cluster cost is the average of the sum of the distances between the centroid and other data in the cluster in each cluster.

At this time, in the cost graph according to the k value, the k value at the point where the cost sharply bends is called elbow.

If the difference in cluster cost is not significant, the elbow may not be present and is not always clearly identifiable with the naked eye (Fig. 2).

Fig. 2 Example for Elbow method



3 Remote Image Dispenser System

3.1 Location Selection Analysis

The location analysis of the image dispenser is divided into a primary location analysis premised on installation only in front of a pharmacy and a secondary location analysis premised on installation outside the pharmacy. In the first analysis, pharmacies with a 3 km radius from the Holiday Keeper pharmacies that are open even on weekends and holidays and 1 km from the 24 h safe over-the-counter drug store are selected as candidates for the pharmacies with the image dispenser. As the distance between drugstores selected as candidates, the optimal cluster for each city is determined using the Elbow Method for each Sinan-gun, Hampyeong-gun, and Damyang-gun selected in the city-gun setting. After clustering pharmacies for each county, a score is given based on the number of elderly population and accessibility, and the pharmacy with the highest score for each cluster is selected as the final location. In the secondary location analysis, in addition to pharmacies, the candidate group was expanded to include health facilities such as public health centers, public health offices, and health clinics, as well as leisure facilities for the elderly such as senior citizens, village halls, and senior welfare centers, including places not limited to in front of pharmacies as candidates for installation. The rest of the process is the same as the first location analysis.

3.2 Local Setting Using Elbow Method

Data on financial independence, the elderly population ratio, the number of pharmacies per 100,000, and the number of cars per 100,000 were used, and the standard scale was used to standardize the data because of the large variance in the data.

PCA is performed to select two dimensions in which the cumulative value of variance explanatory power is 80% or more.

After that, to determine the number of clusters in the dataset, the optimal number of clusters was 3 as shown in <Fig. 4> with the Elbow method (Fig. 3).

As a result of K-Means Clustering based on this, cluster 0 is Mokpo-si, Yeosu-si, Suncheon-si, Naju-si, Gwangyang-si, Hwasun-gun, Muan-gun, and cluster 1 is Wando-gun, Jangseong-gun, Hampyeong-gun, Yeongam-gun, Gangjin-gun, Jindo-gun, Gurye-gun, Damyang-gun, Jangheung-gun, Shinan-gun, Cluster 2 was Gokseong-gun, Goheung-gun, Boseong-gun, Haenam-gun, and Yeonggwang-gun.

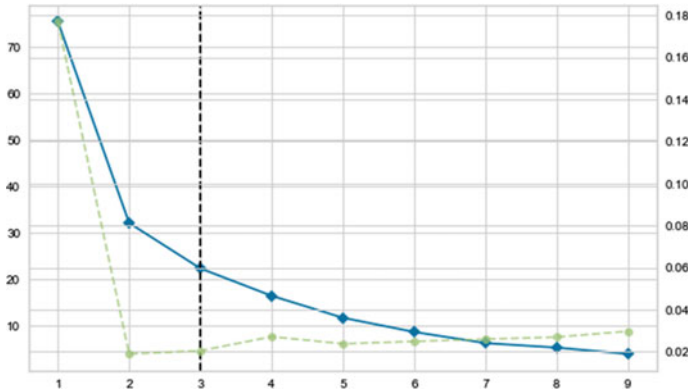


Fig. 3 Result of local setting using elbow method



Fig. 4 Visualization of cluster analysis in Damyang-gun, Jeollanam-do

3.3 Final Location Selection Result for Remote Video Dispensing Machine

Secondary location analysis includes places not limited to pharmacies as installation candidates. While only pharmacies were selected as candidate sites in the first location analysis, in the second round, the candidate group was expanded to include pharmacies, health facilities such as public health centers, health branch offices, and health clinics, and leisure facilities for the elderly such as senior citizens' centers, village halls, and senior welfare centers. As with the first location analysis, after selecting subjects outside the radius of 3 km for the holiday keeper and 1 km from the safety store sales office as candidates, the elbow method was used to derive the cluster, and the K-means algorithm was used for clustering. Gae, Hamyang-gun, and Sinan-gun were classified into four clusters.

4 Conclusion

According to the results of this study, it can be confirmed that when the area where the image dispenser can be installed is not limited to the pharmacy, the accessibility of the image dispenser to the target audience is increased.

When candidate sites in the island and mountainous regions corresponding to the outlier are removed and analyzed again, the sum of the elderly population within a 3 km radius of Sinan-gun is 601. Therefore, it is more than the first 521.33 people, so it is consistent with the assumption of the study. However, since the image dispenser machine selected in the previous process is more valuable in terms of public interest, it is included as it is in the final installation site.

Since this study was conducted before the government's special demonstration project was carried out, a methodology for site selection of the burn injector was presented based on limited data that did not include empirical data. By referring to the empirical data obtained from the special demonstration project to be carried out in the second half of this year, it will be possible to select the correct user base and region and to derive a clear analysis result. With this in mind, this study intends to present the basis of a technical methodology using QGIS based on the location conditions of the burner dispenser. It is expected that such a location methodology for burn dispensers will overcome regional medical disparities according to economic levels and expand consumer access to medicines.

References

1. Yu J (2014) Prospects and legal challenges of remote imager
2. McQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, pp 281–297
3. K-means clustering (2022). https://en.wikipedia.org/wiki/K-means_clustering
4. Ketchen DJ Jr, Shook CL (1996) The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manage J* 17(6):441–458
5. Park S (2021) Electric mobility demand prediction by region using k-means clustering

Passive Indoor Human Tracking Using Commodity Wi-Fi



Lingchao Guo, Jun Ma, and Yue Xu

Abstract Recently, with the rapid growth of the aging population, the public has attached substantial significance to in-home health care. However, most current solutions are either costly or complicated. Commodity Wi-Fi is a promising technology which is cost-effective and non-intrusive. Hence, in this paper, leveraging the Wi-Fi Channel State Information (CSI), we propose a Wi-Fi-based human tracking approach for in-home monitoring. We use the ratio of two adjacent antennas to denoise the CSI measurements and then perform Principal Component Analysis (PCA) to further separate the signal and noise from the subcarriers. In addition, we present a robust signal calibration algorithm to improve tracking accuracy by eliminating static components. Experiment results show that our method can achieve decimeter tracking accuracy, demonstrating the feasibility of the proposed method for in-home monitoring.

Keywords In-home monitoring · CSI · Wi-Fi · Passive human tracking

1 Introduction

Nowadays, better living conditions and healthcare system lead to the aged tendency of population. Global population aged over 60 was about 962 million in 2017, and it is projected to reach 2.1 billion by 2050 [1]. The phenomenon of empty nest is also ubiquitous that over 50% of the elders in China live alone [2]. In addition, many elders are in poor health condition and approximately 80% of them have at least one chronic disease [3]. Thus, the health problem of the elderly arouses widespread concerns. For those elders who live alone, it is difficult for their children to keep abreast of their health condition, so in-home monitoring for the elderly is of vital significance and necessity.

L. Guo (✉) · J. Ma · Y. Xu

China Academy of Information and Communications Technology, Beijing 100191, China

e-mail: guolingchao@caict.ac.cn

So far, most in-home monitoring systems are based on wearable devices or embedded sensors. Wearable devices are inconvenient and burdensome for long-time wearing. The camera-based solutions are susceptible to lighting conditions and raise privacy concerns. Compared with them, Wi-Fi-based human sensing is a promising technique for in-home monitoring which is non-intrusive and more robust against dark and obstacles. Utilizing Channel State Information (CSI) gained from off-the-shelf Wi-Fi cards [4], we can conduct human tracking after processing and analyzing CSI amplitudes and phases. Thus, caregivers can monitor the elderly to deal with possible emergencies even detected underlying diseases.

The main contributions of this paper are summarized as follows:

- We use the ratio of CSI of two adjacent antennas to denoise the original CSI and then perform Principal Component Analysis (PCA) to further separate the signals and noises.
- We propose a robust signal calibration algorithm to improve tracking accuracy by eliminating static components, and we realize human tracking by analyzing the relationship between the phase of CSI ratio and the dynamic path length.
- We implement a prototype system and carry out extensive experiments. Evaluation results verify that the proposed scheme can achieve indoor human tracking at the decimeter level.

2 Preliminaries

2.1 Overview of CSI

Wi-Fi signals suffer multi-path distortions propagating from a transmitter (T_x) to a receiver (R_x) by reflection, absorption, and scattering. CSI represents wireless channel through the samples of Channel Frequency Response (CFR) which is the superposition of signals from all paths at R_x in each Orthogonal Frequency Division Multiplexing (OFDM) subcarrier:

$$H(f, t) = \sum_{i=1}^L \alpha_i e^{-2\pi \frac{d_i(t)}{\lambda}} \quad (1)$$

where L is the number of paths, λ represents the wavelength, α_i is the amplitude attenuation, and $d_i(t)$ means the propagation length of path i . According to [5], the paths are composed of static paths and dynamic paths. Static paths which consist of Line of Sight (LoS) path and reflection paths from static objects do not change over time. Dynamic paths cover the paths reflected from the moving objects with a Doppler frequency shift (DFS) and variable lengths.

When conducting human tracking, the signals reflected from human body propagate over multiple paths, leading to multiple time-varying components. The direct

reflection signal which goes straight to the R_x after reflection is stronger than other signals, and thus, it is dominant among dynamic components. Based on this, considering only one reflection path does not lose generality. Then, CSI can be represented as follows:

$$H(f, t) = H_s(f, t) + H_d(f, t) = H_s(f, t) + \alpha(f, t)e^{-j2\pi\frac{d(t)}{\lambda}} \quad (2)$$

where $H_s(f, t)$ refers to the static component, $H_d(f, t)$ refers to dynamic component, $\alpha(f, t)$ is the complex attenuation, and $e^{-j2\pi\frac{d(t)}{\lambda}}$ means the phase shift of $H_d(f, t)$.

2.2 CSI Error Correction

The origin CSI cannot be used directly for human tracking since CSI samples extracted from commodity NICs [4] have rich distortions in amplitudes and phases in practice. Amplitude noises are consistent across different antennas at the same R_x . Therefore, we can eliminate the noises by applying the amplitude ratio between two adjacent antennas [6]. For phase noises, Carrier Frequency Offset (CFO) generated by asynchronous oscillators of the transceivers hardware results in phase offsets β :

$$H(f, t) = \left(H_s(f, t) + \alpha(f, t)e^{-j2\pi\frac{d(t)}{\lambda}} \right) e^{-j\beta}. \quad (3)$$

Since the receiving antennas on the same Wi-Fi NIC are time synchronous, the phase errors are the same between the receiving antennas for a particular subcarrier. Thus, we calculate phase difference between adjacent antennas to eliminate phase noises and we can obtain $H_{\text{ratio}}(f, t)$ through denoised amplitudes and phases, which is exactly the ratio of CSI between two adjacent antennas.

3 Methodology

3.1 Analysis of $H_{\text{ratio}}(f, t)$

Inspired by [6], we demonstrate $H_{\text{ratio}}(f, t)$ is feasible for human tracking. We regard the difference between the lengths of the two reflection paths on the two adjacent receiving antennas (i.e., $\Delta d = d_2(t) - d_1(t)$) as a constant and express $H_{\text{ratio}}(f, t)$ as:

$$H_{\text{ratio}}(f, t) = \frac{\left(H_{s,1} + \alpha_1 e^{-j2\pi\frac{d_1(t)}{\lambda}} \right) e^{-j\beta}}{\left(H_{s,2} + \alpha_2 e^{-j2\pi\frac{d_2(t)}{\lambda}} \right) e^{-j\beta}} = \frac{H_{s,1} + \alpha_1 e^{-j2\pi\frac{d_1(t)}{\lambda}}}{H_{s,2} + \alpha_2 e^{-j2\pi\frac{d_1(t)}{\lambda}} e^{-j2\pi\frac{\Delta d}{\lambda}}}, \quad (4)$$

where $e^{-j2\pi \frac{d_1(t)}{\lambda}}$ is a unit circle that rotates clockwise as $d_1(t)$ rises. $e^{-j2\pi \frac{d_1(t)}{\lambda}}$ can be $H_{\text{ratio}}(f, t)$ through translation, rotation, and inversion.

Due to the moving of human body, CSI changes on different subcarriers are correlated. In this paper, PCA is performed on 30 subcarriers of OFDM to retain the first principal component and further separate $H_{\text{ratio}}(f, t)$ and noises. The results of $H_{\text{ratio}}(f, t)$ amplitudes after PCA are shown in Fig. 1, and the results of $H_{\text{ratio}}(f, t)$ phases after PCA are shown in Fig. 2. It can be found that CSI amplitudes and phases after PCA are smoother than before. Similarly, $H_{\text{ratio}}(f, t)$ can be represented by static and dynamic paths:

$$\begin{aligned}
 H_{\text{ratio}}(f, t) &= H_{\text{ratio},s}(f, t) + H_{\text{ratio},d}(f, t) \\
 &= H_{\text{ratio},s}(f, t) + \alpha(f, t)e^{-j2\pi \frac{d(t)}{\lambda}}
 \end{aligned}
 \tag{5}$$

As shown in Fig. 3, when the human moves from point A to point B, the length of the dynamic path increases by Δd . In theory, $H_{\text{ratio}}(f, t)$ can be represented as point

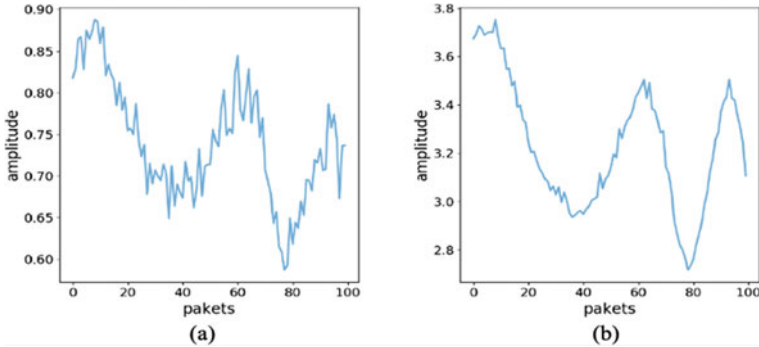


Fig. 1 Amplitude of $H_{\text{ratio}}(f, t)$ before and after PCA **a** Raw amplitude, **b** Amplitude after PCA

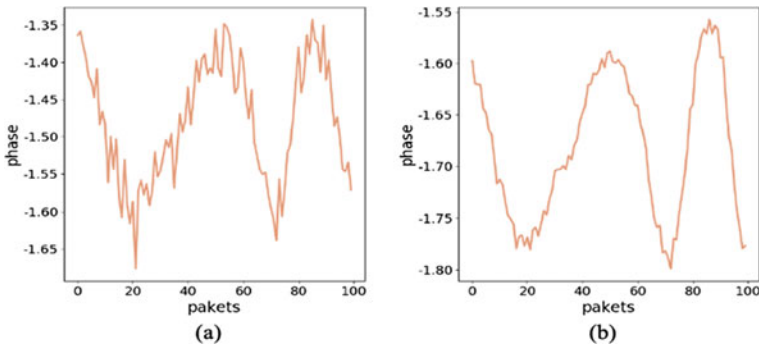


Fig. 2 Phase of $H_{\text{ratio}}(f, t)$ before and after PCA **a** Raw phase, **b** Phase after PCA

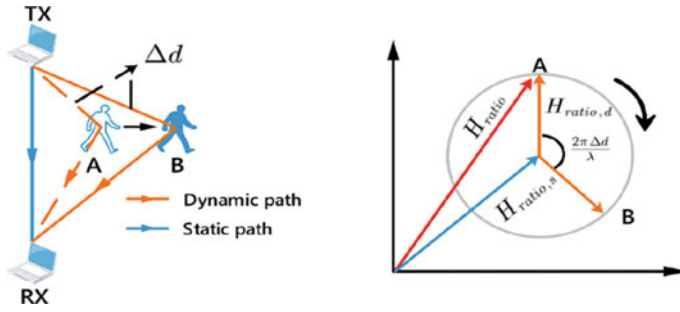


Fig. 3 Schematic diagram of human tracking model using $H_{ratio}(f, t)$

$A \left(\alpha(f, t) \cos\left(2\pi \frac{d(t)}{\lambda}\right), -\alpha(f, t) \sin\left(2\pi \frac{d(t)}{\lambda}\right) \right)$ on the complex plane and rotated clockwise to the point $B \left(\alpha(f, t) \cos\left(2\pi \frac{d(t)+\Delta d}{\lambda}\right), -\alpha(f, t) \sin\left(2\pi \frac{d(t)+\Delta d}{\lambda}\right) \right)$.

Specifically, for the case that the LoS path is not blocked, the strength of the static component is much stronger than that of the dynamic component. When the dynamic path length increases by $k\lambda$, $H_{ratio}(f, t)$ rotates clockwise by $2k\pi$, while when the dynamic path length decreases by $k\lambda$, $H_{ratio}(f, t)$ rotates counterclockwise $2k\pi$. As shown in Fig. 4a, b, the human moves from A to B along the vertical bisector of LoS. The raw CSI is very disorderly in the complex plane, but $H_{ratio}(f, t)$ is very regular and the trace rotates counterclockwise. Similarly, when the human moves from B to A, $H_{ratio}(f, t)$ rotates clockwise as shown in Fig. 5c, d. Thus, the length of the dynamic path can be obtained through the phase mapping of $H_{ratio}(f, t)$, which verifies that $H_{ratio}(f, t)$ can be used to realize human tracking.

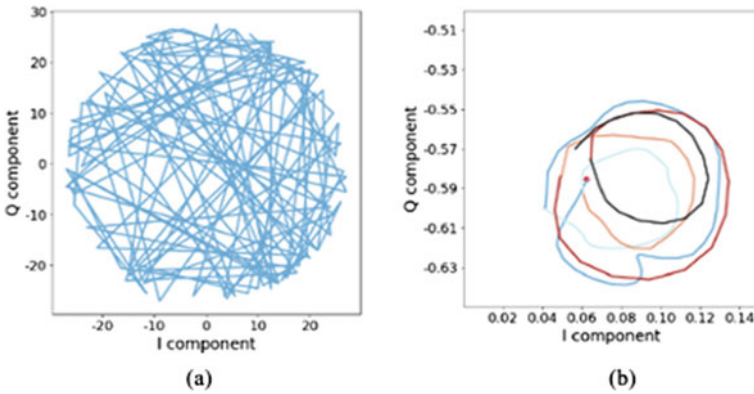


Fig. 4 Raw CSI and $H_{ratio}(f, t)$ when moving from A to B

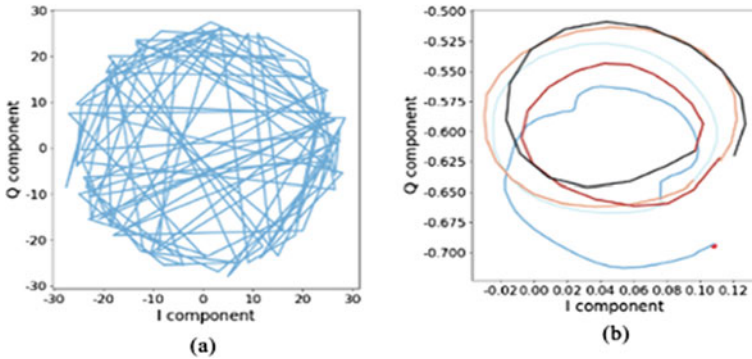


Fig. 5 Raw CSI and $H_{\text{ratio}}(f, t)$ when moving from B to A

3.2 Static Component Elimination

Theoretically, the static path does not change, and the length of the dynamic path changes continuously in inverse proportion to the radius of the circle. Therefore, the traces obtained in this paper should be concentric circles with different radii. As shown in Fig. 6a, the traces are not concentric circles and the center of the circle is far from the origin, deviating from the theoretical model. This is because different parts of the human body reflect signals in practice and moving body blocks the static path reflected by static objects.

We can eliminate the strong static components by subtracting the average value of the $H_{\text{ratio}}(f, t)$ to move the centers to the origin, and the processed trace results and their phases are marked with \mathcal{C} and \mathcal{A} , respectively. Due to the property of the arctangent function, there are multiple peaks and valleys in the phase, and in theory, the trace mapped between each two peaks corresponds to a complete circle. Directly unwrapped \mathcal{A} over time is shown in Fig. 6c, implying that the human moves close to the transceivers.

However, the tracking performance of unwrapped \mathcal{A} is not satisfactory since the slowly changed static paths exert influence on unwrapped \mathcal{A} which may cause some

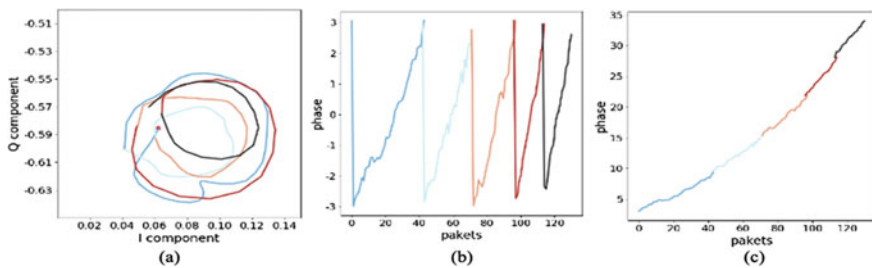


Fig. 6 $H_{\text{ratio}}(f, t)$, phase \mathcal{A} , and unwrapped phase **a** $H_{\text{ratio}}(f, t)$, **b** phase \mathcal{A} , **c** unwrapped phase

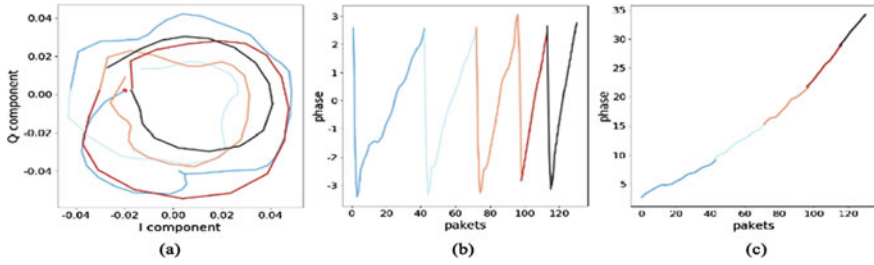


Fig. 7 Calibrated $H_{ratio}(f, t)$, phase \mathcal{A}' , and unwrapped phase **a** Calibrated $H_{ratio}(f, t)$, **b** Calibrated phase \mathcal{A}' , **c** Unwrapped phase

losses of radians on \mathcal{C} . These tiny shifts on CSI tracks are shifts of multiple circular centers to varying degrees in fact, so we deal with this issue by decomposing the entire trace. From Fig. 6b, we can note that there are several peaks on account of phase flips and each set of adjacent peaks corresponds to a complete circular spiral. Utilizing multi-peak search algorithm, we decompose \mathcal{C} into two arcs at the head and tail and some complete circles. We calibrate samples on each circle \mathcal{C}_n by subtracting the center of corresponding circle. According to proximity principle, we mark the two calibrated arcs with \mathcal{C}_1 and \mathcal{C}_n , respectively, and mark the circles with $[\mathcal{C}_2, \mathcal{C}_3, \dots \mathcal{C}_{n-1}]$. The calibrated trace is shown in Fig. 7a, and the calibrated phase \mathcal{A}' and unwrapped \mathcal{A}' are shown in Fig. 7b, c. We can note that the circle is more complete and most of the static components have been eliminated.

Thus, the relationship between dynamic path and $H_{ratio}(f, t)$ can be expressed as $\Delta d = -\frac{\mathcal{A}'}{2\pi} \lambda$.

As shown in Fig. 8a, when a pair of transceivers is placed, the length of the dynamic path does not change when the human moves along the tangent of the ellipse with focus at the T_x and R_x . The length of the reflected path changes only if the motion has a component in the normal direction of the ellipse. To detect signal changes along the normal direction of the ellipse, as shown in Fig. 8b, we place another R_x which is perpendicular to the first pair. For a pair of transceivers, with T_x and R_x as the foci, many ellipses, namely Fresnel region, can be construed [7], and different ellipses correspond to different reflection path lengths. Therefore, we use the properties of ellipse to solve the problem.

We set up a Cartesian coordinate system with T_x as the origin, T_x-R_{x1} as the x-axis, and T_x-R_{x2} as the y-axis to solve the coordinates of the trace. The LoS between $T_x - R_{x1}$ and $T_x - R_{x2}$ is L_1 and L_2 . Given the initial position coordinate, we can get the lengths and of the dynamic path at the initial position of R_{x1} and R_{x2} , respectively.

$$\begin{cases} \sqrt{x^2 + y^2} + \sqrt{(x - L_1)^2 + y^2} = p \\ \sqrt{x^2 + y^2} + \sqrt{x^2 + (y - L_2)^2} = q \end{cases} \quad (6)$$

Since the change of the path length can be estimated by calculating the radians of arcs in $H_{ratio}(f, t)$, we can obtain the change of the dynamic path received by R_{x1}

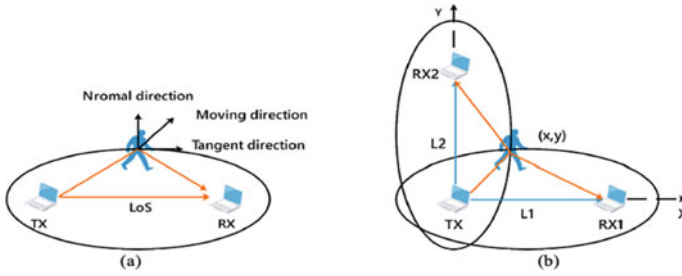


Fig. 8 Placement of the transceivers

relative to the initial position at time t , and likewise, we can gain $\Delta d_{RX2}(t)$. Then, Eq. (8) can be rewritten as:

$$\begin{cases} \sqrt{x(t)^2 + y(t)^2} + \sqrt{(x(t) - L_1)^2 + y(t)^2} = p_0 + \Delta d_{RX1}(t) \\ \sqrt{x(t)^2 + y(t)^2} + \sqrt{x(t)^2 + (y(t) - L_2)^2} = q_0 + \Delta d_{RX2}(t) \end{cases} \quad (7)$$

At this stage, we can get the coordinates of the human at different times and realize human tracking accordingly.

4 Implementation

Experiments are carried out on three mini-PCs equipped with Intel 5300 Wi-Fi NICs. Among them, the T_x is equipped with one omnidirectional transmitting antenna, and two R_x s are respectively equipped with three omnidirectional antennas, placed as Fig. 8a. We configure the NICs to run at 5.68 GHz center frequency with 40 MHz bandwidth and set the sampling frequency of CSI to 300 Hz. The sensing area is $6 \text{ m} \times 6 \text{ m}$, i.e., $L_1 = L_2 = 6 \text{ m}$.

We design three kinds of trajectory, namely, square, line, and arc, to evaluate the tracking accuracy, and their initial positions are set to be (1, 3), (1.6, 1.6), and (1.6, 1.6), respectively. The tracking results are shown in solid red line in Fig. 10. We also use Widar2.0 [8] for comparison, whose tracking results are shown in blue dashed line in Fig. 10.

5 Evaluation

The cylindrical error analysis method is used to calculate the tracking error: the human is assumed as a cylinder with 0.25 m radius, and the minimum distance

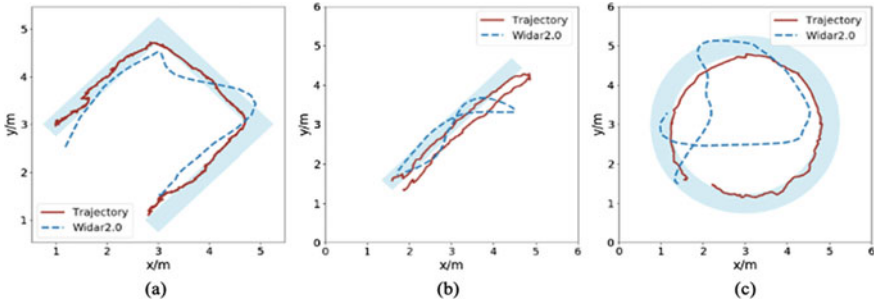


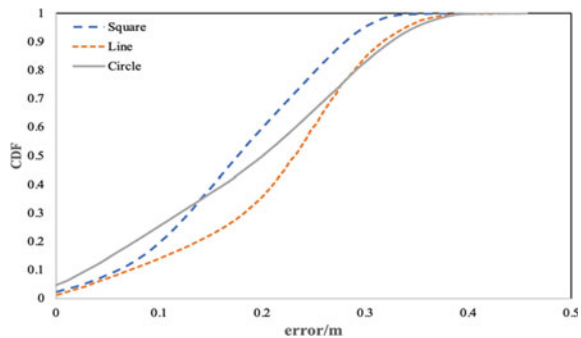
Fig. 10 Tracking results: a Square, b Line, c Arc

between the estimated position and the cylinder surface is considered as the tracking error of this position.

Through extensive experiments, the Cumulative Distribution Function (CDF) of the three trajectories at the given initial position is shown in Fig. 11. The median tracking errors of square, line, and arc is 0.172 m, 0.214 m, and 0.190 m, respectively, and the average errors are 0.183 m, 0.229 m, and 0.200 m, achieving decimeter tracking accuracy. The median tracking errors of Widar2.0 are 0.582 m, 0.532 m, and 0.827 m, respectively, and the average errors are 0.652 m, 0.631 m, and 0.916 m.

We find that tracking results of arc with Widar2.0 are significantly less effective. This is because the human is not a perfect cylinder nor is its surface smooth, and the MIMO-OFDM signal is reflected to the R_x through the human body. When the human body moves along an arc with a velocity of about 1 m/s, due to the changes of the human reflection points, Time of Flight (ToF) and Angle of Arrival (AoA) estimated by Space Alternating Generalized Expectation Maximization (SAGE) algorithm and Graph-based Path Matching (GPM) algorithm appear unsynchronized fluctuations, which lead to the overlaps on the tracking results. Note that this phenomenon is also seen in the case of sharp turns in square and straight tracks. From the above analysis, we can come to a conclusion that the robustness and tracking accuracy of our system are better than that of Widar2.0, achieving a relatively high tracking accuracy.

Fig. 11 CDF of the three kinds of trajectory



6 Conclusion

In this paper, we propose a passive indoor human tracking scheme based on commodity Wi-Fi. Extensive experiments carried out on the prototype system demonstrate that our scheme can achieve decimeter tracking accuracy and the tracking performance is better than Widar2.0.

Acknowledgements This work is supported by National Key R&D Program of China under Grant 2022YFE0101000.

References

1. Department of Economic and Social Affairs Population Division: World population ageing 2017 (2017) United Nations, New York
2. Fan B, Liu C (2003) A new predicting model for China's gross population. *Chinese J. Population Sci* 6(2):73–76
3. Key Facts: Mental health of older adults. <https://www.who.int/new-room/fact-sheets/detail/mental-health-ofolder-adults>. Last accessed 2017
4. Halperin D, Hu W, Sheth A et al (2011) Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Comput Commun Rev* 41(1):53–53
5. Li X, Zhang D, Lv Q et al (2017) IndoTrack: Device-free indoor human tracking with commodity Wi-Fi. *Proc ACM Interact Mobile Wearable Ubiquit Technol* 1:72:1–72:2
6. Zeng Y, Wu D, Xiong J et al (2019) FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proc ACM Interact Mobile Wearable Ubiquit Technol* 3(3):1–26
7. Wu D, Zhang D, Xu C et al (2017) Device-free WiFi human sensing: From pattern-based to model-based approaches. *IEEE Commun Mag* 55(10):91–97
8. Qian K, Wu C, Zhang Y et al (2018) Widar2.0: Passive human tracking with a single Wi-Fi link. In: Proceedings of the 16th annual international conference on mobile systems, applications, and services, pp 350–361

Research and Application on 5G-V2X-Based Enhanced Automated Valet Parking Key Technologies



Janghyun Baek, Yongcheol Ro, Junhyek Jang, Daekyo Shin, Sanghun Yoon,
and Soohyun Jang

Abstract Recently, a global carmakers are trying to have a stable Automated Valet Parking (AVP) solution to cope with a various parking environments. In spite of their efforts, there is a limit to completely autonomous parking by relying only on autonomous vehicle sensors in complex and diverse parking environments. But, the autonomous vehicle can effectively cope with an abnormal parking scenarios by using the parking environment information analyzed from the smart infrastructure installed in the parking lot, and finally, the AVP service area can further be expanded. In this paper, we defined some abnormal behavior scenarios that can occur in the parking lot and propose two ways to detect such misbehaviors by analyzing vehicle Controller Area Network (CAN) data and surveillance camera images.

Keywords Automated valet parking · Misbehavior · Anomaly detection

J. Baek · Y. Ro · J. Jang · D. Shin · S. Yoon · S. Jang (✉)
Korea Electronic Technology Institute (KETI), Seongnam 13488, Gyeonggi, Korea
e-mail: shjang@keti.re.kr

J. Baek
e-mail: baek0307@keti.re.kr

Y. Ro
e-mail: nyc0421@keti.re.kr

J. Jang
e-mail: junjang9327@keti.re.kr

D. Shin
e-mail: dukeshin@keti.re.kr

S. Yoon
e-mail: shyoon11@keti.re.kr

1 Introduction

Recently, as self-driving cars and road infrastructure technologies for autonomous driving has been advanced together, the last mile service of autonomous vehicle is also spotlight. To this end, a global car makers are trying to have a stable Automated Valet Parking (AVP) solution to cope with a various parking environments. Also, Korea is preparing for legal safety standards to vitalize the AVP-related services in parking lots.

However, there is a limit to completely autonomous parking by relying only on autonomous vehicle sensors in complex and diverse parking environments. In particular, when autonomous vehicle park inside a parking lot, there is a limitation to be perfectly perceptive of indoor parking environments due to poor Global Positioning System (GPS) signals and a short sensors' coverage. Therefore, it is difficult to deal with properly a various abnormal parking events that cause accident.

In addition, the autonomous driving process at crowded intersection does not recognize perfectly the traffic environment by itself, but research on efficient and safe autonomous driving technology that use together traffic information analyzed in smart infrastructure on intersection is in progress. Similarly, the autonomous vehicle can effectively cope with an abnormal parking scenarios by using the parking environment information analyzed from the smart infrastructure installed in the parking lot, and finally, the AVP service area can further be expanded.

On the other hand, the parking guidance information should be applied in real time to the AVP process. For this purpose, the transmission delay of parking convenience information can be minimized by using 5G-V2X (5 Generation-Vehicle to Everything) technology. Also, the precision and reliability of the autonomous vehicle's sensor are very important in the AVP process. Therefore, it is necessary to solve also the problem of external cyber-security through the V2X as well as sensor-level malfunction-inducing attacks.

Therefore, this paper proposed the method of detecting an intrusion that cause autonomous vehicle's malfunction and defines the misbehavior in the autonomous parking process. Also, by analyzing these misbehaviors in smart infrastructure, we developed finally a safe autonomous parking service that can provide feedback to vehicles.

2 Definition of Anomalies in Process of AVP

An autonomous vehicle uses several sensors such as camera, Laser Imaging Detection And Ranging (LIDAR) and GPS to self-drive usually in an outdoor environment. Inside a parking lot environment, aforementioned sensors have difficulties in distinguishing drivable areas and estimating the vehicle's location due to poor GPS signals, dim lights, and missing lanes. In a such setting, autonomous car may perform abnormally and lead to a fatal accident. To prevent potential accidents, we first define

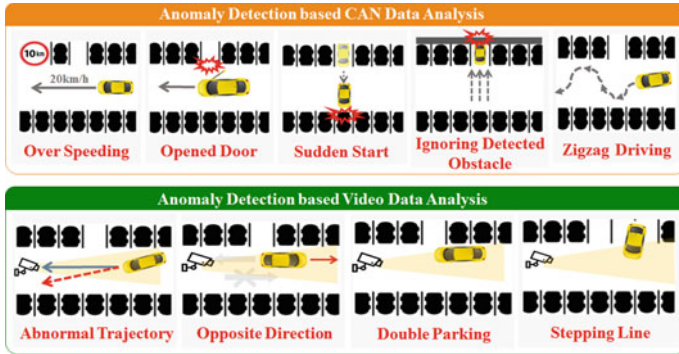


Fig. 1 Abnormal behavior scenario in parking lot

possible unwanted scenarios that autonomous cars can encounter inside a parking lot depicted in Fig. 1.

However, the parking lot is different from the outdoor autonomous driving environment (inside the building, underground, no traffic lights, and no lanes). Therefore, it is difficult to distinguish drivable areas and to estimate vehicle’s locations using GPS sensors inside the underground parking lot. In this situation, autonomous car may perform abnormal behaviors which cause terrible accident.

We defined anomaly scenarios of vehicles that can occur in parking lots. There are five scenarios can be detected based on in-vehicle CAN data and four scenarios that can be detected based on video data.

3 5G-V2X-Based Anomaly Detection Techniques

To detect abnormal scenarios defined in the second chapter, we approach it by using CAN data analysis and video data analysis as shown in Fig. 2. And, we operate detection analysis and data communication using a RSU (Road Side Unit) Edge platform as shown in Fig. 3. The platform system is divided into five processes, and each process performs the following tasks.

Image Analysis, by analyzing the images received from the surveillance cameras, we can infer the location of pedestrians and vehicles inside the parking lot, and determine available parking spaces [1]. Using the inferred information, we further detect abnormal behaviors such as abnormal driving trajectory and incorrect parking of an autonomous vehicle.

Vehicle Re-ID, we additionally detect the license plate number of cars. Using inferred license plate number and location of cars, we perform a Re-ID(Re-Identification) [2] algorithm to track cars in multiple surveillance cameras.

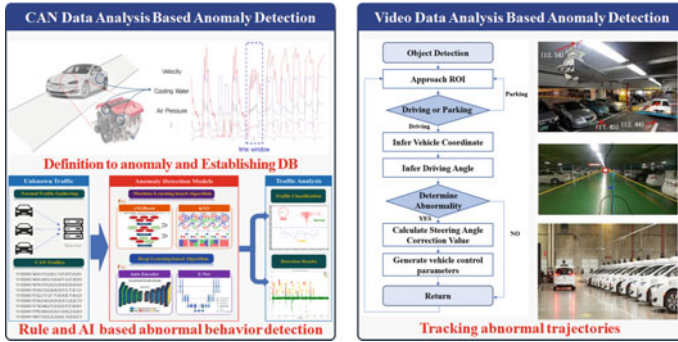


Fig. 2 CAN and video data analysis-based anomaly detection flow

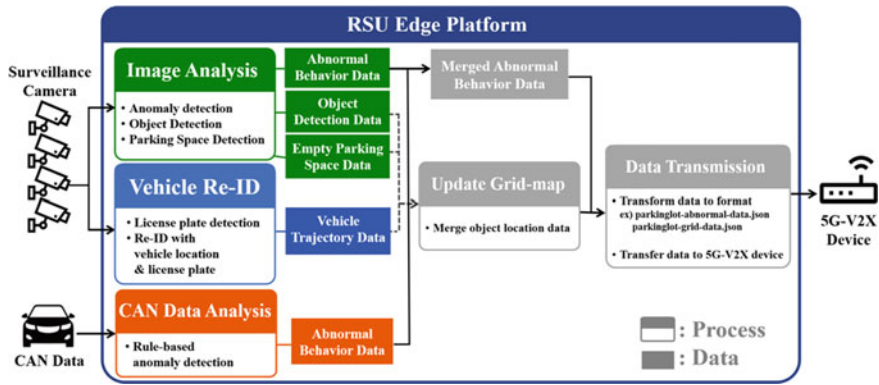


Fig. 3 RSU edge platform system flow for anomaly detection

CAN Data Analysis, after receiving CAN traffic from a vehicle, we analyze CAN traffic to classify abnormal actions using a rule-based algorithm.

Update Grid map, using the results from the above analysis processes, we update a grid map [3] that can be used to monitor the entire state of the parking lot.

Data Transmission, analyzed above results are converted into a JSON format file, and are transmitted to the 5G-V2X devices.

3.1 Method for Detecting Abnormal AVP Events Using Surveillance Image

The workflow to detect vehicle’s anomalistic behaviors using surveillance cameras is summarized in Fig. 4, and can be broken down into three steps [4, 5].

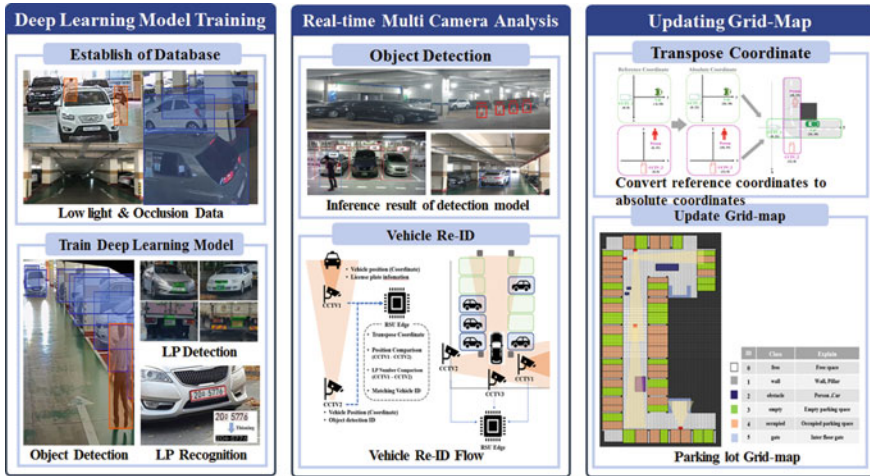


Fig. 4 Vehicle abnormal monitoring in surveillance camera

Train AI Model, in this paper, we trained three models, each performing tasks on object detection, license plate detection, and license plate recognition.

Analyze Multi Camera Data in Real time, we inference images from multiple cameras in real-time, and extract each vehicle’s location and license plate number. Using the inferred information, we track the cars inside the parking lot. In between cameras, cars are continuously tracked through Re-ID algorithm.

Update Grid map, grid map is updated based on the transformed location of cars to provide real-time monitoring of parking lot.

3.2 Method for AVP Anomaly Detection Using Sensor Data

Abnormal sensor data can cause the autonomous cars to misbehave, and lead to a severe accident. We develop algorithms to detect unexpected sensor data in AVP vehicle that can lead to undesired situations defined in the second chapter. Specifically, we detect both incorrectly measured and intruded sensor data. The overall workflow is shown in Fig. 5.

Faulty Sensor Data Detection, to detect abnormal behaviors defined in the second chapter, we utilize CAN sensor data collected from an autonomous vehicle, AVANTE CN7 model [6]. Using the analyzed result of CAN data based on (Database for CAN (DBC) [7], we detect abnormal sensor values using a rule-based algorithm as depicted in Fig. 5.

Intrusion Detection, The common types of intrusion in V2X systems are flooding, fuzzing, spoofing, and replay [8]. In recent studies, intrusion detection is mainly

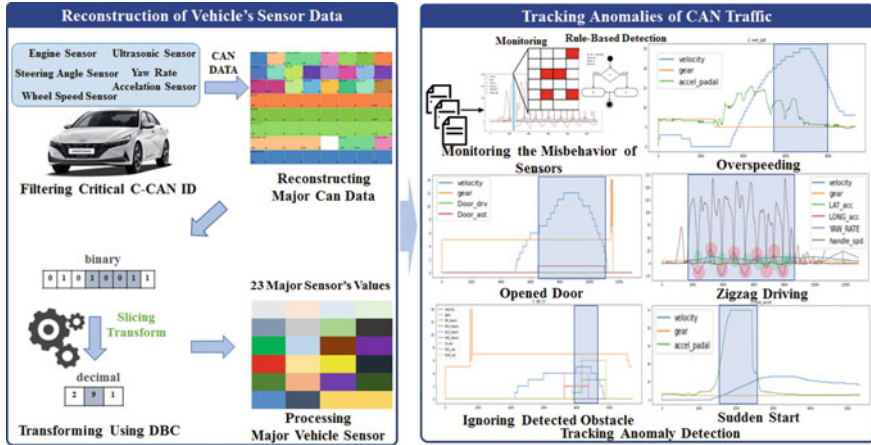


Fig. 5 CAN data analysis flow based on anomaly detection

performed using un-supervised learning methods [9]. However, these methods have problems with overfitting and difficulties processing in real time. To overcome such problems, we use a hybrid algorithm [10] that our team previously developed. In this paper, the algorithm shown in Fig. 6 is evaluated on the Cyber Security Challenge 2020 dataset [11]. As a result, we achieved more than 99% of accuracy [10].

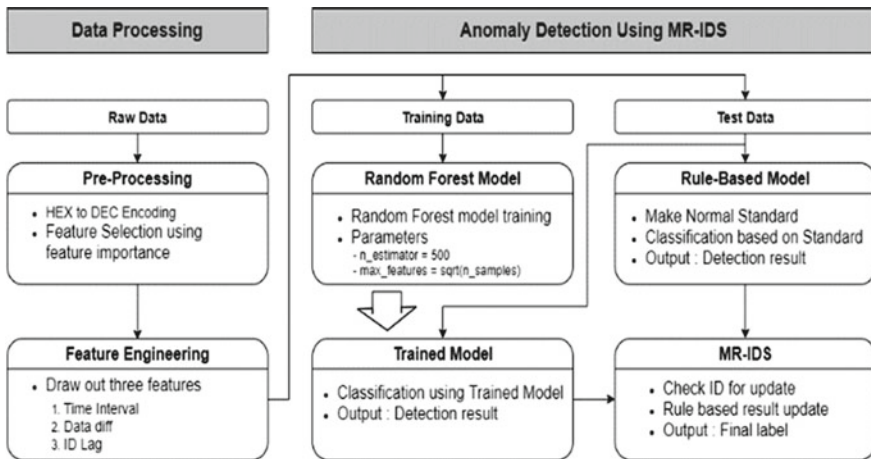


Fig. 6 Hybrid IDS workflow [11]

4 Conclusion

This paper proposed the method of detecting an intrusion that cause autonomous vehicle's malfunction and defines the misbehavior in the autonomous parking process. Also, by analyzing these misbehaviors in smart infrastructure, we developed a safe autonomous parking service that can provide feedback to vehicles. The proposed smart infrastructure based AVP system can be a good a solution for autonomous vehicle required for Last-mile service.

Acknowledgements This work was supported under the framework of international cooperation program managed by National Research Foundation of Korea (NRF-2021K1A3A1A61003229).

References

1. Ma Y, Liu Y, Shao S, Zhao J, Tang J (2022) Review of research on vision-based parking space detection method. *Int J Web Serv Res* 19:1–25
2. Peri N, Khorramshahi P, Rambhatla SS, Shenoy V, Rawat S, Chen J-C, Chellappa R (2020) Towards real-time systems for vehicle re-identification, multi-camera tracking, and anomaly detection. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), pp 622–623
3. Feng D, Haase-Schutz C, Rosenbaum L, Hertlein H, Glaser C, Timm F, Wiesbeck W, Dietmayer K (2021) Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges. *IEEE Trans Intell Transp Syst* 22:1341–1360
4. Absolute coordinates. http://wiki.gis.com/wiki/index.php/Absolute_coordinates. Last accessed 17 Oct 2022
5. Luis CE, Vukosavljev M, Schoellig AP (2020) Online trajectory generation with distributed model predictive control for multi-robot motion planning. *IEEE Rob Autom Lett* 5:604–611
6. Hyundai AVANTE CN7 model. <https://www.hyundai.com/kr/ko/e/vehicles/avante/intro>. Last accessed 19 Oct 2022
7. Open DBC homepage. <https://github.com/commaai/opendbc>. Last accessed 17 Oct 2022
8. Han ML, Kwak BI, Kim HK (2018) Anomaly intrusion detection method for vehicular networks based on survival analysis. *Veh Commun* 14:52–63
9. Seo E, Song HM, Kim HK (2018) Gids: Gan based intrusion detection system for in-vehicle network. In: 2018 16th Annual conference on privacy, security and trust (PST)
10. Kang DM, Yoon SH, Shin DK, Yoon Y, Kim HM, Jang SH (2021) A study on attack pattern generation and hybrid MR-ids for in-vehicle network. In: 2021 International conference on artificial intelligence in information and communication (ICAIC)
11. Car hacking: attack & defense challenge 2020 Homepage. <https://ocslab.hksecurity.net/Datasets/carchallenge2020>. Last accessed 16 Oct 2022

Sensitivity Analysis and Multi-optimization on Building Performance in Tube-Style Dwellings in the Rural Area of Northern Zhejiang Province



Li Wenxuan, Yin Zilu, Fan Bingxin, Wang Meiyuan, and Long Jiang

Abstract Due to the rapid development of rural tourism in China, more and more ordinary rural buildings renovate into rural tourism buildings. Taking the typical tube-style dwellings in the rural area of northern Zhejiang Province as a case, the factors affecting the building performance and the optimal renovation strategies are systematically analyzed and obtained. First, field survey is carried out and reveals that the electricity consumption of rural tourism buildings is large and the requirement of indoor thermal comfort is high. Then, the benchmark model is built and validated, and the parameters from three aspects: building, personnel use characteristics, lighting, and equipment are set for further simulation. Second, standardized regression coefficient algorithm, a global sensitivity analysis method is used to analyze the sensitivity of each factor, and the key parameters affecting building performance are identified. Third, multi-optimization simulation for cool electricity consumption and annual discomfort hours is carried out. Finally, 22 Pareto front solution sets are generated, and the optimal renovation solutions are sorted.

Keywords Tube-style dwelling · Rural tourism building · Sensitivity analysis · Multi-optimization · Energy consumption

1 Introduction

With the rapid development of rural tourism (RT) in China, more and more ordinary rural buildings (ORBs) have been turned into rural tourism buildings (RTBs) to increase farmers' income by providing tourism dwellings, catering services, leisure, and entertainment. It is also an effective way to protect and reuse the traditional dwellings [1]. However, RTBs are not only used for farmers' self-occupation, but

L. Wenxuan · Y. Zilu · F. Bingxin · W. Meiyuan (✉) · L. Jiang
Zhejiang A&F University, Hangzhou 311300, China
e-mail: wangmeiyuan0606@163.com

L. Wenxuan
e-mail: liwenxuanpog@163.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_47

355

also provide accommodation and recreational activities for tourists, which require higher indoor comforts [2], but also lead to higher building energy consumption than ORBs [3].

In recent years, many studies focus on the design practice or theoretical research on RTBs design. Zhu et al. studied the shape-related variables of RTBs in north China through multi-objective optimization [2]. However, few studies have systematically analyzed the influence factors and optimal strategies of RTBs renovated from ORBs.

Zhejiang is famous for RT, and the development of RTBs is at the forefront in China. In northern Zhejiang Province, the tube-style dwelling (TD) is a unique type of rural residential buildings formed during the period of “Emulating Da-Zhai on Agriculture Campaign”, which is the historical witness of the rural reform at the beginning of the founding of the People’s Republic of China. Taking TD in Zhanguwan Village of Deqing County as an example, scientifically and systematically study on the influence factors of building performance will provide optimal solutions and effective strategies for future RTBs renovation design.

2 Field Survey and Model Establishment

2.1 Field Survey on EC of RTBs

Due to the COVID-19 pandemic, RT business is greatly impacted. Therefore, the monthly electricity consumption (EC) of 30 RTBs before 2019 is collected and analyzed. Figure 1a shows the peak-valley fluctuation trend of EC is obvious. According to field survey, RT business mainly operate from April to November. Among them, July and August are the peak periods of EC, which are also the hottest months and the peak season of RT in a year. The EC at the peak is $5.49 \text{ kW}\cdot\text{h}/\text{m}^2$, about 4.8 times higher than that at the low peak in January ($1.43 \text{ kW}\cdot\text{h}/\text{m}^2$). Additionally, it also can be seen that the EC of different RTBs varies significantly. This is due to the combined effects of differences in outdoor climate conditions, architectural features, passenger flow, energy management, and energy-saving measures between different RTBs.

Description of TDs. Most of the existing TDs in Zhanguwan Village are two-story brick concrete structures, with the width about 3.6 m and length more than 20 m. As shown in Fig. 2a, the interior space is divided into three parts by two courtyards, and connected by the corridor inside. Each TD is an independent living unit as a household. In order to save land, most of them are built side by side.

Benchmark building. The benchmark model was obtained by field survey. The typical TD is orientated north and south, with the length of 33 m, covers an area of 252 m^2 , and the total building area of 504 m^2 . The building plane layouts after renovation are shown in Fig. 2b. The equipment, lighting, and personnel activity parameters of different zones are set according to the field survey results and standards

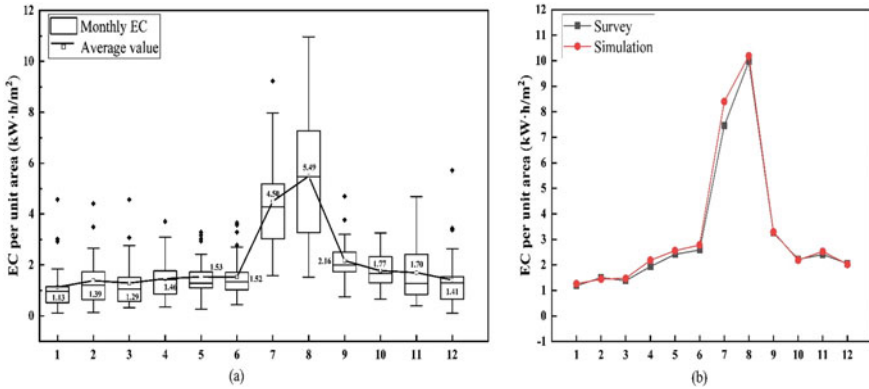


Fig. 1 Monthly EC of 30 RTBs surveyed (a) and model verification (b)

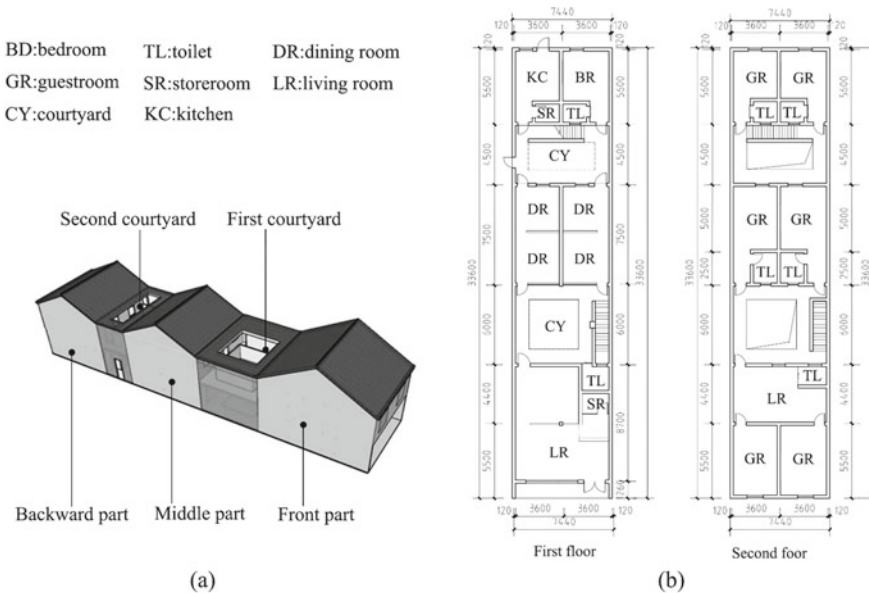


Fig. 2 3D simulation model (a) and Benchmark building plane layouts (b)

[4, 5], as shown in Table 1. The construction of envelopes are created based on field survey, and the materials and calculation methods are adopted according to standard [6]. Finally, the base values and variation range of each parameter for simulation are listed in Table 2.

Model verification and validation. Figure 1a reveals the EC of RTBs varies greatly. Finally, monthly EC data of five RTBs with larger passenger flow and relatively stable operation are set as the verification reference. After adjustments, the simulation

Table 1 Interior loads and temperature setting of main zones

Zone	People (P/room)	Metabolic rate (W/P)	Temperature (°C)		Light (W/m ²)	Equipment (W/m ²)
			Summer	Winter		
GR	3	100	26	18	7	15
BR	2	100	26	18	7	15
LV	10	110	26	18	7	15
DR	10	110	26	18	7	15
KC	3	180	/	/	7	70

Table 2 Input parameters for simulation

Parameter	Units	Base value	Range	Distribution
Wall U value	W/(m ² ·K)	2.2	0.4–2.6	Uniform
Windows U value	W/(m ² ·K)	6.4	1.5–3.5	Uniform
Windows Solar heat gain coefficient (SHGC)	—	0.8	0.2–0.8	Uniform
Roof U value	W/(m ² ·K)	6.0	0.5–6.0	Uniform
External door U value	W/(m ² ·K)	2.8	2.0–3.0	Uniform
Window to wall rate	%	30	15–65	Uniform
Air tightness	ach/h	1.0	0.3–1.5	Uniform
Lighting load	W/(m ²)	7	4–8	Uniform
Equipment load	W/(m ²)	15	3–15	Uniform
Occupancy rate of GR	P/room	3	1–4	Uniform
Cooling set-point	°C	18	16–22	Uniform
Heating set-point	°C	26	22–28	Uniform

results are compared with the survey data, as Fig. 1b shows. The results show that the maximum monthly error occurs in July, which is 12.7%, the annual error is 4.0%, and the mean square error coefficient is 9.0%, which means the three statistical indices are within the acceptable tolerances of Chinese standard [7], the ASHRAE Guideline [8] and the Federal Energy Management Program Guidelines [9].

3 Sensitivity Analysis

A global sensitivity analysis method—standardized regression coefficient algorithm—is adopted to identify the key parameters for further optimization. Latin hypercube sampling, a highly efficient sampling method, is applied. Twelve parameters about building, personnel use characteristics, lighting and equipment are set for simulation list in Table 2. After 600 runs, the standardized regression coefficients

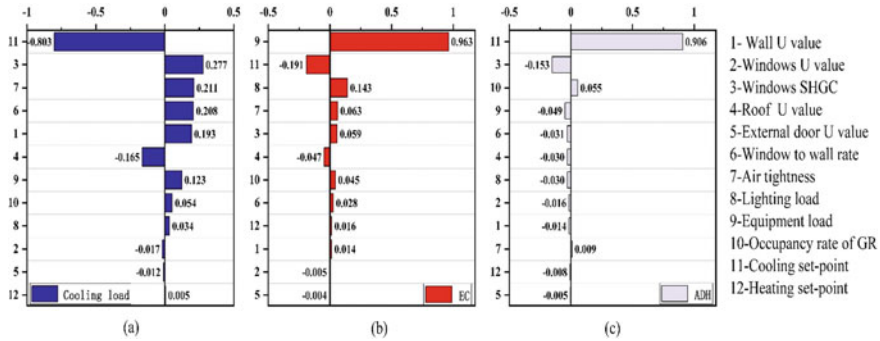


Fig. 3 SRCs of input parameters for cooling load (a), EC (b), and ADH (c)

(SRCs) of the parameters for cooling load, EC, and annual discomfort hours (ADHs) are obtained and shown in Fig. 3.

Figure 3a, c indicate that the sensitivity of personnel use characteristics to cooling load and EC is higher than others, especially the cooling set-point temperature, whose SRC value is the largest. Next is building parameters. Roof *U* value, wall *U* value, and windows *U* value have a greater influence than external door *U* value. Window-to-wall ratio and windows SHGC play more important role than window *U* value.

It also can be observed from Fig. 3a, c that lighting load and equipment load are not key parameters for cooling load and ADH. However, as shown in Fig. 3b, the SRCs of them are high, which means the EC of lighting and equipment in terminal energy consumption of RTBs is large, and should not be ignored.

4 Multi-objective Optimization

Multi-objective optimization is carried out by using the genetic algorithm of NSGA-II equipped with the optimization module in *Designbuilder* software. The field survey indicated EC of RTBs is mainly in summer. Therefore, cooling EC and ADH are selected as the optimization goals. Five key parameters are chosen for further simulation. The genetic iterations, individuals per generation, mutation rate, and crossover rate are 200, 20, 0.4, and 1.0, respectively. Finally, the Pareto optimal solution sets and the parameters are shown in Fig. 4.

Figure 4a shows the Pareto front and previous generations. Twenty-two Pareto optimal solutions were generated after simulation. The most comfortable solution in on the right, and the cooling EC is 3911.6 kWh, the ADH is 1987 h. The most energy-efficient result in on the left: the cooling EC is 2868.2 kWh and the ADH is 2037 h.

The parameters of the 22 Pareto optimal solution sets are shown in Fig. 4b. The roof *U* value and wall *U* value are 0.4 and 1.0 W/(m²·K), respectively. The other

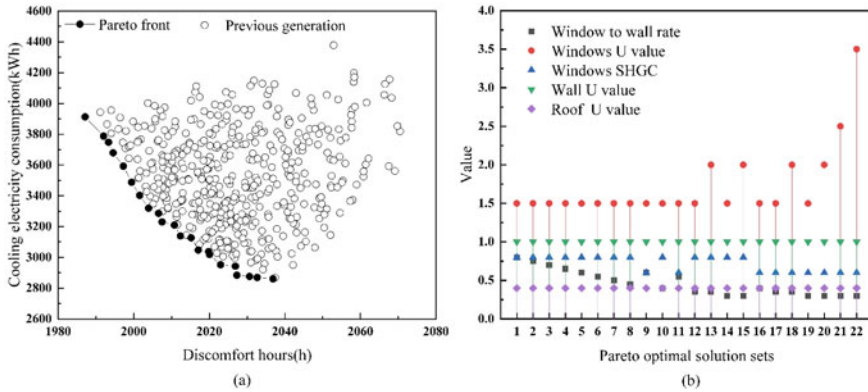


Fig. 4 Pareto front solution sets (a) and GC of every optimal solutions (b)

parameters are various. The range of window to wall rate is from 0.3 to 0.8, the larger the rate the higher the EC is. Six types windows were obtained in the optimal sets, which are composed of four different windows U values, and three different windows SHGC. The distribution of windows U values and SHGC shows that the lower U value and higher SHGC are benefit for comfort. Conversely, the higher U value and lower SHGC is helpful to reduce EC.

5 Conclusions

TD is a traditional residential building with Chinese characteristics, which is distributed in many parts of China. In this study, a renovation scheme and optimal strategies have been proposed. In order to understand the actual situation of EC status of RTBs and to verify simulation model, 30 RTBs in different regions of Zhejiang Province were investigated. Then benchmark model of typical TD after renovation has been built and verified according to the field survey results. A global sensitivity analysis method was used to systematically investigate the impacts of 12 influential parameters on EC and indoor thermal comfort, and five key factors were obtained for further optimization. Through Pareto-based multi-objective optimization of EC and ADH, 22 optimal results were obtained. The results can provide a reference for renovation of TDs into RTBs.

Acknowledgements This research work was sponsored by Zhejiang Provincial Natural Science Foundation of China (No.LQ19E080022), Scientific Research Development Fund Project of Zhejiang A&F University (No.2017FR043) and Student Research Training Program (NO.S202210341083).

References

1. Wang Z, Zheng Y, Chen C, Qian ZL (2016) The micro-activated organic renewal of a village of tube-style houses: a case study of Zhangluwan village in Deqing County, Zhejiang province. *Architectural J* 08:79–83
2. Zhu L, Wang BH, Sun Y (2020) Multi-objective optimization for energy consumption, daylighting and thermal comfort performance of rural tourism buildings in north China. *Build Environ* 176:106841
3. Shen SY, Wang H, Quan QH, Xu J (2019) Rurality and rural tourism development in China. *Tourism Manage Perspect* 30:98–106
4. Ministry of housing and urban-rural development of the people's republic of china, design standard for energy efficiency of rural residential buildings (2013) (GB/T50824-2013). China Architecture & Building Press, Beijing, 2013
5. Ministry of housing and urban-rural development of the people's republic of china. General code for energy efficiency and renewable energy application in buildings (2021) (GB 55015-2021). China Architecture & Building Press, Beijing, 2021
6. Ministry of housing and urban-rural development of the people's republic of china. Code for thermal design of civil building (2016) (GB 50176-2016). China Architecture & Building Press, Beijing, 2016
7. Ministry of housing and urban-rural development of the people's republic of china. Technical code for the retrofitting of public building on energy efficiency (2009) (JGJ 176-2009). China Architecture & Building Press, Beijing, 2009
8. American society of heating, refrigerating and air-conditioning engineers, Inc (ASHRAE) (2002) ASHRAE Guideline 14-2002: Measurement of Energy and Demand Savings, ASHRAE Standards Committee, 2002
9. U.S. Department of energy (2015) Federal energy management program (FEMP), M&V guidelines: measurement and verification for performance-based contracts, Version 4.0, United States, 2015. <https://www.osti.gov/servlets/purl/1239725>

Study on the Cooling Effect of Green Space in Shandong Peninsula Coastal Cities Based on Multi-source Remote Sensing Data



Yansu Qi, Xuefei Li, and Weijun Gao

Abstract As the main measure to regulate the heat island effect in cities, the influence mechanism of the green space spatial distribution on the cooling effect has important significance for the study of the thermal environment in coastal cities. This study takes the green space of seven coastal cities in Shandong Peninsula as the study area. Multi-source remote sensing data are used to quantitatively analyze the spatial distribution of green space and land surface temperature in coastal cities and explore the distribution pattern of the cooling effect of green space in coastal cities. The results show that the cooling effect of green space in coastal cities is influenced by spatial distribution. Linear regression results of NDVI and LST indicated that the LST gradually decreases with the increase of NDVI during the daytime. At night, with the increase of NDVI index, the LST changes to a lesser extent relative to the day, indicating that the temperature at night is in a more stable state. As the distance from the coastline increased, the cooling effect of the green space became stronger in daytime. On the contrary, at night, the seven dot lines show a different trend from the daytime; except for DY and QD, the slope becomes bigger as the distance from the coastline increases, which indicates that the farther the green area is from the coastline, the weaker the cooling effect is.

Keywords Green space · Coastal city · Remote sensing

Y. Qi · W. Gao (✉)

Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology, Qingdao 266033, China

e-mail: gaoweijun@me.com

Y. Qi · X. Li · W. Gao

School of Environment and Municipal Engineering, Qingdao University of Technology, Qingdao 266033, China

Y. Qi · W. Gao

Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu 808-0135, Japan

1 Introduction

With the gradual development of urban heat island effect research, the exploration of urban heat island effect mitigation measures has become the highlight of research [1]. The effects of greenspace on cooling can be measured using remote sensing at all scales from a greenspace patch to a city and beyond [2]. At present, research on the relationship between urban landscape pattern and land surface temperature (LST) have been studied, among which the most attention has been paid to the influence of landscape components on LST, and the conclusions are relatively clear [3–6]. Current studies on the cooling effect of urban green spaces are mainly based on two types of data. One is based on air temperature obtained from the field observations at set points [7]. Another is the LST inversion based on remote sensing thermal infrared band data [8]. It has been shown that the traditional in-field observation method is costly, heavy workload, and covers a small spatial scale [9]. Therefore, the LST data inversion by remote sensing has been well time-synchronized. The obtained LST can be effectively connected with the landscape pattern indices to describe the spatial morphology and layout of green areas, thus becoming the main method to study the correlation between landscape patterns and land surface temperature of green areas. Coastal cities have a high level of economic development in China, and with the economic development, the heat island effect has gradually become one of the major urban problems that need to be solved urgently. It is considered that increasing green space is one of the effective means to reduce the heat island effect [10], but less research has been done on this aspect in existing studies for coastal urban areas. The urban morphological specificity of coastal cities and the moderating effect of marine climate on the inland thermal environment are neglected [11]. And the effect of coastline distance on the cooling effect is also ignored.

This study takes coastal cities in Shandong Peninsula, China, as the study area and uses multi-source data to carry out urban thermal environment characteristics study in coastal areas. For the summer of 2020, the spatial evolution characteristics of the urban thermal environment were analyzed by combining remote sensing image inversion of urban land surface temperature. For the analysis method, the method of spatial analysis is proposed to take the coastline as the benchmark to set up a buffer zone to study the cooling effect of green spaces. The land surface temperature difference between daytime and nighttime is used to represent the cooling effect of green spaces, and the results of fitting NDVI to LST difference are used to present the influence of distance from the coastline on the cooling effect of green spaces.

2 Materials and Methods

2.1 Study Area

Located in northeastern China, as Fig. 1a shows, Shandong Peninsula is the largest peninsula in China and is surrounded by Bohai Sea and Yellow Sea. The coastline of Shandong Peninsula is distributed in seven cities, namely Binzhou (BZ), Dongying (DY), Weifang (WF), Yantai (YT), Weihai (WH), Qingdao (QD), and Rizhao (RZ). Their locations are shown in Fig. 1a. The climate of Shandong Peninsula belongs to the warm temperate monsoon climate type, with January is the coldest month and August is the hottest. In order to explore the cooling effect of green areas in coastal areas, the coastal area was divided into seven areas with a 5 km buffer distance from the coastline (Fig. 1) to explore the correlation between the cooling effect of green areas.

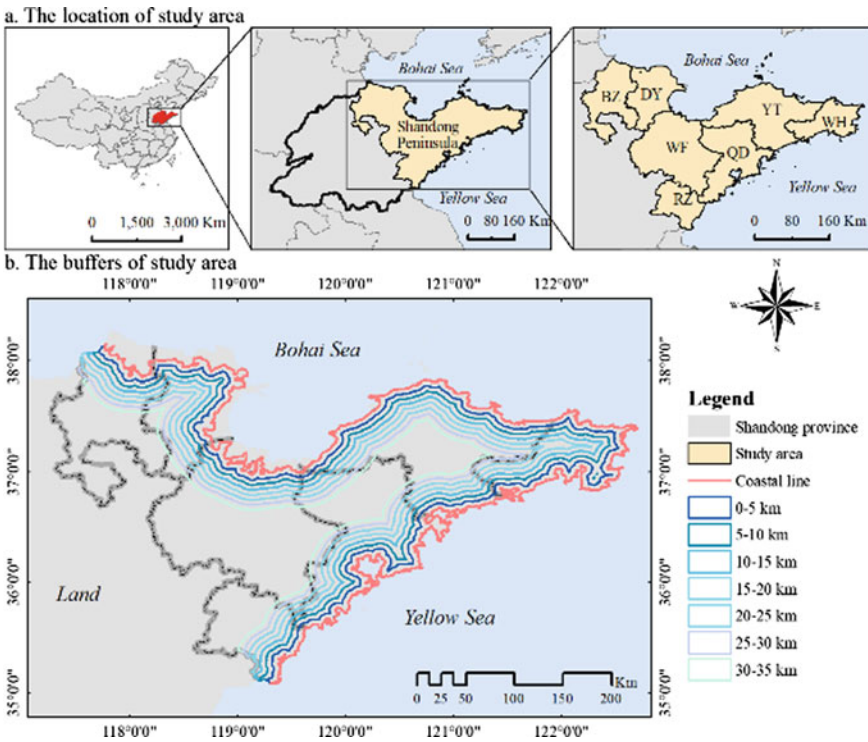


Fig. 1 Location of study area

2.2 Data Sources and Pre-processing

The dataset of remote sensing in this study included land cover products, Landsat 8, and the Moderate Resolution Imaging Spectroradiometer (MODIS) data. Land-cover products are used to extract the green space, and Landsat 8 satellite images are used to extract the coastal line. The application of each data is shown in the Fig. 2. Furthermore, the land surface temperature (LST) and normalized difference vegetation index (NDVI) of this study area were calculated based on the MODIS dataset. Details of these datasets are described in the following sections.

The GlobeLand30 datasets from 2020 (<http://www.globallandcover.com>) were used in this study. The data include 10 land-cover types such as artificial land and bare land, and cover the land within the range 80°S–80°N. Green space was used as the research target, so the range of green space (including the forest land, grassland, and urban green space) was extracted using the land cover categories in the GlobeLand 30 data. Zhang et al. [12] concluded that global 30 m impervious-surface dynamic dataset is an accurate and promising product and could provide vital support in monitoring regional or global urbanization as well as in related applications.

Landsat 8 satellite data can present the land image previously and is used to extract the coastal line. Information on Landsat8 data covering the study area is presented in Table 1. The Modified Normalized Difference Water Index (MNDWI) was used to extract water information from densely built areas [13]. The MNDWI is expressed as Eq., and the range in MNDWI is [− 1, 1].

$$MNDWI = (\rho_{green} - \rho_{mir}) / (\rho_{green} + \rho_{mir})$$

where ρ_{green} represents the reflectance of green band of Landsat OLI data, ρ_{mir} represents the reflectance of short wave infrared band of Landsat 8 OLI data [14]. According to MNDWI value, areas with MNDWI values greater than 0.9 are land and are assigned a value of 1 for each pixel, while areas less than 0.9 are water and are assigned a value of 0. Extract the “1” value from the binarized image and vectorize the area.

Fig. 2 Diagram of multi-source remote sensing data application

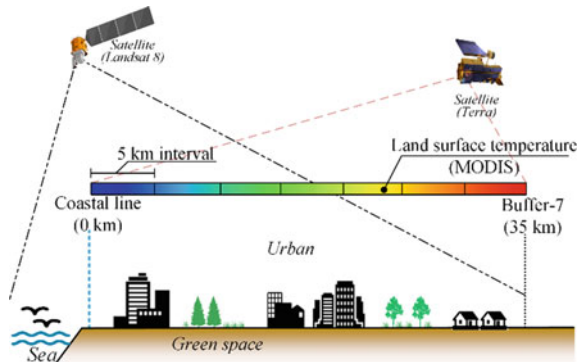


Table 1 Landsat 8 satellite data of study area

Landsat 8 Scene ID	Acquisition date	Cloud (%)
LC81190342019121LGN00	2019-05-01	0.92
LC81190352019121LGN00	2019-05-01	0.03
LC81200342019240LGN00	2019-08-28	0.47
LC81200352019240LGN00	2019-08-28	0.16
LC81210342020202LGN00	2020-07-20	3.82
LC81210352019183LGN00	2019-07-02	0.46
LC81210362020138LGN00	2020-05-17	3.48
LC81220342019158LGN00	2019-06-07	2.52
LC81220352019158LGN00	2019-06-07	3.74
LC81220362020241LGN00	2020-08-28	0.38
LC81230342021138LGN00	2021-05-18	0.14

MODIS is a key instrument aboard the Terra (originally known as EOS AM-1) and Aqua (originally known as EOS PM-1) satellites. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1–2 days, acquiring data in 36 spectral bands, or groups of wavelengths. The MDOIS Terra (MOD11A2) provided, the average surface temperature data in the past 8 days with a resolution of 1000 m during the day (10:30 a.m.) and night (10:30 p.m.) at the same time. This study selected the 1 km MOD11A2 (MODIS/Terra Land Surface Temperature/3-Band Emissivity 8-Day) as the basic data of LSTs calculation. The data row band number is H27V06. NDVI values are calculated based on 1 km MOD13Q1 (MODIS/Terra Vegetation Indices 16-Day). MODIS Reprojection Tool (MRT) is the software targeted for processing MODIS data, such as coordinate transformation, image mosaic, and map projections. The images processed by MRT were input into ArcGIS for data calculation.

3 Results

3.1 Land Surface Temperature Distribution

Based on the MODIS remote sensing data of the mean LST over 8 days in summer, the daytime land surface temperature (LST_{day}) and nighttime land surface temperature (LST_{night}) of the study area are shown in Fig. 3. The difference between daytime and nighttime temperatures in the study area indicates the cooling effect. According to Fig. 3a, the highest temperature was 36.71 °C, the lowest temperature was 21.75 °C, and the difference of LST_{day} was 14.96 °C at the same time, the higher temperature areas were distributed on the construction land, and the lower temperature areas were distributed on the green space. Figure 3b shows the distribution of LST_{night} , the highest temperature was 26.07 °C, the lowest temperature was 11.45 °C, and

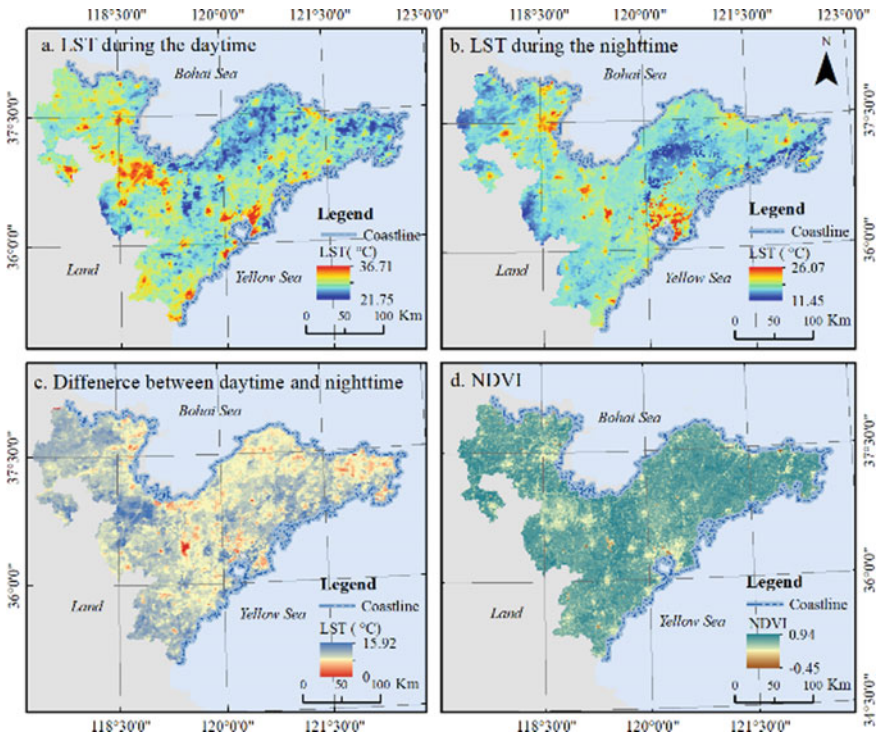


Fig. 3 Land surface temperature distribution of coastal cities in Shandong peninsula

the temperature difference at night was 14.62 °C. At the same time, the higher temperature was mainly distributed in water bodies and built-up areas of urban areas, and the areas with lower temperatures were distributed in green spaces. Figure 3c was the distribution of cooling intensity of LST from day to night. The spatial distribution of temperature differences in the green space of the study area is also different. The mean LST_{day} and LST_{night} are different for each buffer, and there is a decreasing trend with increasing distance, both during the day and at night.

3.2 Cooling Effect in Coastal Cities

In order to explore the effect of distance from the coastline on the cooling effect of vegetation, NDVI was linearly fitted to LST of daytime and nighttime for three buffer zones at distances of 5, 10, and 15 km. The effect of cooling was analyzed by observing the slope of the fitted NDVI to LST at the daytime and nighttime in different buffers (Table 2).

The slope indicates the vegetation cover of the forest, and the negative values of the slope in Table 2 showed that LST is affected by NDVI negatively. The slope

Table 2 Linear regression results between NDVI and LST of coastal cities

City	Distance from coastline (km)	Day		Night	
		R ²	Slope	R ²	Slope
Binzhou	5	0.11	- 1.23	0.37	- 4.18
	10	0.11	- 1.42	0.28	- 2.52
	15	0.29	- 2.36	0.11	- 2.14
Dongying	5	0.15	- 2.48	0.36	- 4.25
	10	0.23	- 3.66	0.37	- 4.18
	15	0.43	- 6.28	0.45	- 4.55
Weifang	5	0.22	- 4.21	0.38	- 2.92
	10	0.17	- 5.14	0.12	- 2.18
	15	0.24	- 8.91	0.06	- 1.36
Yantai	5	0.23	- 4.66	0.24	- 2.35
	10	0.25	- 5.77	0.20	- 2.27
	15	0.32	- 6.18	0.14	- 1.89
Weihai	5	0.09	- 2.75	0.07	- 2.73
	10	0.20	- 4.99	0.03	- 1.88
	15	0.14	- 5.22	0.0006	- 0.18
Qingdao	5	0.30	- 6.69	0.17	- 2.72
	10	0.52	- 9.59	0.21	- 4.07
	15	0.56	- 10.18	0.23	- 4.09
Rizhao	5	0.47	- 5.76	0.31	- 2.98
	10	0.57	- 7.26	0.28	- 3.07
	15	0.6	- 8.24	0.08	- 1.24

represents the degree of influence of the green space NDVI on LST. The cooling range decreases with the increase of NDVI, which indicates that the dense degree of green space has a certain effect on the regulation of the large temperature difference, and can effectively regulate the urban heat island effect. It can be seen from the linear regression results of NDVI and LST for the seven coastal cities that the LST gradually decreases with the increase of NDVI during the daytime. This is mainly because the tree canopy has a shading effect, which can effectively block solar radiation and thus reduce the surface temperature, and the denseness of the woodland increases with the increase of NDVI, making the temperature lower. At night, with the increase of NDVI index, the LST changes to a lesser extent relative to the daytime, indicating that the temperature at night is in a more stable state.

According to Fig. 4, it can be seen that during the daytime, the seven dotted lines show a decreasing trend, and the slope becomes smaller as the distance from the coastline increases, indicating that the farther the green area is from the coastline, the stronger the cooling effect. On the contrary, at night, the seven dot lines show a different trend from the daytime; the slope becomes bigger as the distance from

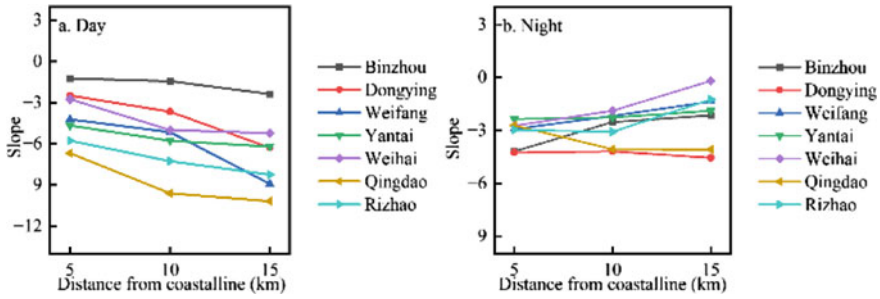


Fig. 4 Slope value of linear regression in buffer zones (5–10 km) of study area

the coastline increases, which indicates that the farther the green area is from the coastline, the weaker the cooling effect is.

4 Conclusion

This study investigated the characteristics of the cooling effect of green space in coastal areas by using multi-source remote sensing data. Firstly, the spatial distribution of green space in the study area was obtained by GlobeLand30 datasets from 2020. The coastline was extracted by Landsat 8 data through MDWI method to constitute the spatial division for green spaces. Secondly, the daytime and nighttime land surface temperature and NDVI distributions of study area in the summer of 2020 were obtained based on MODIS data. Finally, linear regression analysis of NDVI and land surface temperature of green space was applied by spatial analysis to obtain the scale of cooling effect (slope in the linear regression equation) of green space in different zones and compared. The slope of linear regression results represents the degree of influence of the green space NDVI on LST. LST is affected by NDVI negatively, and as NDVI increases, LST decreases. It indicates that the dense degree of green space has a certain effect on the regulation of the large temperature difference and can effectively regulate the urban heat island effect. The slope change curve in different buffers of day and night shows that the farther the green area is from the coastline, the stronger the cooling effect in the daytime. On the contrary, at night, the seven dot lines show a different trend from the daytime. The slope becomes bigger as the distance from the coastline increases which indicates that the farther the green area is from the coastline, the weaker the cooling effect is.

References

1. Gago EJ, Roldan J, Pacheco-Torres R, Ordóñez J (2013) The city and urban heat islands: a review of strategies to mitigate adverse effects. *Renew Sustain Energy Rev* 25:749–758
2. Kong F, Yin H, James P, Hutyrá LR, He HS (2014) Effects of spatial pattern of greenspace on urban cooling in a large metropolitan area of eastern China. *Landscape Urban Plan* 128:35–47
3. Li J, Song C, Cao L, Zhu F, Meng X, Wu J (2011) Impacts of landscape structure on surface urban heat islands: a case study of Shanghai China. *Remote Sens Environ* 115(12):3249–3263
4. Tran DX, Pla F, Latorre-Carmona P, Myint SW, Caetano M, Kieu HV (2017) Characterizing the relationship between land use land cover change and land surface temperature. *Isprs J Photogramm* 124:119–132
5. Guha S, Govil H, Gill N, Dey A (2020) Analytical study on the relationship between land surface temperature and land use/land cover indices. *Ann GIS* 26(2):201–216
6. Chen L, Wang X, Cai X, Yang C, Lu X (2022) Combined effects of artificial surface and urban blue-green space on land surface temperature in 28 major cities in China. *Remote Sens-Basel* 14(3):448
7. Lee S, Lee K, Jin W, Song H (2009) Effect of an urban park on air temperature differences in a central business district area. *Landsc Ecol Eng* 5(2):183–191
8. Sun R, Chen L (2017) Effects of green space dynamics on urban heat islands: mitigation and diversification. *Ecosyst Serv* 23:38–46
9. Wu J, Liu C, Wang H (2022) Analysis of Spatio-temporal patterns and related factors of thermal comfort in subtropical coastal cities based on local climate zones. *Build Environ* 207:108568
10. Li X, Zhou W, Ouyang Z, Xu W, Zheng H (2012) Spatial pattern of greenspace affects land surface temperature: evidence from the heavily urbanized Beijing metropolitan area China. *Landscape Ecol* 27(6):887–898
11. Cheng X, Nizamani MM, Jim CY, Balfour K, Da L, Qureshi S et al (2020) Using SPOT data and FRAGSTAS to analyze the relationship between plant diversity and green space landscape patterns in the tropical coastal City of Zhanjiang China. *Remote Sens-Basel* 12(21):3477
12. Zhang X, Liu L, Zhao T, Gao Y, Chen X, Mi J (2022) GISD30: global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series landsat imagery on the google earth engine platform. *Earth Syst Sci Data* 14(4):1831–1856
13. Xu H (2005) A study on information extraction of water body with the modified normalized difference water index (MNDWI). *J Remote Sens* 9(5):589–595
14. Min M, Lin C, Duan X, Jin Z, Zhang L (2019) Spatial distribution and driving force analysis of urban heat island effect based on raster data: a case study of the Nanjing metropolitan area China. *Sustain Cities Soc* 50:101637

Quantification Korean Skin Color Using Multi-color Space for Objective Skin Tone Classification



Seula Kye and Onseok Lee

Abstract Recently, due to the development of computer vision and artificial intelligence, research on skin image analysis has been actively conducted, and high classification performance was achieved. However, there are few studies that classify skin tone and evaluate performance. Skin tone is different due to various environmental factors, despite the same part of a person. Therefore, it is evaluated based on visual diagnosis or experience from dermatologists, which leads to subjective and inconsistent evaluation. So, it is necessary to establish a quantitative standard to ensure objective classification of skin tones. In this study, a skin tone quantitative value analysis system was developed using multi-color space to objectively classify the skin tone grade of Koreans. First, we constructed a dataset with images of skin captured using a digital camera, and then visually classified the dataset according to the skin tone. Subsequently, for the RGB, HSV, CIE $L^*a^*b^*$, YCbCr, and CMY color spaces from the Region of interest (ROI) of skin images classified as the same grade, the average of the color values for each channel extracted was calculated. As a result, the higher the skin tone grade, the lower the value of V of HSV, L^* of CIE $L^*a^*b^*$, and Y of YCbCr, which means brightness. Additionally, the a^* , b^* channel values increased in the CIE $L^*a^*b^*$ color space, which demonstrated that the skin becomes increasingly red and yellow with an increase in the skin tone grade. It can be used for skin diagnosis, and for the development of cosmetic technology.

Keywords Skin image · Skin tone · Multi-color space · Quantification · Classification

S. Kye · O. Lee (✉)

Department of Software Convergence, Graduate School, Soonchunhyang University, 22, Soonchunhyang-Ro, Asan City 31538, Chungcheongnam-Do, Republic of Korea
e-mail: leeos@sch.ac.kr

O. Lee

Department of Medical IT Engineering, College of Medical Sciences, Soonchunhyang University, 22, Soonchunhyang-Ro, Asan City 31538, Chungcheongnam-Do, Republic of Korea

1 Introduction

Skin is an essential organ that protects the body from external agents. It varies according to race, gender, age, and body parts, and it has been studied in various fields such as face recognition, tracking, and skin color division with the recent development of computer vision [1, 2]. In particular, external changes occur depending upon health conditions, increasing interest in skin color, which is used as an indicator to determine the progression of diseases, diagnose diseases, and verify treatment responses in the medical field [3]. However, the evaluation of skin is subjective because it depends upon the doctor's visual diagnosis and medical experience and is inconsistent because the evaluation results vary depending upon the specialists. It is difficult to objectively evaluate skin color, which is an important parameter in skin evaluation, because it appears differently even for the same person due to various environmental factors such as distribution of light sources and illuminance level [4]. Therefore, it is necessary to establish a quantified standard to objectively evaluate skin color.

In this study, a system was developed to measure the color quantitative values of each grade using multi-color space to establish criteria for the classification of Korean skin images according to skin tone. First, a dataset was constructed from images of skin captured using a digital camera, and the skin tone grades were classified by visual judgment. Each channel value of the multi-color space was extracted for the region of interest of the classified skin image, and the channel value average of the same grade images was calculated to obtain the color quantification value of each grade. The proposed system will be a fundamental study for the establishment of criteria for skin tone classification. Furthermore, it can be utilized for the development of cosmetic technologies, methods for skin diagnosis and disease assessment, and forensic science.

2 Methods

2.1 Overview of Proposed System

Figure 1 shows a series of steps for skin tone quantification. First, a dataset is created by photographing the human skin surface using a digital camera, and grades are classified by skin tone. For this, a skin image is cropped, and color quantitative value analysis is performed on the skin tone of the cropped image. Through this, skin tone can be quantitatively compared and changes can be observed.

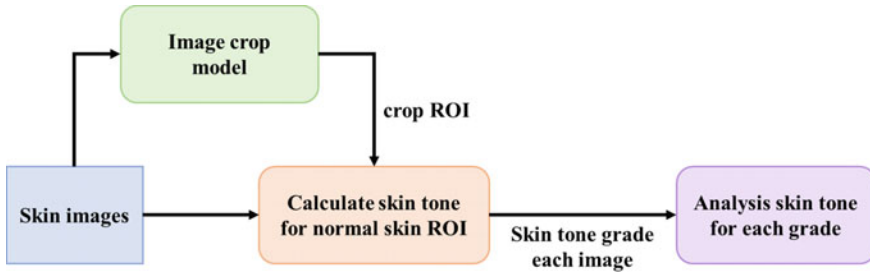


Fig. 1 Overview of multi-color space-based color quantification for skin tone grade classification

2.2 Acquisition of Skin Image

In this study, human skin surface was photographed using a digital single-lens reflex (DSLR) camera (Canon EOS 30D; Canon Inc., Tokyo, Japan) equipped with a macro lens (Tamron SP 60 mm f/2 Macro lens; Tamron Co., Ltd., Saitama, Japan) and a polarizing filter (Kenko 55 mm; Kenko Co. Ltd., Tokyo, Japan). An LED ring light (Aputure Amaran AHL-C60 Halo LED Ring Flash; Aputure Imaging Industries Co. Ltd., Shenzhen, China) was used to ensure uniform illuminance. It can be observed in Fig. 2 that a cylindrical light blocking module was installed in the ring light to block the interference caused due to ambient light. The length of the light blocking module was 12 cm, and a gap was absent between the module and skin surface because the module was in contact with the skin while capturing an image. Therefore, the skin was illuminated only by the LED ring light, and it was possible to capture an image while maintaining a certain distance without being disturbed by the ambient light. In addition, while photographing each part of a subject, a tape with a circular hole of diameter 5 mm was attached to the skin surface for capture the same part.

2.3 Skin Tone Grading

In this study, photos were taken without limiting body parts to collect various skin tones, and a dataset was constructed from the captured images. The dataset consisted of 500 normal skin digital images without skin diseases. Prior to skin color quantification, the dataset was graded by skin tone. The skin tone of a skin image was classified by visual judgment, and consistency was maintained by performing the classification twice by one person.

In this study, we implemented a multi-color space-based color quantitative value analysis system using Matlab 2021a (MathWorks Inc., MA, USA) to quantify the skin tone in skin images. Figure 3 shows the implemented analysis system. To quantify skin tone, each channel of RGB, HSV, CIE $L^*a^*b^*$, YCbCr, and CMY color space was used as a measurement index, which can be calculated simply as an image to ensure that skin tone can be classified practically. Additionally, as shown in Fig. 4,



Fig. 2 Composition of skin photography equipment

the skin tone is analyzed by cropping the circular hole part of the tape attached to the skin in the skin image.

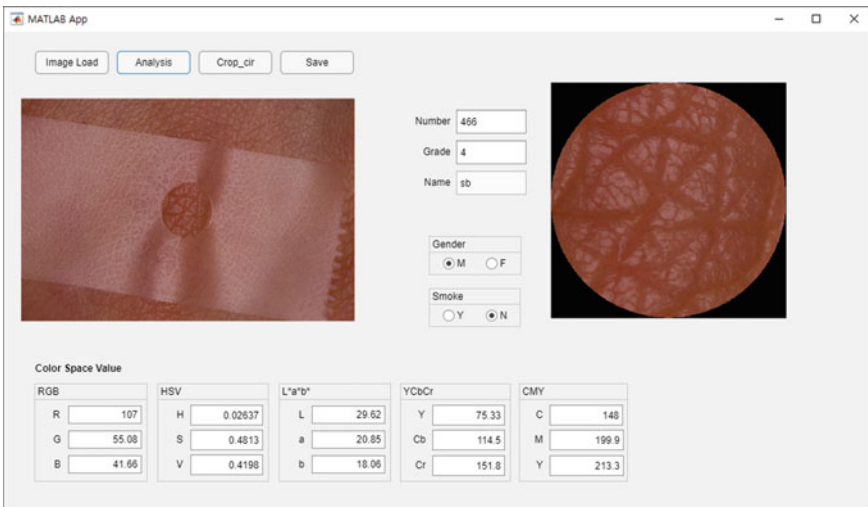


Fig. 3 Implementation of skin tone color quantification analysis system

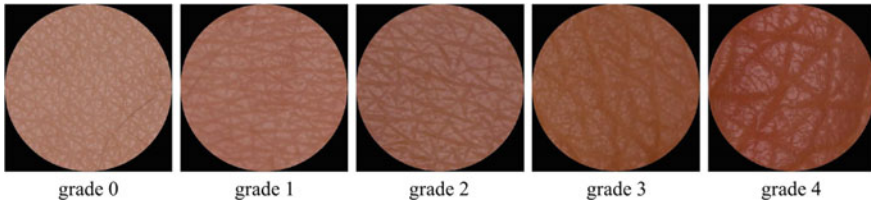


Fig. 4 Examples of each skin tone grade [5]

3 Results

Table 1 shows the color quantified values of each color space channel from grades 0 to 4 of skin tone. V of HSV, L^* of CIE $L^*a^*b^*$, and Y of YCbCr are channels that indicate brightness, and there is a decrease in the value with an increase in the grade, indicating that the skin tone becomes darker. As for the S channel of HSV, which means the degree of mixing with white, the value of S increases as the grade increases, which means that as the value increases, the skin color becomes closer to the primary color. Furthermore, in CIE $L^*a^*b^*$, a^* channel is close to green for negative and red for positive, and b^* channel is close to blue for negative and yellow for positive. The quantified CIE $L^*a^*b^*$ values show that the positive a^* , b^* channel values increase with an increase in grade, demonstrating that red and yellow skin becomes increasingly reddish and yellowish with an increase in grade. The results in Table 1 are consistent with the skin tone changes shown in Fig. 4. Therefore, the multi-color space-based color quantification method developed in this paper can be utilized as a quantitative standard.

4 Conclusions

This study confirmed the difference in skin color according to the skin tone grade of Koreans. A dataset was established by taking skin images using a digital camera, which is easy for the general public to obtain data, and a multi-color space-based skin tone quantification system was developed. Using the skin tone quantification system, each color quantification value was analyzed for the skin images classified by visual judgment. It was confirmed that it gradually became biased toward red and yellow. In the future, we will develop an automated skin tone quantification and classification system by securing various skin tone data and performing a classification model-based skin tone classification and performance evaluation.

Table 1 Average of quantitative values of color space channels by skin tone grade

Grade	RGB			HSV		
	R	G	B	H	S	V
0	123.5650	86.4237	76.8282	0.0290	0.2968	0.4846
1	123.6403	83.5495	73.5981	0.0284	0.3169	0.4849
2	122.3603	79.6545	67.2448	0.0294	0.3536	0.4798
3	118.4047	74.4482	63.3435	0.0265	0.3651	0.4643
4	113.0216	64.8579	53.0095	0.0825	0.4192	0.4432
Grade	CIE $L^*a^*b^*$			CMY		
	L^*	a^*	b^*	Y	C	M
0	40.0114	13.4914	11.2710	98.8232	118.3736	145.0300
1	39.2435	14.7781	12.1062	97.0763	117.7519	146.3538
2	37.9933	15.7325	14.1319	94.1611	116.2616	147.6936
3	36.1573	16.6632	13.8795	90.1379	116.7003	148.1207
4	32.9470	18.8399	15.6574	82.9089	115.7107	149.9986
Grade	YCbCr			CMY		
	Y	Cb	Cr	C	M	Y
0	98.8232	118.3736	145.0300	131.4350	168.5763	178.1718
1	97.0763	117.7519	146.3538	131.3597	171.4505	181.4019
2	94.1611	116.2616	147.6936	132.6397	175.3455	187.7552
3	90.1379	116.7003	148.1207	136.5953	180.5518	191.6565
4	82.9089	115.7107	149.9986	141.9784	190.1421	201.9905

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1010170) and the BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048).

References

1. Xiao K, Yates JM, Zardawi F, Sueeprasan S, Liao N, Gill L, Li C, Wuerger S (2017) Characterising the variations in ethnic skin colours: a new calibrated data base for human skin. *Skin Res Technol* 23(1):21–29
2. Rahmat RF, Chairunnisa T, Gunawan D, Sitompul OS (2016) Skin color segmentation using multi-color space threshold. In: 2016 3rd International conference on computer and information sciences. IEEE, Malaysia, pp. 391–396
3. Cho M, Lee DH, Doh EJ, Kim Y, Chung JH, Kim HC, Kim S (2017) Development and clinical validation of a novel photography-based skin erythema evaluation system: a comparison with the calculated consensus of dermatologists. *Int J Cosmet Sci* 39(4):426–434
4. Li CR, Tsai HY, Yang CC, Lin MY, Huang KC, Lin YH (2020) Quantifying the color changes in bruised skin using a color-calibrated imaging system. In: 2020 IEEE international symposium on medical measurements and applications. IEEE, Italy, pp 1–5
5. Kye S, LEE O (2022) Skin color classification of Koreans using clustering. *Skin Res Technol* <https://doi.org/10.1111/srt.13201>

Segmentation of UV Images Based on U-Net for Precisely Analyzing Cutaneous Pigmentation: A Pilot Study



Yejin Wee, Taeyeon Gil, and Onseok Lee

Abstract This study focuses on applying a suitable treatment method and precisely guiding a treatment area of a lesion for accurate diagnosis and severity assessment of cutaneous pigmentation. In particular, Melanoma, one of the types of pigmentation, is a skin cancer accompanying a very high mortality rate; therefore, it becomes important to apply accurate treatment to the lesions. However, visual inspection, mainly used in clinical practice, is unable to precisely diagnose lesion areas. Therefore, studies have been proposed to enable objective diagnosis and severity assessment through deep learning using lesion images. Thus, precise segmentation of the lesion area is required. The inspection using UV images shows the difference between the normal skin and lesion area more accurately than RGB images. However, when compared to the research studies on RGB images, the number of studies using UV images is relatively few. Therefore, we segmented lesions in UV images using a CNN model combining ResNet50 and U-Net after constructing a dataset that pre-processed and augmented images obtained using UV lamps and a filter. In this study, like melanoma, nevus exposed to the skin due to hypertrophy of melanocytes was targeted for segmentation prior to melanoma lesions. After verifying the performance of segmentation by comparing the segmentation prediction result and the ground truth, the proposed method obtained an accuracy of 0.9996 and a Jaccard Index of 0.7868. Therefore, the segmentation of UV images for cutaneous pigmentation was possible with this method.

Keywords Medical imaging · Ultraviolet imaging · Skin lesion · Pigmentation · Segmentation · Deep learning · U-Net

Y. Wee · T. Gil · O. Lee (✉)

Department of Software Convergence, Graduate School, Soonchunhyang University, 22, Soonchunhyang-Ro, Asan City, Chungcheongnam-Do, Republic of Korea
e-mail: leeos@sch.ac.kr

O. Lee

Department of Medical IT Engineering, College of Medical Sciences, Soonchunhyang University, 22, Soonchunhyang-Ro, Asan City, Chungcheongnam-Do, Republic of Korea

1 Introduction

Cutaneous pigmentation is an immune reaction to prevent ultraviolet (UV) light from penetrating the body. It shows symptoms such as melanocytes producing melanin that absorbs UV light or cells themselves increasing excessively. It is important to objectively and consistently detect the pigmentation lesions and quantitatively assess the severity to apply the appropriate treatment method to the area of pigmentation lesion accurately. Also, the precise segmentation of the boundary of the lesion that is unclear and uneven can be used as a guide for the area of the treatment to promote efficient and accurate treatment. In clinical practice, it is usually diagnosed by checking the distribution of pigmentation with the naked eye and confirming the incidence tendency in patients. However, with a visual inspection, it is difficult to diagnose and distinguish the lesion and objectively detect its distribution, so the application of the appropriate treatment to the precise lesion area may not be accurate. To overcome these limitations, biopsy, an invasive method is used to diagnose cutaneous pigmentation. This method identifies melanocytes in the dermal layer and a more accurate diagnosis can be obtained. However, since it is a method that directly contacts the skin, there is a risk of an injury or infection. Therefore, inspection methods using Wood's light, that is, UV light, are preferred as a non-invasive method for diagnosing pigmentation. The difference in pigmentation is emphasized by the absorption of UV light from the melanin-pigmented area, making it appear darker than the normal area. Therefore, this method has the advantage of determining the clinical boundary of the lesion and monitoring its changes [1]. However, numerous studies have been conducted on how to segment and diagnose pigmentation using RGB images [2, 3], while those using UV images were relatively few. Additionally, UV images have fuzzy characteristics in themselves, which makes lesion segmentation difficult. Consequently, overcoming these problems, a method is required to segment the lesion and the normal skin surrounding them, considering the boundaries of various pigmentation diseases.

In particular, among cutaneous pigmentation, melanoma is a skin malformation that accounts for 75% of skin cancer mortality among all skin cancers. Diagnosing melanoma early becomes very important because the mortality rate of people with a severe case is significantly higher than that with a mild case [4]. Melanoma is a skin cancer that originating in melanocytes, which can occur in normal skin or existing nevus, and one-third of melanomas occur in existing nevus.

Therefore, we aim to perform robust segmentation of lesions using UV images targeting cutaneous pigmentation. This study was conducted with a nevus, like melanoma, exposed to the skin due to the hyperplasia of melanocytes. In this study, we used a non-invasive method of UV light, and segmentation of lesions was performed by a U-Net-based CNN model that can perform robustly despite the fuzzy characteristics, which is the limit of UV images. In addition, a UV image dataset was constructed by pre-processing and augmenting the images taken in the experimental environment for UV imaging designed in this study, and more accurate segmentation will be achieved by focusing on skin imaging of local UV images.

2 Methods

2.1 Acquisition of UV Images

A 16-bit UV image with 1024 * 1024 resolution was taken with PIXIS 1024 air cooling CCD (Princeton Instruments, Inc., NJ, USA) camera equipped with a Macro lens (Tamron SP 60 mm f/2 DI II 1:1 Macro Lens for Canon EF, Canon Inc., Tokyo, JP). We investigated the skin with a square-shaped UV light produce by eight 8W BLB lamps (F8T5BLB, Sankyo Ultraviolet Co., Ltd., Seoul, ROK) with a peak of 368 nm and a wavelength of UVA 315-400 nm, which is the same UV wavelength band of the Wood's lamp used in the clinic for diagnosing cutaneous pigmentation. To minimize the confusion with other wavelength bands, a UV bandpass filter (BP365-D40, Midwest Optical Systems, Inc., Palatine, IL, USA) that allows only UV wavelength bands was installed on the lens, and images were acquired in a darkroom space manufactured using blackout cloth and iron sticks. A circular polarizer filter (Zeta EX CPL, Kenko Tokina Co., Ltd., Tokyo, JP) was used to obtain a clean image by reducing unnecessary reflected light from the skin. Figure 1 indicates the graph of the quantum efficiency of the CCD camera (a), the transmission of the UV filter (b), and the spectral energy distribution of the UV lamp (c). The environment of acquiring UV images in darkroom space manufactured are shown in Fig. 2.

After preheating the UV lamp for about 5 min, the skin was cleaned to reduce noise caused by makeup and oil, and acquiring UV images of nevus was carried out 15 min later. The CCD camera lens and UV lamp were placed on the same line, and the distance between the UV lamp and the subject was about 11 cm, and the focal length of the camera was 0.3 m. The UV images of a nevus on various regions of the body with various curves and skin colors were taken. A total of 51 UV images of lesions were taken from a total of 11 subjects (3 men and 8 women; age: 25.27 ± 2.34).

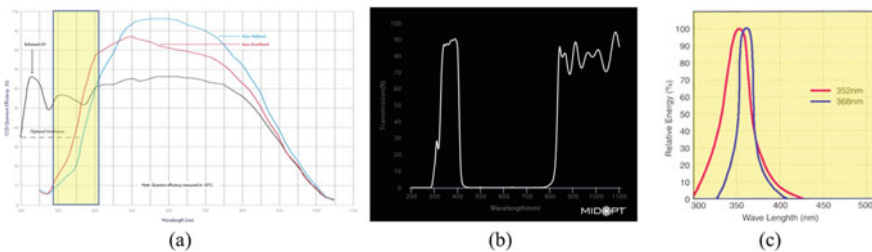


Fig. 1 This shows the spectral specificity graphs of products for acquiring UV images, **a** is Quantum Efficiency (QE) Curve of the CCD camera, **b** is Transmission of a UV bandpass filter, and **c** is Spectral energy distribution of UV lamp

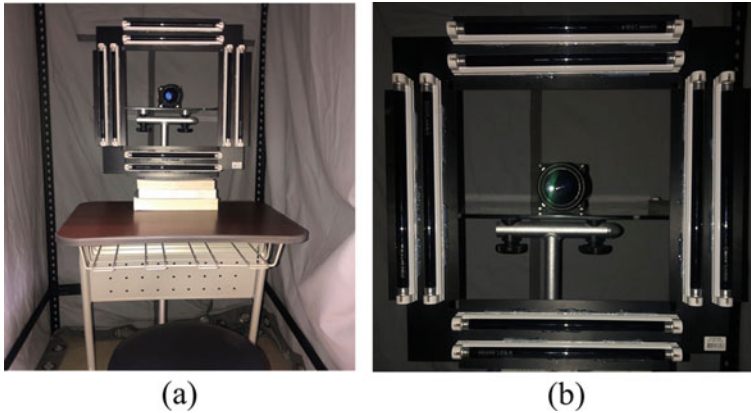


Fig. 2 It shows an experimental setup for acquiring UV images **a** is a darkroom experimental setup designed in this study and **b** is the placement of UV light and CCD camera

2.2 Image Preprocessing

Among the 51 UV images obtained, the experiment was conducted with a total of 42 UV images, classifying the available images. The segmentation ground truth was generated for 42 images. Subsequently, a dataset consisting of 126 UV images was constructed by augmenting the obtained images and ground truth images with random horizontal and vertical flipping, random rotation in the degree (90° , 180° , 270°), and perspective transformation method. Additionally, all images were scaled to $416 * 618$ pixels before being used as input to the model. After uploading the constructed dataset to GitHub, it was imported and used for training as a model.

2.3 Model Architecture Based ResNet50 and U-Net

In this study, we used the CNN model that the base model is ResNet50 and the segmentation model is U-Net for training the lesion segmentation. ResNet is a model that resolves the instability of learning due to gradient vanishing when the neural network is deepened and trained a lot by minimizing the residual of the skip connection. ResNet50 has 50 layers and repeats different residual block forms for each layer. U-Net, which is used as a segmentation model, is widely used in medical imaging because high-accuracy segmentation is possible even with a small number of data by combining the context of the image captured in the contracting path and the feature map upsampling in the expansive path. Its effectiveness has been verified in several studies.

The model training and the pre-processing were performed using Google Colaboratory, Python 3.7.11, and TensorFlow 2.5.0. Also, when training the model, the batch

size was set to 10, and steps per epoch were set to 512, and then it was trained for 30 epochs. To optimize the model, we used the Adam optimization algorithm, which allows stable optimization by combining the strengths of RMSProp and momentum and properly modifying the direction of change and step size.

3 Results

This study conducted verification of the performance of segmentation using Matlab R2021a (MathWorks Inc., MA, USA). We calculated Accuracy, Jaccard Index (JI), Dice Coefficient (DIC), Hausdorff Distance (HD), and Structural Similarity Index Measure (SSIM) between all classified pixels and their corresponding ground truth pixels.

The performance test showed a predicted segmentation result image as shown in Fig. 3 obtained for the nevus observed in the arm, and the segmentation performance results are shown in Table 1. The JI and DIC, which represent the ratio of the aggregated region between ground truth and predicted image, and the overlapping region of the two images, showed low results of 0.7868 and 0.8282, respectively, but showed high similarity with HD and SSIM indices of 2.4646 and 0.9999, respectively.

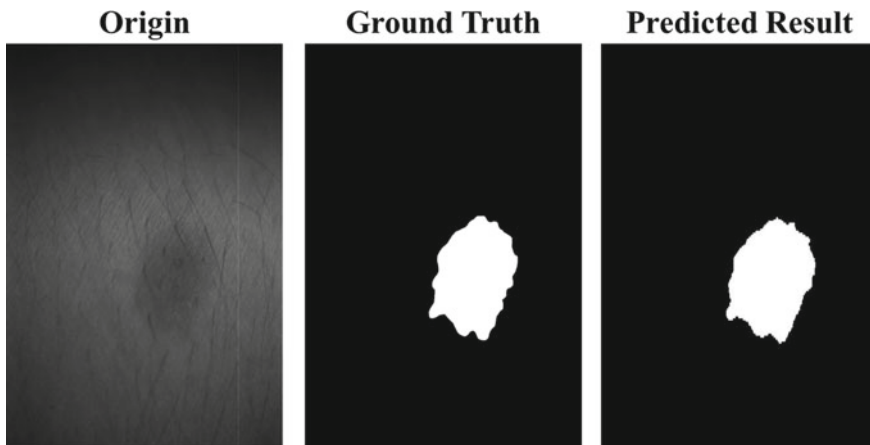


Fig. 3 These figures are the original UV image, ground truth, and predicted result image of the arm that nevus is observed

Table 1 The performance evaluation results with 5 indices

	Accuracy	JI	DIC	HD	SSIM
Value	0.9996	0.7868	0.8282	2.4646	0.9999

4 Discussion and Conclusion

Recently, interest in skincare has been increased resulting in a growing demand for segmentation technology that quantitatively analyzes the skin and removes objects of interest, not only from medical aspects but also in terms of cosmetic aspects. In particular, studies or techniques that visually represent the results can provide high confidence and satisfaction to both the person performing the procedure as well as the patient.

Therefore, by using UV imaging, we intended to conduct a base study on segmentation methods and analysis for skin lesions to supplement the limitations of visual inspection. We segmented the cutaneous pigmentation lesion and normal area in the UV image using CNN based on U-Net architecture after acquiring the UV image of the nevus and constructing an augmented dataset. After verifying the segmentation performance by comparing the predicted segmentation results with the ground truth, it showed an accuracy of 0.9996, a JI of 0.7868, and an SSIM value of 0.9999. Thus, it is confirmed that cutaneous pigmentation lesions in UV images can be effectively segmented with a CNN model combining ResNet50 and U-Net. Therefore, for the planning of the treatment methods, this study will be available as a preceding study for the treatment area of cutaneous pigmentation and its severity assessment.

In addition, the recognition of the pattern and distribution of melanin in cutaneous pigmentation skin lesions provides helpful information for confirming the hyperplasia severity of melanocytes and distinguishing between the malignant and normal lesions. The melanin distribution map of the lesion can be obtained using the UV image, and which predicts the possibility of a transition of an existing nevus to a melanoma. In the future, we will assess the severity of the lesions by generating and analyzing a melanin distribution map of the segmented pigmentation lesion.

Acknowledgements This work was funded by BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2022R1A2C1010170).

References

1. Paraskevas LR, Halpern A, Marghoob A (2005) Utility of the wood's light: five cases from a pigmented lesion clinic. *Br J Dermatol* 152(5):1039–1044
2. Deepa J (2021) A Convolutional neural network based skin lesion segmentation from dermoscopy images. *Turkish J Comput Math Educ (TURCOMAT)* 12(10):7557–7564
3. Liang Y, Sun L, Ser W, Lin F, Thng STG, Chen Q, Lin Z (2018) Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis. *Comput Biol Med* 99:123–132
4. Tao S, Jiang Y, Cao S, Wu C, Ma Z (2021) Attention-Guided network with densely connected convolution for skin lesion segmentation. *Sensors* 21(10):3462

Korean-Vietnamese Neural Machine Translation at Sub-word Level



Van-Hai Vu, Quang-Phuoc Nguyen, and Cheol-Young Ock

Abstract Recently, the explosion of deep learning has brought development opportunities to various fields such as automatic speech recognition, image recognition, financial fraud detection, etc. For the field of natural language processing (NLP) and more specifically for the field of machine translation (MT), advanced MT methods are also continuously being published based on the development of deep learning. However, most of the research focuses on popular languages such as English, French, and Chinese. Research related to low-level languages such as Korean or Vietnamese is still quite limited. We try to enrich the library of Vietnamese-Korean MT studies by applying Byte Pair Encoding (BPE) into MT systems. The results are evaluated by the method of Bi-Lingual Evaluation Understudy (BLEU) and Translation Error Rate (TER).

Keywords Neural machine translation · Vietnamese-Korean NMT · BPE

1 Introduction

MT is a branch of natural language processing that combines language, translation, and computer science. An automatic translation machine translates one language (called the source language) into one or more other languages (called the target language) automatically, without human intervention during the translation process. There are many studies on MT of popular languages with many approaches such as SMT [1–3] or NMT [4–6]. Recently, the introduction of Bi-direction Encoder Representation from Transformer (BERT) has made a breakthrough in the aspects of NLP, and this model also makes a positive contribution to improving the quality of

V.-H. Vu · C.-Y. Ock (✉)

Department of Electrical, Electronic and Computer Engineering, University of Ulsan,
Ulsan 44610, Republic of Korea
e-mail: okcy@ulsan.ac.kr

Q.-P. Nguyen

Chainos Global, Seoul 1338-12, Republic of Korea

MT systems [6–8]. For low-language pairs also, some studies have been published [9, 10], but the number of these publications is still limited compared to common language pairs.

In terms of the relationship between Vietnam and Korea, according to the statistics of the Korean Statistical Information Service¹ in 2020, the number of Vietnamese holding Korean nationality ranks second among foreigners holding Korean nationality. In the opposite direction, according to the statistics in February 2020, there are also about 170,000 Korean living and working in Vietnam. Besides, the demand for Vietnamese to learn Korean and for Koreans to learn Vietnamese is also increasing. However, translation machines like Google Translate or Papago have not yet provided Vietnamese-Korean translations as expected for users.

For the above reasons, we believe that it is necessary to contribute to the Vietnamese-Korean MT research community. Our study tried to improve the performance of Vietnamese-Korean NMT by parsing words in the parallel corpus into sub-word by BPE technique. The performance of Vietnamese-Korea NMT systems is evaluated by BLEU [11] and TER [12] scores and the result shows that training the NMT model at the sub-word level can improve the Vietnamese-to-Korean NMT system and the reverse direction by both BLEU and TER evaluation methods.

2 Neural Machine Translation

NMT applies neural network models to learn a statistical model for MT. This is a sequence-to-sequence model that has a sequence as input and a string as output. One of the effective approaches is Encoder-Decoder.

2.1 Encoder and Decoder

The encoder is responsible for receiving the input sequence and encoding it into a fixed-length vector, while the decoder decodes the above vector and predicting the output sequence.

Figure 1 is a general description of the encoder-decoder model in NMT.

The encoder consists of numerous RNN layers (or LSTM or GRU layers), which receive the signal of an element of the input sequence and transmit it to the end of the network. In NMT, the input signal is the embedding of the input sentence. Each word is represented by x_i where i is the word order.

The hidden states are calculated by the corresponding input x_t and the previously hidden state h_{t-1} . The information of the input will be calculated from beginning to end, and it is concatenated into vector c . This vector will be used as the first hidden state of the decoder.

¹ https://kosis.kr/statHtml/statHtml.do?orgId=111&tblId=DT_1B040A6.

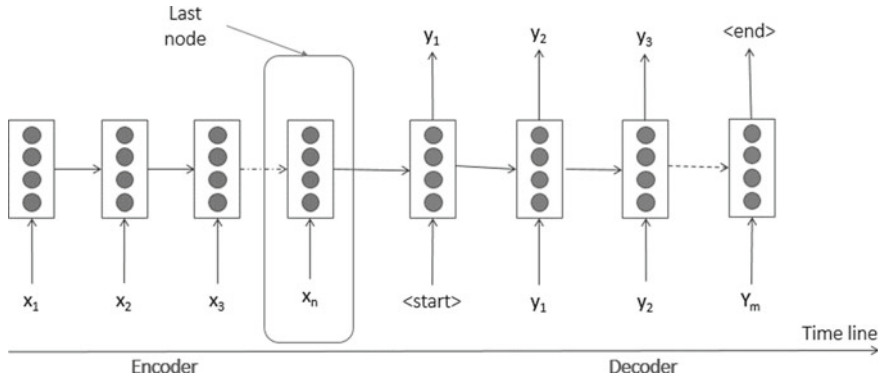


Fig. 1 Encoder-decoder in NMT

The decoder also includes many layers of the RNN. It predicts the output y_t at time step t . Each of these elements takes as input the previous hidden state and produces its own hidden state as output. In MT, the output is a sentence in the target language. Each word in this sentence is represented by y_i where i is the word order.

An important problem with the encoder-decoder model is that the input sentence information is only stored at the last node. Therefore, when the machine learning model reaches the last word in the input string, the information of the first words in the input string will gradually disappear. The attention mechanism was born to solve this problem.

2.2 Attention Mechanism

The introduction of the attention mechanism [4] in deep learning has brought significant effects to various models, it has been and continues to be an indispensable component in the most advanced models.

Figure 2 demonstrates the attention mechanism in NMT. When the decoder generates words of the target sentence, instead of using only the hidden state of the previous time-step, it uses both the hidden state of the previous time-step and the context vector that is accumulated at the encoder. The context vector will be calculated by taking the sum of the products between the hidden state vector and the weight vectors.

3 Byte Pair Encoding

BPE [12] is a data compression algorithm that increases the efficiency of all recent advanced NLP models. Usually, when training data, words that do not appear in the dictionary (out-of-vocabulary: OOV) are usually represented by the character. To

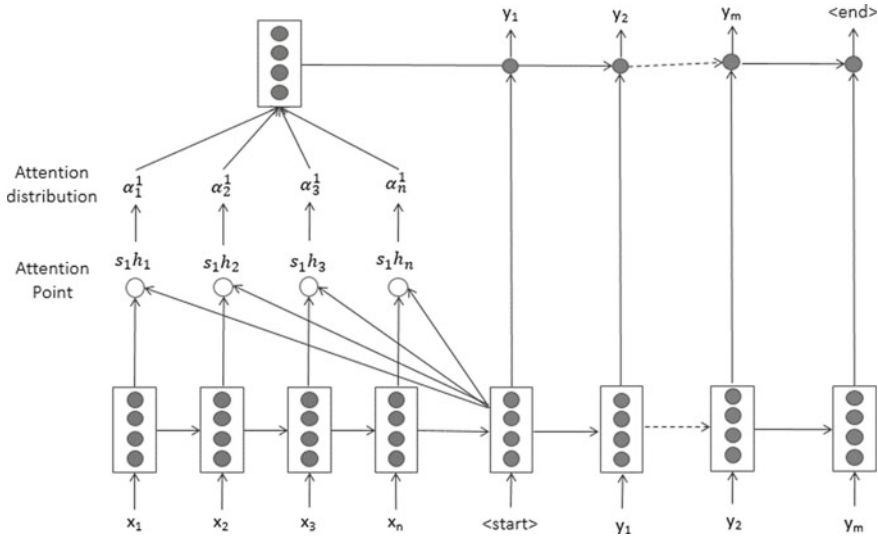


Fig. 2 NMT with attention mechanism

solve this problem, the BPE method was born with the idea of analyzing words into sub-words.

For example, the words *low*, *lower*, *lowest*, *high*, *higher*, and *highest* are all made up of *low*, *high*, and the suffix *er*, and *est*. When encrypting these words, instead of encrypting all 6 above words, we can encrypt four tokens *low*, *high*, *er*, and *est*.

The BPE method will calculate the frequency of occurrence of the sub-words together and find a way to group them if their frequency is the greatest. Specifically, the process of splitting words into sub-words starts with initializing a dictionary, then splitting the word into a character sequence; these characters will be added to the character of the dictionary. In the next step, the symbol pairs that frequently appear together are selected and grouped. This process is repeated until the vocabulary size is reached.

4 Experimentation and Result

4.1 Vietnamese-Korean Parallel Corpus

Our studies are based on data sets that we have collected and processed. This dataset consists of more than 454 k sentence pairs. We found that most Vietnamese words exist in the form of compound words; however, in the initial data, Vietnamese words exist in a single-word form. This leads to difficulty in determining the border of words and reduces the quality of MT. Therefore, to increase the quality of MT, we

Table 1 Korean-Vietnamese parallel corpus

	Number of sentences	Average length	Number of tokens	Number of vocabulary
Korean	454,751	12.0	5,435,686	397,130
Vietnamese		19.3	8,790,197	40,090

used word segmentation to define the border for the Vietnamese. The details of the parallel corpus that we used in this study are described in Table 1.

4.2 Experiment, Evaluation and Result

We have used the open source OpenNMT [13] to build a Vietnamese-Korean bi-direction NMT system. These translators are described as follows:

- *Baseline*: Uses the original Korean and Vietnamese texts; Vietnamese texts were normalized and tokenized using the Moses tokenizer² and converted to lowercase.
- *BPE*: texts of Korean in the parallel corpus were tokenized into sub-word by [14] (we set BPE vocabulary size to 32 K words for both source and target languages in all models).

In this research, the two most popular evaluation methods were used to evaluate the quality of MT systems including BLEU and TER. BLEU is the most common method for assessing the accuracy of MT systems, and this indicates the similarity between the candidate text and the reference texts. This method counts the number of matching n-grams of the candidate and reference. The TER is used to measure the amount of correction that a translator needs to produce a perfect translation. This is based on the edit distance between the candidate and the reference.

The results of Vietnamese-Korean NMT systems are shown below.

4.3 Discussion

According to Table 2, the application of BPE improves Vietnamese-Korean NMT in both dimensions through both assessment methods. For the Korean-Vietnamese dimension, BPE improved by 0.72 BLEU points and 4.58 TER points, respectively. In the reverse direction, BPE improves the NMT system by 11.91 BLEU points and 7.37 TER points, respectively.

The BPE method segmented Korean words into sub-words (a sequence of characters) based on the statistic of the most frequent character sequences. This will solve the OOV problem leading to minimizing the possibility that the data in the

² <http://www.statmt.org/ Moses>.

Table 2 Results of Vietnamese-Korean NMT systems

Systems		BLEU	TER
Korean-to-Vietnamese	Baseline	24.66	66.32
	BPE	25.38	61.74
Vietnamese-to-Korea	Baseline	10.76	70.94
	BPE	22.67	63.57

test set does not exist in the training set. Moreover, a word in Korean can have many different endings such as *ha-da*, *hae-yo*, and *hab-ni-da*. The word “*cảm ơn*” (*thank you*) in Vietnamese can be translated into *gam-sa-ha-da*, *gam-sa-hae-yo*, and *gam-sa-hab-ni-da* in Korean. If the candidate text is *gam-sa-ha-da*, but the reference text is *gam-sa-hae-yo*, the word has been mistranslated. However, when the word is split into sub-words, the candidate text and the reference texts have the same token *gam-sa*. This will increase the accuracy of the evaluation methods.

5 Conclusions and Future Work

In this research, we gave a brief presentation on NMT and BPE. We also applied BPE to convert Korean words to sub-word before training translation model. The experimental results show that applying BPE can increase the quality of Vietnamese-Korean NMT by BLEU and TER scores.

In the future, we plan to do Vietnamese-Korean NMT experiments at document level. We hope that this method can continue to improve the quality of Vietnamese-Korean NMT systems.

Acknowledgements This work was supported by a grant from the Institute for Information & Communications Technology Planning & Evaluation (IITP), funded by the Korean government (MSIT) (No. 2013-0-00131, Development of Knowledge Evolutionary WiseQA Platform Technology for Human Knowledge Augmented Services).

References

1. Cho SW, Lee E-H, Lee J-H (2018) Phrase-level grouping for lexical gap resolution in Korean-Vietnamese SMT. In: Hasida K, Pa WP (eds). Computational linguistics. pp 127–136. Springer Singapore, Singapore
2. Koehn P, Hoang H, Birch A, Callison-Burch C, Federico M, Bertoldi N, Cowan B, Shen W, Moran C, Zens R, Dyer C, Bojar O, Constantin A, Herbst E (2007) Moses: open source toolkit for statistical machine translation. In: Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions. Association for Computational Linguistics, Prague, Czech Republic, pp 177–180
3. Koehn P (2005) Europarl: A parallel corpus for statistical machine translation. In: Proceedings of machine translation summit X: papers. Phuket, Thailand, pp 79–86

4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Proceedings of the 31st international conference on neural information processing systems. Curran Associates Inc., Red Hook, NY, USA. pp 6000–6010
5. Voita E, Serdyukov P, Sennrich R, Titov I (2018) Context-aware neural machine translation learns Anaphora resolution. In: Proceedings of the 56th annual meeting of the association for computational linguistics (vol 1: Long Papers). Association for Computational Linguistics, Melbourne, Australia. pp 1264–1274. <https://doi.org/10.18653/v1/P18-1117>
6. Guo Z, Nguyen ML (2020) Document-level neural machine translation using BERT as context encoder. In: Proceedings of the 1st conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop. Association for Computational Linguistics, Suzhou, China. pp 101–107
7. Devlin J, Chang M-W, Lee K, Toutanova K (2019) 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, Minneapolis, Minnesota, pp 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
8. Zhu J, Xia Y, Wu L, He D, Qin T, Zhou W, Li H, Liu T-Y (2020) Incorporating BERT into Neural Machine Translation. In: [arXiv:2002.06823](https://arxiv.org/abs/2002.06823) [cs]. , Addis Ababa, Ethiopia
9. Östling R, Tiedemann J (2017) Neural machine translation for low-resource languages. [arXiv:1708.05729](https://arxiv.org/abs/1708.05729) [cs]
10. Vu V-H, Nguyen Q-P, Shin J-C, Ock C-Y (2020) UPC: an open word-sense annotated parallel corpora for machine translation study. *Appl Sci* 10:3904. <https://doi.org/10.3390/app10113904>
11. Papineni K, Roukos S, Ward T, Zhu W-J (2002) Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp 311–318. <https://doi.org/10.3115/1073083.1073135>
12. Snover M, Dorr B, Schwartz, Micciulla L, Makhoul J (2006) A study of translation edit rate with targeted human annotation. In: Proceedings of the 7th conference of the association for machine translation in the Americas: technical papers. Association for Machine Translation in the Americas, Cambridge, Massachusetts, USA, pp 223–231
13. Klein G, Kim Y, Deng Y, Senellart J, Rush A (2017) OpenNMT: open-source toolkit for neural machine translation. In: Proceedings of ACL 2017, system demonstrations. Association for Computational Linguistics, Vancouver, Canada, pp 67–72
14. Sennrich R, Haddow B, Birch A (2016) Neural machine translation of rare words with subword units. In: Proceedings of the 54th annual meeting of the association for computational linguistics (vol 1: Long Papers). Association for Computational Linguistics, Berlin, Germany, pp 1715–1725. <https://doi.org/10.18653/v1/P16-1162>

Classification of the Factors Influencing Center of Pressure Using Machine Learning and Wavelet Analysis



Xueli Ning, Young Kim, Jong Gab Ho, Xin Guo, Rathna Damsmoun, A. Hyun Jung, and Se Dong Min

Abstract Postural control is influenced by somatosensory, visual, and vestibular information. This study investigated the effects of open/closed eyes and external vibration stimuli on Achilles tendon on postural control in healthy adults. The visual and vibrational inputs were selected to induce somatosensory changes in quiet standing posture. Machine learning and wavelet analysis were used to classify the factors that influence postural control. Fifteen healthy subjects performed a quiet standing task under four different conditions: eyes closed, eye opened, with Achilles vibration, and without Achilles vibration. Results showed that both visual and vibration input conditions caused a significant difference in the mean velocity of anteroposterior direction in Center of Pressure (COP) shifting. This indicated that visual and somatosensory information changes have a significant role on postural control. The feasibility of using machine learning for the detection and classification of the factors influencing postural control function has been confirmed in this study. Combination of the features extracted by wavelet packet analysis and time-domain features showed to have the best classification outcome with F1-score of 96.77%.

Keywords Postural control · Balance · Center of pressure · Wavelet analysis · Machine learning

X. Ning · Y. Kim · J. G. Ho · X. Guo · R. Damsmoun · S. D. Min (✉)
Department of Software Convergence, Soonchunhyang University, Asan, Republic of Korea
e-mail: sedongmin@sch.ac.kr

Y. Kim
e-mail: ykim02@sch.ac.kr

A. Hyun Jung · S. D. Min
Department of Medical IT Engineering, Soonchunhyang University, Asan, Republic of Korea

1 Introduction

Postural control requires the integration of sensory input from the somatosensory, visual, and vestibular systems to maintain balance [1]. Understanding the effects of different sensory inputs on balance and the causes of balance loss is important for diagnosis, prevention, and management of falls [1].

Changes in balance are often quantified by time-domain Center of Pressure (COP) analyses such as trajectory length of COP, velocity, and area of postural sway [2]. However, frequency-domain COP analyses using power spectrum, information entropy, and discrete wavelet analysis are found to obtain more in-depth information [3, 4]. Discrete wavelet analysis can decompose COP signals into multiple independent frequency bands, which can be classified into the factors used for different postural controls [5].

Effects of sensory inputs on balance and postural control mechanism have been explored in existing studies where somatosensory changes were achieved by applying different external vibrations to the Achilles tendon [6–8]. The functions of postural control can be influenced by the changes of COP when the sensory inputs such as visual, vestibular, proprioceptive information are changed [3]. For this reason, many researchers analyzed the changes in balance with different sensory inputs, but studies classifying the factors affecting balance functions are lacking.

Decreased balance and postural control can increase the risk of falls. Howcroft et al. [9] investigated static standing postures with eyes open and eyes closed in older adults, and the COP data obtained using static posture measurements to calculate relevant metrics could distinguish between older fallers and non-fallers. They found out that visual removal can lead to increased risk of falls in older adults [9]. Lee et al. [10] compared acceleration data versus COP combined with acceleration data using machine learning and found out that the addition of COP improved accuracy in fall detection [10]. By using machine learning, COP time-domain and frequency-domain indicators can be extracted for the detection and classification of different postural control mechanisms.

In this study, visual and somatosensory functions in postural control were examined through four different static standing conditions. The aim of this study was first to determine the factors for maintaining balance without falling and then to test the feasibility of COP analysis for postural control factor classification by using machine learning.

2 Method

2.1 Participants

A total of 15 healthy young adults (age: 24.2 ± 3.8 years, height: 168.0 ± 10.1 cm, weight: 64.5 ± 16.1 kg) were participated in this study. None of the subjects had any

balance-impairing orthopedic disorders, injuries to the lower extremities, or visual impairment (corrected normal vision was considered within the screening range). We recruited healthy young adults for this experiment to exclude the effects of other diseases or aging factors on postural control.

2.2 *Experimental Procedures*

Data for the plantar center of pressure were obtained using a Force Plate (Tekscan Inc., Boston, MA, USA) with a sampling rate of 100 Hz. Vibration motor (Qinfeng Motor Inc., Guangdong, China) was strapped around the ankle to externally stimulate Achilles tendon.

The four experimental conditions were as follows: (1) standing on the force plate with eyes open (EO); (2) standing on the force plate with eyes closed (EC); (3) standing on the force plate with eyes open with Achilles tendon vibration (OV); and (4) standing on the force plate with eyes closed with Achilles tendon vibration (CV). When the subjects' eyes were open, they were instructed to look at a black spot 2 m away at eye level. Each upright condition lasted for 60 s and the tasks were assigned in a random order. Resting times given at the end of the vibration condition and non-vibration condition were 5 min and 1 min, respectively.

2.3 *Data Processing*

The force plate acquires COP coordinates, mediolateral velocity (V_{ml}), antero-posterior velocity (V_{ap}), and total velocity (V_t) signals at a sampling rate of 100 Hz. The average velocity of COP is widely used to quantify the postural sway, which is calculated by dividing the total length of the sway by the recorded time. In this study, the average velocity of COP is calculated for the mediolateral direction, anteroposterior direction, and total direction.

Ten common time-domain features (Mean, Variance, Standard Deviation, Root Mean Square, Skewness, Cliffness, Waveform, Peak, Pulse, and Margin) were derived for the COP's V_t signal. Nine-level Symlet-8 wavelet discontinuous wavelet decomposition and wavelet packet decomposition were used to extract two separate frequency-domain features. The signal can be split into ten frequency bands using the nine-level Symlet-8 wavelet discrete wavelet transform, and the energy of the ten frequency bands is determined as the frequency-domain feature, denoted as *Wave_feature*. Whereas different frequency bands are thought to correspond to various aspects of postural control, (1) medium frequency (1.56–6.25 Hz) is correlated with muscle body sensation, (2) low frequency (0.39–1.56 Hz) is correlated with the cerebellum, (3) sub-low frequency (0.10–0.39 Hz) is correlated with the vestibule, and (4) ultra-low frequency (< 0.10 Hz) is correlated with vision [5].

Wavelet packet decomposition retains the high frequency part of the information, and the n th layer decomposition yields 2^n frequency bands. In wavelet decomposition, the frequency bands of layer 4 correspond to muscle body sensation, so in wavelet packet decomposition, the 16-node energy of layer 4 is extracted as the frequency-domain feature, noted as Layer4_feature.

To achieve classification of vibration and non-vibration conditions, a binary support vector machine (SVM) is trained with a Gaussian kernel. Due to the small amount of data, the leave-one-out cross-validation method (LOOCV-SVM) is used as the performance evaluation of the model. The COP speed data are divided into two categories: Achilles tendon vibration and there is no Achilles tendon vibration; there are 30 samples in each of the two types, the time-domain and frequency-domain features are extracted from the COP velocity data V_t , and the classification effects of different features are compared.

2.4 Statistical Analysis

The three mean velocity data (total, mediolateral, and anteroposterior) of the COP were analyzed using one-way repeated measures ANOVA to compare the difference between experimental conditions. SPSS statistical software was used for analysis with an alpha level of 0.05.

3 Results

3.1 COP Average Velocity

Table 1 shows the mean and standard deviation of COP velocity in mediolateral direction (M-L), anteroposterior direction (A-P), and the combination of the two, which is total velocity for the four conditions.

The results of mean COP velocity difference in mediolateral direction were as follows: EO-EC: $1.44-1.47 = -0.03$ cm/s, $p = 0.81$; EO-OV: $1.44-1.95 = -$

Table 1 Average velocity of COP in each condition

Velocity (cm/s)	Eyes open	Eyes closed	Eyes open + vibration	Eyes close + vibration
M-L	1.44 ± 0.50	1.47 ± 0.13	1.95 ± 0.49	1.99 ± 0.83
A-P	1.86 ± 0.48	1.54 ± 0.14	2.91 ± 0.84	2.47 ± 0.87
Total	3.18 ± 1.00	2.83 ± 0.28	4.50 ± 1.09	4.03 ± 1.34

M-L: Mediolateral **A-P:** Anteroposterior

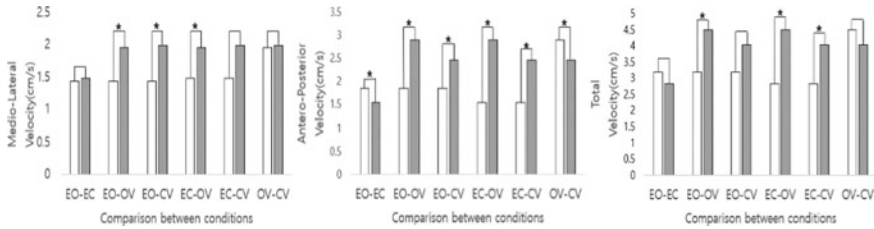


Fig. 1 Comparison of COP velocity between different conditions. EO: eyes open, EC: eyes closed, OV: eyes open + vibration, CV: eyes closed + vibration. Asterisks (*) denote $p < 0.05$

0.51 cm/s, $p = 0.01$; EO-CV: 1.44–1.99 = - 0.55 cm/s, $p = 0.02$; EC-OV: 1.47–1.95 = - 0.48 cm/s, $p = 0.01$; EC-CV: 1.47–1.99 = - 0.52 cm/s, $p = 0.07$; OV-CV: 1.95–1.99 = - 0.04 cm/s, $p = 0.8$. The results of mean COP velocity difference in anteroposterior direction were EO-EC: 1.86–1.54 = - 0.32 cm/s, $p = 0.04$; EO-OV: 1.86–2.91 = - 1.05 cm/s, $p = 0.00$; EO-CV: 1.86–2.47 = - 0.61 cm/s, $p = 0.04$; EC-OV: 1.54–2.91 = - 1.37 cm/s, $p = 0.00$; EC-CV: 1.54–2.47 = - 0.93 cm/s, $p = 0.01$; OV-CV: 2.91–2.47 = 0.44 cm/s, $p = 0.04$. The results of mean total COP velocity difference were EO-EC: 3.18–2.83 = 0.35 cm/s, $p = 0.24$; EO-OV: 3.18–4.50 = - 1.32 cm/s, $p = 0.00$; EO-CV: 3.18–4.03 = - 0.85 cm/s, $p = 0.08$; EC-OV: 2.83–4.50 = - 1.67 cm/s, $p = 0.00$; EC-CV: 2.83–4.03 = - 1.2 cm/s, $p = 0.01$; OV-CV: 4.50–4.03 = 0.47 cm/s, $p = 0.12$ (Fig. 1).

3.2 Classification of Vibration and Non-vibration Conditions

The five features were put into LOOCV-SVM training test, and Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), F1-score (F1_score) were calculated for each approach. The results are shown in Table 2.

Using F1_Score, classification performance was evaluated comprehensively. The F1_Score scores for different input features were 72.41% (Time), 79.41% (Wave), 75.86% (Time + Wave), 80.65% (Layer4_wave), and 96.77% (Time + Layer4_wave).

Table 2 Training results for different features

Feature	ACC (%)	SEN (%)	SPE (%)	F1_score (%)
Time	73.33	70.00	76.67	72.41
Wave	76.67	90.00	63.33	79.41
Time + wave	76.67	73.33	80.00	75.86
Layer4_wave	80.00	83.33	76.67	80.65
Time + Layer4_wave	96.67	100.00	93.33	96.77

4 Discussion

In this study, the effects of vision and vibration on postural control were examined in four different conditions. The results of the three directional indicators of COP mean velocity showed that the postural sway may be reduced due to the increase in body stiffness caused by the subject's tension. Although the average velocity of COP is widely used in balance assessment, it is most reliable in the A-P direction [2, 11]. Analysis of variance of the A-P direction indicator and the mean velocity change in A-P direction were statistically significant in both visual and vibration conditions.

The results of this study based on machine learning for the classification of vibration and non-vibration conditions can provide a reference for the detection and classification of the factors influencing postural control. The discrete wavelet analysis obtained in different independent frequency bands can characterize the visual and somatosensory changes [5], and for this reason, the classification performance is superior to that of the time-domain features. The classification performance of the features extracted by wavelet packet analysis is better than wavelet analysis, because wavelet packet analysis adds a large amount of high-frequency information to wavelet analysis [12]. The classification of vibrational versus non-vibrational conditions suggests that the use of COP frequency-domain features characterizing the frequency band of the somatosensory signal can be used to classify the presence or absence of changes in somatosensory function.

In the future research, the sample size will be increased for the classification of multiple sensory inputs, which will provide a feasible basis for detecting injuries to the sensory nervous system. Broad analysis on postural control-related indicators is believed to further indicate potential postural control risks, intervene in advance for treatment, reduce the risk of falls, and prevent the deterioration of injuries.

5 Conclusion

Both visual condition changes (eyes open or closed) and Achilles tendon vibration had significant effects on the average velocity in the anteroposterior direction of COP. The features extracted using COP wavelet analysis and wavelet packet analysis can better achieve COP signals in vibration and non-vibration conditions, providing a reference for using machine learning to analyze postural control functions.

Acknowledgements This research was funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2021R1I1A3059769) and funded by BK21 FOUR (Fostering Outstanding Universities for Research) (No.: 5199990914048) and funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2018R1D1A1B07050037).

References

1. Osoba MY, Rao AK, Agrawal SK et al (2019) Balance and gait in the elderly: contemporary review. *LaryngosCOPE Investigative Otolaryngol* 4(1):143–153
2. Quijoux F, Nicolaï A, Chairi I et al (2021) A review of center of pressure (COP) variables to quantify standing balance in elderly people: algorithms and open-access code. *Physiol Rep* 9(22):e15067
3. Jiang BC, Yang WH, Shieh JS et al (2013) Entropy-based method for COP data analysis. *Theor Issues Ergon Sci* 14(3):227–246
4. Chagdes JR, Rietdyk S, Haddad JM et al (2009) Multiple timescales in postural dynamics associated with vision and a secondary task are revealed by wavelet analysis. *Exp Brain Res* 197(3):297–310
5. Liang Z, Clark R, Bryant A et al (2014) Neck musculature fatigue affects specific frequency bands of postural dynamics during quiet standing. *Gait Posture* 39(1):397–403
6. Billot M, Handrigan GA, Simoneau M et al (2015) Reduced plantar sole sensitivity induces balance control modifications to compensate ankle tendon vibration and vision deprivation. *J Electromyogr Kinesiol* 25(1):155–160
7. McKay SM, Wu J, Angulo-Barroso RM (2014) Effect of Achilles tendon vibration on posture in children. *Gait Posture* 40(1):32–37
8. Dumas M, Valkanidis TC, Hatzitaki V (2019) Putting proprioception for balance to the test: contrasting and combining sway referencing and tendon vibration. *Gait Posture* 67:201–206
9. Howcroft J, Lemaire E D, Kofman J et al (2017) Elderly fall risk prediction using static posturography. *PLoS ONE* 12(2):e0172398
10. Lee CM, Park J, Park S et al (2020) Fall-detection algorithm using plantar pressure and acceleration data. *Int J Precis Eng Manuf* 21(4):725–737
11. Quek J, Brauer SG, Clark R et al (2014) New insights into neck-pain-related postural control using measures of signal frequency and complexity in older adults. *Gait Posture* 39(4):1069–1073
12. Chen G, Li QY, Li DQ et al (2019) Main frequency band of blast vibration signal based on wavelet packet transform. *Appl Math Model* 74:569–585

A Local Assessment Approach for Diagnosis of Thermal Imaging-Based Palmar Hyperhidrosis



Jiwon Lee, Juhyun Kim, and Onseok Lee

Abstract Hyperhidrosis is classified as a skin disease that causes excessive sweating due to emotional stress and lowers the quality of life according to the degree of discomfort. The most common palmar hyperhidrosis is diagnosed by the questionnaires in the clinic, and the necessary digital infrared thermal imaging and severity assessment also depend on the clinician's knowledge and experience. Therefore, to overcome the subjective evaluation of the diagnosis of skin diseases such as hyperhidrosis, we intend to present an objective diagnostic method by performing a quantitative approach. In this study, the evaporation heat of the hand, not the hand, was analyzed using a thermal imaging camera, and the fingers and palm areas were locally divided and evaluated. As the body temperature residual heat and sweat evaporated, the diffusion phenomenon with slow rates was captured as the amount of water increased, and the condition of hyperhidrosis can be diagnosed according to the degree of diffusion. These analysis data were used as the machine learning classifier training data to diagnose the state about presence or absence of sweat, and K-nearest neighbors, one of the classifiers used, showed the best performance with an accuracy of 96%. This study can be a basic study that quantitatively presents the criteria for determining the drug treatment injection site for palmar hyperhidrosis and can be presented as a new diagnostic evaluation method for thermal imaging-based palmar hyperhidrosis through the proof of the water evaporation heat of sweat.

Keywords Palm hyperhidrosis · Thermal image · Image processing · Evaporation heat · Machine learning · Diagnosis

J. Lee · J. Kim · O. Lee (✉)

Department of Software Convergence, Graduate School, Soonchunhyang University, 22, Soonchunhyang-ro, Asan City, Chungcheongnam-do 31538, Republic of Korea
e-mail: leeos@sch.ac.kr

O. Lee

Department of Medical IT Engineering, College of Medical Sciences, Soonchunhyang University, 22, Soonchunhyang-ro, Asan City, Chungcheongnam-do 31538, Republic of Korea

1 Introduction

Hyperhidrosis is classified as a skin disease according to the amount of sweat secreted more than necessary in response to emotional stimuli and the degree of discomfort the patient complains of. The symptoms of hyperhidrosis are most common in the hand part, which is the body part that is used the most, and the quality of life can be reduced due to social discomfort or difficulty in performing work [1]. In clinical practice, questionnaires for diagnosing hyperhidrosis depend on the examinee's knowledge and experience along with the visual observation, and the additional digital infrared thermal imaging is also dominated by subjective evaluation. This method is known as an efficient imaging method that can identify not only the sweat site but also changes in body temperature. However, when comparing the bilateral body heat, if there is a temperature difference of over 0.3° , there is an abnormality, and the more distinct the difference, it is judged that there is a disease. Depending on the research, 0.6° or 1.3° are considered as the standard. That is, the severity evaluation is not made objectively, and as there are various treatment methods for hyperhidrosis such as sympathectomy and injection some drug like Botox, it is necessary to objectify the severity of symptoms in order to treat them according to the constitution of the patient [2, 3].

Drug injection treatment for palmar hyperhidrosis is treated by injecting drugs such as botulinum toxin into various localized areas of the fingers and palms. Although it depends on the research, each finger and palm have many sweat glands, so a total of 15–36 topical areas are subcutaneously injected [4]. An iodine–starch test can be performed to determine sweat emission and treatment site, but it is a method that is not used well because it takes a long time and causes great discomfort. Also, since the points where iodine mixed with sweat and starch reacted do not always become injection sites, a grid of 1 cm apart is referred to determine the injection site. In other words, the criteria for determining the treatment site for drug injection for palmar hyperhidrosis are not clear.

On the other hand, it is ideal for the temperature change to gradually increase or decrease, but if water is included in the materials, the evaporation phenomenon occurs and the hydraulic continuity may be impaired by the evaporation mechanism until thermal equilibrium with the environment is achieved [5]. That is, the slow diffusion that occurs during evaporation represents an unstable temperature change. And, we hypothesized that the more excessive sweating of palmar hyperhidrosis occurs, the more unique diffusion phenomenon according to water and evaporation amount occurs. Based on this, it can be determined that the slower the speed, the more Abnormal the point of occurrence, that is, the place where there was sweat. Therefore, the diagnostic features of hyperhidrosis were analyzed by part division about the thermal imaging sequence images of hand hyperhidrosis. Ultimately, we propose a method to quantitatively determine the treatment site for drug injection at the same time with diagnosis and suggest a novel diagnostic evaluation method for thermal imaging-based palmar hyperhidrosis by proving the water evaporation heat of sweat.

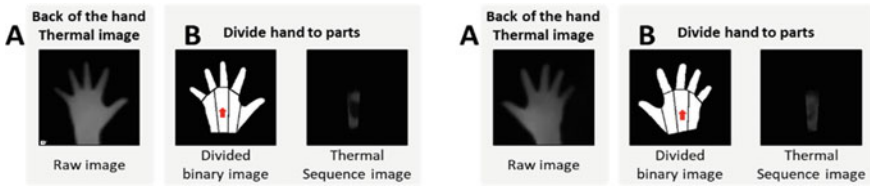


Fig. 1 Processing thermal image data. Abnormal and normal hand data (left/right). Part for representative results (red arrow)

2 Methods

2.1 Data Acquisition

Thermal imaging was performed in a dark room environment (internal temperature $20 \pm 0.5 \text{ }^\circ\text{C}$ and humidity $35 \pm 5\%$) controlled from the external environment. The feature of this experiment is not to capture a body object with a thermal imaging camera, but to capture the delivered temperature marks after being in close contact with the palm for 10 s. During close contact, an image of the back of the hand was collected (Fig. 1a), and 61 thermal image sequence images were collected by shooting at 3-second intervals for 3 min. This study is a pilot study for the diagnosis of hyperhidrosis through the detection of evaporation heat due to sweat. In order to set an arbitrary environment and sweat, the researcher wore a latex glove on the left hand for about 2 h and assumed that the sweat generated at this time was sweat caused by hyperhidrosis (Abnormal). For control data, the right hand wore nothing (Normal).

2.2 Data Processing for Analysis

To detect temperature marks in thermal images, hand object was extracted by performing canny operation on the back of the hand image. Mask images were generated from binary images of each part separated into 5 fingers and 3 palms, and the sum and average of pixel values were obtained by adding them to each thermal image sequence (Fig. 1b). To confirm the width of change in the pixel value of the sequence images on the same point, it was taken the absolute value. Here, the change in data of consecutive thermal images was confirmed by obtaining the pixel value difference of each sequence. In addition, the surface area, volume, cross-sectional area, and perimeter were calculated by creating 2D and 3D alpha shape objects (radius = 5) composed of points by intensity values in the 2D thermal image sequence. These parameters calculated for each part were calculated for the training data of the machine learning classifier to distinguish between Normal and Abnormal. In addition, to present a diagnostic evaluation solution for hyperhidrosis according

to the amount of water loss, we tried to analyze the features, using these parameters, according to the presence or absence of sweat by capturing the evaporation and diffusion phenomena for sweat and residual heat transferred during palm close together. As classifiers, ECOC-SVM (SVM), Decision Tree (DT), K-nearest neighbors (KNN), and Random Forest (RF) were used. Classification performance was evaluated by performing ten-fold cross-validation on the trained classifier. All data processings were performed using MATLAB 2021a.

3 Results

As the skin temperature is transferred and the residual heat evaporates, the pixel value in the thermal image gradually decreases. When sweating due to hyperhidrosis, the skin temperature of the corresponded area is lowered due to the evaporation of sweat. The lower the body temperature, the rapider the changes (Fig. 2a). In addition, in these Normal data, a prominent portion can be seen, and it is considered that evaporation is limited by diffusion at a slower rate as the hands are wetter and sweat is present. In addition, maintaining a constant pixel value as the sequence progresses can be said to be in thermal equilibrium with the background temperature.

Figure 2b shows the difference in pixel values between thermal image sequences. As residual heat and sweat evaporate, the amplitude gradually decreases. However, due to the slow diffusion, it is possible to see portions in which the amplitude of changes significantly than the previous sequence one. In Fig. 2c–f, the value decreases due to evaporation, but in Normal, the value shows the width of large change during a certain sequence due to diffusion. In the case of Abnormal with sweat, diffusion appears slower as it has more water, and data soar to height of lower range. It is considered that the range of change is low because it has a relatively low body temperature due to sweating and shows a rapid change in the initial stage. To generalize this phenomenon, it is necessary to analyze a lot of data.

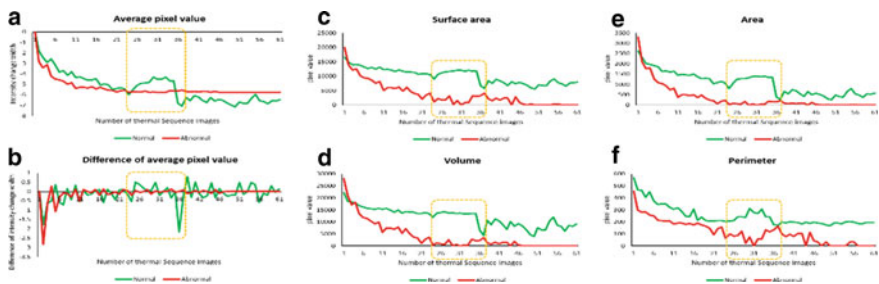
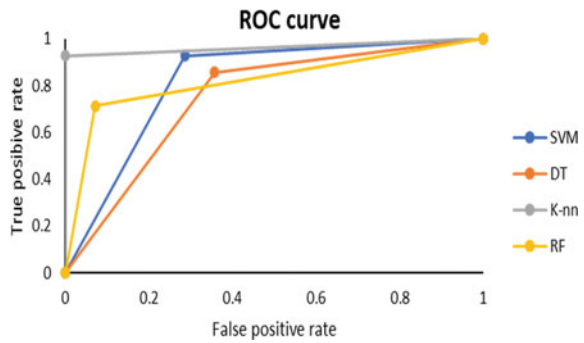


Fig. 2 Results are data on local area including concave area of the palm. It is classified into normal and abnormal data depending on the presence or absence of sweat. These were used as learning data for classification of machine learning

Table 1 Results' classification

	SVM	DT	KNN	RF
Accuracy	0.82	0.75	0.96	0.82
Sensitivity	0.93	0.86	0.93	0.71
Specificity	0.71	0.64	1.00	0.93
Precision	0.76	0.71	1.00	0.91
F1-score	0.84	0.77	0.96	0.80
AUC	0.87	0.75	0.96	0.82

Fig. 3 ROC curve shown as results of cross-validation



In these data, the presence or absence of sweat was binary-classified. Among the four classifiers used, the KNN result showed the best performance with an accuracy of 96% (Table 1). Figure 3 shows the ROC curves of the results of cross-validation on the trained classifier to evaluate the classifier performance equally. All of them were AUC 0.7 or higher, and KNN showed the best performance with an AUC greater than 0.9.

4 Conclusions

This study used a thermal imaging camera to analyze the evaporation heat of the hand, not the hand, and confirmed the difference in features according to the presence or absence of sweat. In palmar hyperhidrosis, there are many sweat glands in the palm of the hand, so subcutaneous injections are made over multiple sites during drug treatment. In other words, thermal imaging-based evaluation of hyperhidrosis may cause local assessment factors that may be missed when performing global analysis of the hand or palm. Therefore, a novel attempt was needed, and in this study, the fingers and palms were divided locally and analyzed, and machine learning with the analysis data was used to diagnose hyperhidrosis, that is, to evaluate the severity. In these data, just the presence or absence of sweat was judged, and all four classifiers

used showed excellent performance. This study can be proposed as the first novel diagnostic solution for quantitative evaluation of palmar hyperhidrosis by objectifying basal body temperature and hand size according to individual constitution. In the future, we plan to secure hyperhidrosis disease data to classify the severity and conduct generalization studies.

Acknowledgements This work was funded by BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1010170).

References

1. Romero FR, Haddad GR, Miot HA, Cataneo DC (2016) Palmar hyperhidrosis: clinical, pathophysiological, diagnostic and therapeutic aspects. *An Bras Dermatol* 91:716–725
2. Kim DH, Kim TH, Lee SH, Lee AY (2017) Treatment of palmar hyperhidrosis with tap water iontophoresis: a randomized, sham-controlled, single-blind, and parallel-designed clinical trial. *Ann Dermatol* 29(6):728–734
3. Jung HS, Lee DY, Park JS (2018) Alternative surgical methods in patients with recurrent palmar hyperhidrosis and compensatory hyperhidrosis. *Yonsei Med J* 59(2):345–348
4. Doft MA, Hardy KL, Ascherman JA (2012) Treatment of hyperhidrosis with botulinum toxin. *Aesthetic Surg J* 32(2):238–244
5. Maurais J, Beaumont É, Bourret J, Dauphinais E, Bouchard NA, Ayotte P (2020) A thermal imaging methodology to study evaporation kinetics in mine tailings. *Environ Sci: Water Res Technol* 6(5):1456–1464

A Study on Internet Annotation Analysis System Using Opinion Mining



Yen-Tae Jeong  and Byung-Hoon Jeon 

Abstract This paper proposes a big data sentiment analysis method and deep learning implementation method to provide a webtoon comment analysis web page for convenient comment confirmation and feedback of webtoon writers for the development of the cartoon industry in the video animation field. In order to solve the difficulty of automatic analysis due to the nature of Internet comments and provide various sentiment analysis information, long short-term memory (LSTM) algorithm, ranking algorithm, and word2vec algorithm are applied in parallel, and actual popular works are used to verify the validity. If the analysis method of this paper is used, it is easy to expand to other domestic and overseas platforms, and it is expected that it can be used in various video animation content fields, not limited to the webtoon field.

Keywords Sentiment analysis · Deep learning · AI · Webtoon · Comment analysis · Video animation · LSTM algorithm

1 Introduction

Recently, the production and consumption of various online video contents are exploding due to the spread of personal smart devices including smartphones and the establishment of fast communication networks. The spread of such infrastructure is a reality that is accelerating. The performance of personal smart devices is taking place as a replacement device for PCs and laptops, and the communication speed is rapidly evolving to 5G beyond LTE, and Wi-Fi can be easily used anywhere in daily life. This phenomenon is common all over the world, and in terms of Korea, it is showing a faster development speed [1]. The growth rate of the online animation market also changes in the surrounding environment, but accurate communication

Y.-T. Jeong · B.-H. Jeon (✉)

Department of Electrical Engineering, Dongguk University, 30, Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea

e-mail: bhjeon@dongguk.edu

is limited due to easy accessibility and easy production and consumption of video animation content. The most important communication channel is the comments of consumers about the works, and in the case of popular works, there are 3–4000 cases per day, and it is practically unreasonable to check them all. Therefore, in this paper, in order to make it easier to communicate opinions between writers and readers about webtoons, which are the field of manga, among online video animation contents, a vast amount of real-time comment data is automatically transmitted through big data sentiment analysis. We will discuss how to provide web services through analytic deep learning models.

2 Big Data Sentiment Analysis for Internet Comments

The process of ‘Big Data Sentiment Analysis for Internet Comments’ proposed in this paper is shown in (Fig. 1). As shown in (Fig. 1), it consists of the processes of data collection, data preprocessing, and data analysis, and detailed explanations are provided in each section.

3 Result of Big Data Sentiment Analysis

Big data sentiment analysis performed two tasks for data analysis using preprocessed data. First, the comments were summarized using the text rank algorithm [2], and then the words related to the author and work keywords were extracted using word2vec [3]. In the comment summary, after using a word graph that calculates the similarity between words using a count vectorizer and a sentence graph that calculates the similarity between sentences using the TF-IDF model [4], a ranking algorithm that outputs sentences and words with high ranking using this graph is used. Three lines of summary and 10 keywords were printed. Word2vec’s skip-gram model was used to extract words most similar to the author’s and work keywords in order to understand the relationship between words for the tokenized words.

In order to verify the validity of the big data sentiment analysis conducted in this paper, the analysis was conducted using “August Blizzard”, one of the most popular works in Naver, and the results are shown. Figure 2 shows the pipe chart as the result of the sentiment analysis for one episode. Positive, negative, and medium represent 1601 times, 2463 times, and 698 times, respectively, and when expressed as a ratio, negative accounts for 51%.

Figure 3 is a screen that provides a list of works provided by the platform in order to proceed with the analysis, and it is configured so that it is possible to search and select by episode or genre as needed. This screen can be replaced with a platform screen as needed.

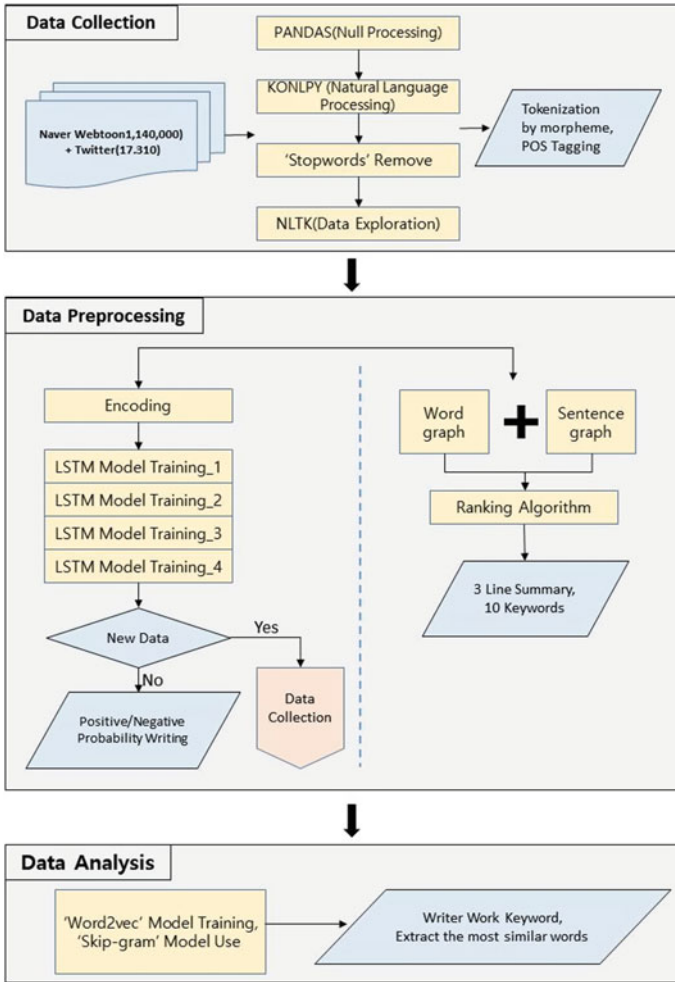


Fig. 1 Process of sentiment analysis

4 Conclusion

Due to the development of these infrastructure conditions, the online market is gradually increasing its share in many areas such as movies, books, music, and art, beyond offline. Among them, the cartoon industry in the field of video animation shows a more prominent phenomenon. According to a statistical survey by the Korea Creative Content Agency (KOCCA), the development of the offline comics industry is on the decline, while the online comics industry is showing a rapid growth of more than 30% in all fields (number of companies, employees, sales, etc.). Recently, the production of movies and dramas based on Webtoon (Web + Carton) [5] is also actively being

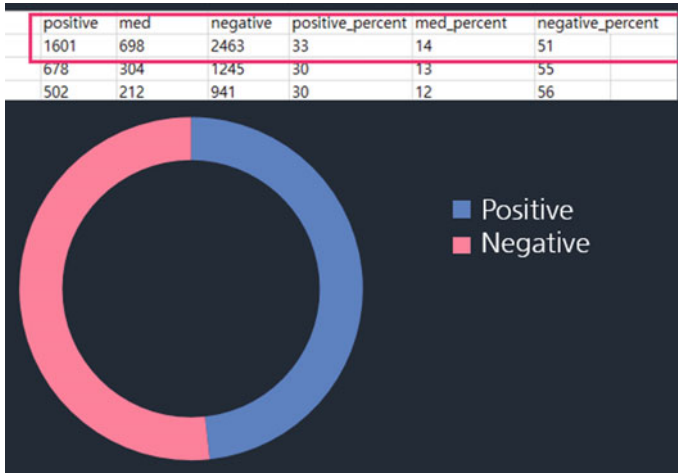


Fig. 2 Emotional analysis result (“August blizzard”)

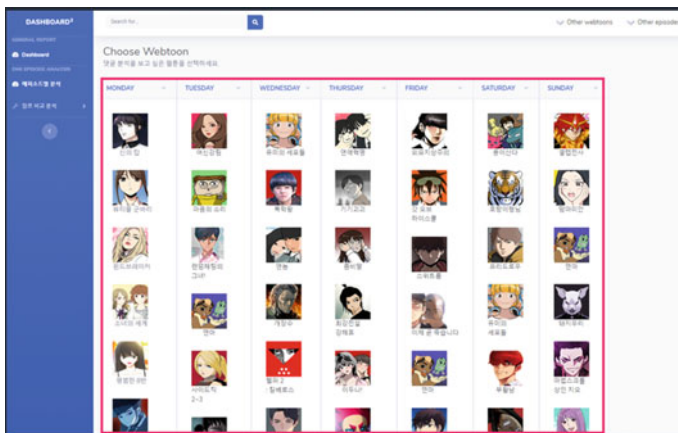


Fig. 3 Result dashboard screen (work selection screen)

made. If this market is included, the growth rate of the online animation market is increasing more rapidly [6].

In this paper, a big data-based sentiment analysis method and implementation method are proposed to provide webtoon comment analysis web pages for convenient comment confirmation and feedback of webtoon writers. Big data analysis method was used to analyze a large amount of comments. In order to solve the difficulty of automatic analysis due to the nature of comments, LSTM algorithm, ranking algorithm, and Word2vec algorithm are applied in parallel to provide various information, and to verify validity, actual popular works were implemented. It was possible to judge the possibility of using a big data analysis. If this analysis method is used,

it is easy to expand to other domestic and overseas platforms, and it is expected that it can be used in various content fields, not limited to the webtoon field.

References

1. Rauschnabel PA, Felix R, Hinsch C (2019) Augmented reality marketing: how mobile ARapps can improve brands through inspiration. *J Retail Consum Serv* 49:43–53
2. Nallapati R, Zhou B, dos Santos C, Gulcehre C, Xiang B (2016) Abstractive text summarization using sequence-to-sequence RNNs and beyond. In: *Proceedings of the 20th SIGNLL conference on computational natural language learning*, Berlin, Germany, pp 280–290
3. Mihalcea R (2004) Graph-based ranking algorithms for sentence extraction, applied to text summarization. In: *Proceedings of the 42nd annual meeting of the association for computational linguistics (ACL 2004) (companion volume)*, Barcelona, Spain
4. Zoph B, Le QV (2016) Neural architecture search with reinforcement learning. arXiv e-prints, art. [arXiv:1611.01578](https://arxiv.org/abs/1611.01578)
5. Nam K, Kim H-J (2020) The determinants of mobile game success in South Korea. *Telecommun Policy* 44:101855
6. Graphic Nevel Industry White Paper, Korea Creative Content Agency (2017)
7. Boemer F, Lao Y, Cammarota R, Wierzynski C (2019) nGraph-HE: a graph compiler for deep learning on Homomorphically encrypted data. In: *ACM international conference on computing frontiers*, pp 1–27 (2019)
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: *31st conference on neural information processing systems (NIPS 2017)*, Long Beach, CA, USA

Real-Time Pressure Feedback Sensor Designed for Manual Therapy—A Pilot Clinical Feasibility Study



Eun Hye Jo, Young Kim, Seungwan Jang, Hyun Young Lee, Seungmin Jeong, Ye Eun Kong, and Se Dong Min

Abstract Manual therapy is an evidence-based treatment technique that manual pressure of physical therapists is applied to the symptomatic body parts such as spine, soft tissue, and joints for treatment purposes. Although the importance of quantifying hand pressure used in manual therapy has been steadily suggested, an actual quantification of the pressure applied through therapists' fingers has not been studied in clinical setting. In this paper, a therapeutic pressure sensor (TPS) was developed with an aim to measure the finger pressure, quantify it, and give real-time feedback to the therapist during therapeutic sessions. Clinical data were collected from actual clinical trials to confirm clinical applicability. Results showed that the mean pressure applied through the thumb of the therapist was 1.48 ± 0.74 , index finger was 1.12 ± 0.63 , and the middle finger was 1.08 ± 0.79 in the left hand during muscle tension release manual therapy for pain caused by stiff gluteal muscle. For hamstring muscle tightness, the average pressure applied with the thumb was 1.50 ± 1.48 , index finger was 1.43 ± 0.82 , and the middle finger was 1.28 ± 0.62 in the left hand. The most often used finger was the thumb, and the level of therapeutic pressure showed to decrease with more treatment sessions. Accumulation of clinical data by more use of TPS for manual therapy would build therapy guidelines, which can improve the quality of treatment for patients as well as the health and safety for therapists.

Keywords Manual therapy · Pressure sensor · Quantification · Wearable

E. H. Jo · Y. Kim · S. Jang · H. Y. Lee · S. Jeong · S. D. Min (✉)
Department of Software Convergence, Soonchunhyang University, Asan, Republic of Korea
e-mail: sedongmin@sch.ac.kr

Y. Kim
e-mail: ykim02@sch.ac.kr

Y. E. Kong · S. D. Min
Department of Medical IT Engineering, Soonchunhyang University, Asan, Republic of Korea

1 Introduction

Manual therapy is an evidence-based treatment technique for musculoskeletal pain, sports injury, neuromuscular disorders, and postural imbalance. For treatment purposes, manual pressure of physical therapists is applied to the symptomatic body parts such as spine, soft tissue, and joints [1].

During manual therapy, the treatment pressure of the hand applied by the therapist differs depending on the treatment area [2]. Treatment area of manual therapy mainly includes the muscles and nerves that are shortened by injury, bad posture, musculoskeletal and neuromuscular diseases. Due to different cause of the symptoms, every patient is treated with different therapeutic approaches and level of pressure and each therapist applies different treatment procedures and techniques [3]. For this reason, many researchers reported the need of therapeutic manual pressure quantification.

Physical therapists apply pressure to the treatment area based on their anatomical, physiological knowledge, and experience [4]. Gagnon et al. [5] compared physical therapists with more than 5 years of experience and physical therapy trainees with 0 years of experience by applying pressure to the second and fourth lumbar spines [5]. Although the frequency of force application was similar in the two groups, it was found that the experienced physical therapists applied higher pressure than physical therapy trainees by about 50% or more, and based on this, the importance of force quantification and force visual feedback was presented [5]. As such, the importance of quantifying hand pressure used in manual therapy has been consistently suggested, but there are very few cases in which quantification has been attempted by measuring the pressure of the finger through a pressure sensor when performing manual therapy in actual clinical practice [6].

Pressure sensors in the form of gloves were mostly used as sensors to measure finger pressure [7, 8]. In the field of rehabilitation medicine, hand pressure sensors have been developed mainly for patient rehabilitation, such as Force Sensitive Resistor (FSR) sensor-based hand exoskeleton pressure sensors for hemiplegic rehabilitation, and glove-type pressure sensors to evaluate upper limb rehabilitation training or muscle stiffness and resistance [8]. Most of such previous studies measured pressure from the perspective of a patient, not from the perspective of a therapist.

Therefore, a therapeutic pressure sensor (TPS) was developed in this study with an aim to measure the finger pressure, quantify it, and give real-time feedback to the therapist during therapeutic sessions.

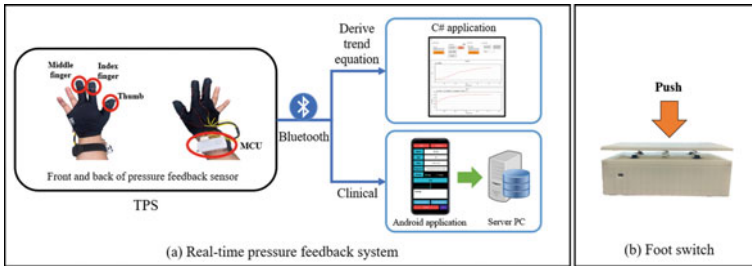


Fig. 1 Therapeutic pressure sensor (TPS) data collection system

2 Method

2.1 Glove-Type Real-Time Pressure Feedback System

A glove-type TPS was developed for the use in clinical setting for physical therapists. TPS was designed with FSR pressure sensors to be located on the finger pads of the thumb, index, and middle finger. For FSR pressure sensor, RA12P (Marveldex, Seongnam, Republic of Korea) was used and Jarduino-UNO-BTmini (JCnet, Daejeon, Republic of Korea) was used as the MCU for data collection, which was designed as a wristband as shown in Fig. 1.

An Android C#-based data collection application was developed to derive a trend equation that converts finger pressure in ADC unit to kilogram (SI unit). TPS transmitted pressure data of the three fingers (thumb, index, and middle) through Bluetooth communication, and the baud rate was set to 115,200. UI was designed for the therapist to select the treatment date from a calendar, patient’s name, age, gender, level of pain using Visual Analog Scale (VAS), symptoms, and diagnosis on the initial screen. In addition, a foot switch was developed for the therapist to step on at every change in the therapy area or postural change of the patient during a treatment session (see Fig. 1).

2.2 Experimental Protocol for Pressure Quantification

A lab-based experiment was conducted to convert the data of TPS collected as Analog–Digital Converter (ADC) value into the unit of kilogram. As a reference device, a commercial push–pull gauge MCT-1150 (A&D, Tokyo, Japan) was used and data were collected using the MCT-lite program (A&D software) through USB communication. Sampling rate for both TPS and push–pull gauge was set at 50 Hz, and the data were simultaneously collected. In order to derive a trend equation to convert ADC data to kg, a push–pull gauge was pressed vertically by the thumb,

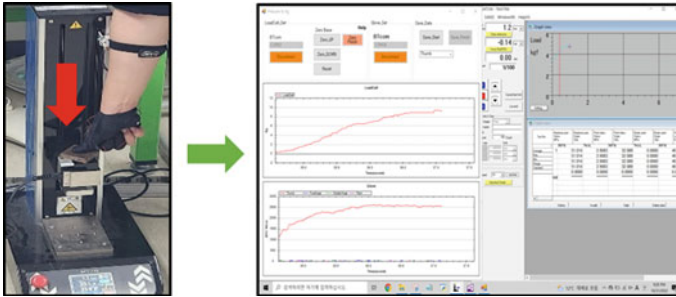


Fig. 2 Experimental environment of finger pressure measurement using push-pull gauge

index finger, and middle finger while wearing TPS (see Fig. 2). Pressure was maintained for 10 s while reaching 5 kg as the pressure value in kg was visibly provided. Resting time of 10 min was given between trials.

2.3 Data Analysis and Trend Equation

To extract kg units from push-pull gauge data collected in N ($\text{kg} \cdot \text{m/s}^2$), the analysis was conducted by dividing them by gravity acceleration g (9.8 m/s^2) and converting them into kg. Regression analysis was performed to derive a trend equation that converts ADC data of TPS into kg, and polynomial regression analysis was applied because finger pressure has nonlinear characteristics. Statistical analysis was performed using SPSS version 25.0.

2.4 Clinical Data Collection and Analysis

Clinical data of a 70-year-old female patient with a history of lumbar spine surgery and diagnosed with primary stage of bilateral gonarthrosis were used for analysis in this study. The patient's initial pain intensity on VAS was 5 [9]. A physical therapist with more than 11 years of clinical experience treated the patient by wearing TPS on both hands. A total of 24 treatment sessions were conducted from October 13, 2021, to November 25, 2021. For pressure analysis, data collected from 19 treatment sessions with more than 30 min of treatment time per session were used. Five of the datasets were excluded because the data had no foot switch markings for postural changes and treatment time of less than 20 min per session. Based on the treatment area and postural changes marked by the therapist, the ADC data of TPS were converted into kg using the trend equation for each finger. Peak detection of the pressure was obtained by MATLAB for the calculation of pressure average, standard deviation, the maximum and the minimum pressures applied through each finger.

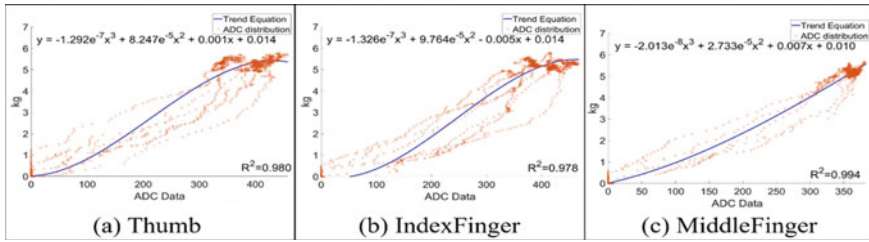


Fig. 3 Trend equation for each finger pressure of TPS

3 Results

3.1 Trend Equation

A tertiary trend equation was derived by regression analysis using the lab-based comparison experiment of repeatedly measuring each finger pressure up to 5 kg for three times in the thumb, index finger, and middle finger (see Fig. 3). The coefficient of determination (R^2) of the trend equation that converts ADC data of TPS to kg showed a high correlation in all three fingers: $R^2 = 0.980$ for the thumb, $R^2 = 0.978$ for the index finger, and $R^2 = 0.994$ for the middle finger.

3.2 Clinical Data

Table 1 shows the average and standard deviation, max, and min of each finger pressure for each treatment area. For gluteal muscle, the average pressure of the left finger was 1.48 ± 0.74 kg for the thumb, 1.12 ± 0.63 kg for the index finger, and 1.08 ± 0.79 kg for the middle finger. The average pressure of the right finger was 0.91 ± 0.43 kg for the thumb, 1.00 ± 0.76 kg for the index finger, and 0.80 ± 0.59 kg for the middle finger. For hamstring muscle, the average pressure of the left finger was 1.50 ± 1.48 kg for the thumb, 1.43 ± 0.82 kg for the index finger, and 1.28 ± 0.62 kg for the middle finger. The average pressure of the right finger was 0.92 ± 0.42 kg for the thumb, 0.71 ± 0.30 kg for the index finger, and 0.71 ± 0.19 kg for the middle finger.

4 Discussion

The purpose of this study was to develop a wearable manual pressure-quantifying sensor for physical therapists and to test its clinical feasibility. The results showed a decreasing trend of therapeutic pressure toward the end of treatment session (i.e.,

Table 1 Treatment pressure applied by thumb, index finger, and middle finger

Treatment area	Application of manual therapy		Mean \pm SD (kg)	Highest peak (kg)	Lowest peak (kg)
Gluteal muscle	Left hand	Thumb	1.48 \pm 0.74	3.17	0.32
		Index finger	1.12 \pm 0.63	2.91	0.36
		Middle finger	1.08 \pm 0.79	3.26	0.32
	Right hand	Thumb	0.91 \pm 0.43	1.84	0.33
		Index finger	1.00 \pm 0.76	2.83	0.31
		Middle finger	0.80 \pm 0.59	2.94	0.35
Hamstring muscle	Left hand	Thumb	1.50 \pm 1.48	5.15	0.49
		Index finger	1.43 \pm 0.82	3.44	0.40
		Middle finger	1.28 \pm 0.62	2.41	0.34
	Right hand	Thumb	0.92 \pm 0.42	1.81	0.37
		Index finger	0.71 \pm 0.30	1.09	0.33
		Middle finger	0.71 \pm 0.19	0.99	0.37

session #15/19) along with decreased level of pain on VAS. In terms of therapist's use of hand for manual therapy, the thumb was the most used finger compared to the index and middle finger. In addition, the therapist was applying higher pressure to treat hamstring muscle compared to the pressure treating gluteal muscle. These findings may build a basis for trend analysis in applying pressure during manual therapy.

Analyzing the treatment data of one patient using FSR sensor with a maximum measurement capacity of 5 kg for quantifying manual therapy pressure may be a limitation of this study. In future studies, larger volume of clinical data would need to be collected to analyze and establish therapeutic guidelines for young and inexperienced therapists. Data of different patient symptoms, diagnosis, age, gender, treatment area, and sessions as well as various therapeutic skills and experience of therapists need to be analyzed. In-depth analysis of the sequence and techniques of therapy and the effective and safe pressure ranges for each treatment area can be a very useful information for both clinicians and researchers.

5 Conclusion

In this paper, a therapeutic pressure sensor was developed with an aim to measure the finger pressure, quantify it, and give real-time feedback to the therapist during therapeutic sessions. Accumulation of clinical data by more use of TPS for manual therapy would build therapy guidelines, which can improve the quality of treatment for patients as well as the health and safety for therapists.

Acknowledgements This work was funded by BK21 FOUR (Fostering Outstanding Universities for Research) (No.: 5199990914048) and funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2021R111A3059769) and supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07050037).

References

1. Chang I, Rivera MJ, Eberman LE (2020) The effect of feedback on manual therapy skill acquisition: a systematic review. *Athl Train Educ J* 15(3):224–234
2. McArthur C, Ziebart C, Laprade J (2021) What do we know about spinal manual therapy for people with osteoporosis? A narrative review. *Phys Ther Rev* 26(1):42–52
3. Kim C-G et al (2016) Economic evaluation of manual therapy for musculoskeletal diseases: a protocol for a systematic review and narrative synthesis of evidence. *BMJ Open* 6(5)
4. Peters R, Schmitt MA, Verhagen AP, Pool-Goudzwaard AL, Mutsaers JH, Koes BW (2020) Comparing the range of musculoskeletal therapies applied by physical therapists with postgraduate qualifications in manual therapy in patients with non-specific neck pain with international guidelines and recommendations: an observational study. *Musculoskelet Sci Pract* 46:102069
5. Gagnon DH, Longtin C, Berbiche D, Gaudreault N (2016) Do experienced physiotherapists and final year physiotherapy trainees apply similar force during posterior-to-anterior lumbar mobilization techniques? *Man Ther* 21:287–291
6. Chaves P, Simões D, Paço M, Pinho F, Duarte JA, Ribeiro F (2020) Pressure applied during deep friction massage: characterization and relationship with time of onset of analgesia. *Appl Sci* 10(8):2705
7. Zhou G, Lu ML, Yu D (2023) Investigating gripping force during lifting tasks using a pressure sensing glove system. *Appl Ergon* 107:103917
8. Amit M, Chukoskie L, Skalsky AJ, Garudadri H, Ng TN (2020) Flexible pressure sensors for objective assessment of motor disorders. *Adv Func Mater* 30(20):1905241
9. Sterling M, Pedler A, Chan C, Puglisi M, Vuvan V, Vicenzino B (2010) Cervical lateral glide increases nociceptive flexion reflex threshold but not pressure or thermal pain thresholds in chronic whiplash associated disorders: a pilot randomised controlled trial. *Man Ther* 15(2):149–153

A Blockchain-Based CCP Data Integrity Auditing Protocol for Smart HACCP



Siwan Noh, Muhammad Firdaus, Zhuohao Qian, and Kyung-Hyune Rhee

Abstract CCP data integrity auditing in South Korea's smart HACCP system is currently only enabled through centralized access control, so its reliability cannot be assured. Since simply logging the history of data access does not provide a way to effectively verify the integrity of the data, so reliable HACCP systems must include the appropriate data integrity auditing schemes on their system. Many researches have been proposed to verify the integrity of data stored on semi-honest cloud servers, but to the best of our knowledge, no research has taken a smart HACCP environment into account. Therefore, in this paper, we propose a blockchain-based data integrity auditing protocol appropriate for contexts with a smart HACCP system. Moreover, we conduct qualitative analysis to show that the computational overhead in our protocol is reasonable. The analysis results show that our protocol will satisfy the requirements of the smart HACCP system regardless of the blockchain platform on which it is implemented.

Keywords Smart HACCP · Cloud storage service · Remote data integrity auditing · Blockchain

S. Noh · Z. Qian

Department of Information Security, Pukyong National University, Busan 48513, Republic of Korea

e-mail: nosiwan@pukyong.ac.kr

M. Firdaus

Department of Artificial Intelligence Convergence, Pukyong National University, Busan 48513, Republic of Korea

e-mail: mfirdaus@pukyong.ac.kr

K.-H. Rhee (✉)

Division of Computer Engineering, Pukyong National University, Busan 48513, Republic of Korea

e-mail: khrhee@pknu.ac.kr

1 Introduction

Hazard analysis and critical control points (HACCP) is an approach to systematically managing food safety by analyzing hazards (e.g., biological, chemical, and physical) and identifying critical control points (CCP) in the food production process. In South Korea, the ministry of food and drug safety issues HACCP certificates for certified food manufacturing and food sales companies. Traditional HACCP certification relied on data gathered manually by the production supervisor during the production process. However, there is a problem in that manually collected data can be fraudulently manipulated.

Smart HACCP [1] is developed to automate and digitize the production process through digital automation solutions such as IoT and CPS to ensure data reliability and provide collected data to third parties efficiently. Sensor data is collected on the server in the smart HACCP system by using a standard protocol as shown in Fig. 1. Only authorized users can access this server, and the server privately keeps a record of when data has been modified so that data forging in the system can be detected. However, this approach is solely based on trust in a database management system (DBMS) and does not consider third-party audits. Furthermore, when production supervisors use cloud services to archive CCP data, they lose direct control of their data, and when natural disasters and disk damage are considered, trusted DBMS cannot always guarantee data integrity [2]. Therefore, there is a need for a data integrity auditing mechanism that allows authorized third parties to remotely verify the integrity of the data stored in the server.

The rest of this paper is organized as follows. In Sect. 2, we give background knowledge, which can help readers understand this paper. In Sect. 3, we introduce our protocol for data integrity auditing, and qualitative analysis is presented in Sect. 4. Finally, Sect. 5 concludes the paper.

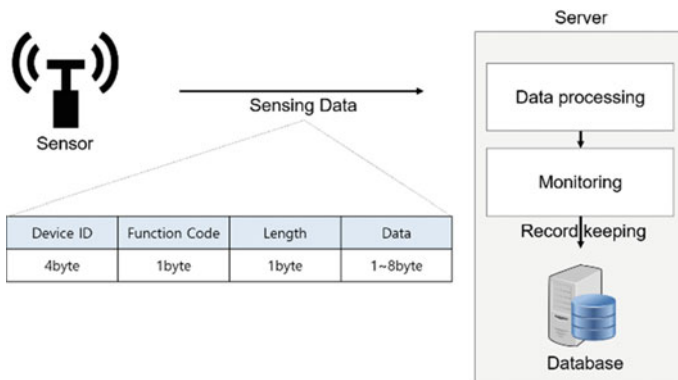


Fig. 1 Smart HACCP system architecture

2 Remote Data Integrity Auditing Scheme

Downloading the data to be verified is the most obvious method to verify the data integrity stored on external storage. In further detail, the user uses the hash-based message authentication code (HMAC) [3] to calculate the message authentication code (MAC) for the uploaded data and keeps this value before uploading it to the repository. To check the data's integrity, he calculates a new MAC from the downloaded data and compares it with the value he stored. However, since this approach requires the user to download the original data for verification, both storage space and communication costs are needed to store and download the data.

To address this issue, the remote data integrity auditing (RDIA) scheme [4] is proposed, which leverages a challenge-proof-verify mechanism between a verifier and a prover without requiring the entire original data to be retrieved. The RDIA process is as follows:

1. **Challenge:** The user chooses a random index i from one of the blocks split by the data to be verified and sends it to the server.
2. **Proof:** The server calculates a verifiable proof from the corresponding block and auxiliary information stored in the storage and returns it.
3. **Verify:** The user verifies this proof and checks data integrity.

However, there is a limit to a large amount of computation required on the user's side for verification every time, thus a Third Party Auditor (TPA)-based RDIA scheme [5], which delegates auditing tasks to a TPA, has been proposed. Regardless of that TPA can handle tasks more efficiently than resource-constraint users with sufficient computing resources, a single point of failure (SPOF) risk occurred caused to dependency on trusted TPA. A blockchain-based DIA (BDIA) scheme [6–9] has been proposed as a way to mitigate the single-point risk of TPA-based RIDA by replacing a blockchain with the trusted TPA.

Instead of the TPA, the prover in BDIA publishes the proof of the verifier's auditing request over the blockchain. The verifier leverages a cryptographic tag that has already been stored in the immutable blockchain to verify this proof [6–8], or the blockchain nodes verify the proof instead of the TPA, and the verifier checks the result of the consensus [9].

Although BDIA can guarantee the reliability of verification results and mitigate the risk of SPOF by eliminating the dependency on the trusted TPA, there are still many security issues [10], thus related research is ongoing.

3 Blockchain-Based Data Integrity Auditing Protocol

3.1 System Model

In this section, we define a smart HACCP system architecture we take into consideration before discussing in more detail about our protocol. Our smart HACCP system includes the following entities: *data owner*, *cloud service provider*, and *data user*.

- **Data owner DO** uploads the CCP data collected during the production process with the corresponding tag to the cloud server periodically as a production supervisor of the factory.
- **Data user DU** is an organization that uses CCP data stored on a cloud server for issuing HACCP certificates to a factory, for example. They utilize only the verified data and execute the integrity verification protocol before downloading data from the server.
- **Cloud service provider CSP** generates proof to verify the integrity of the corresponding data upon the DU's request and then publishes it to the blockchain.

Before we describe our protocol in detail, we assume the following.

- Before being uploaded to the cloud server, CCP data is encrypted using an appropriate encryption algorithm. That is, DO and DU already have established a shared secret key, and the proposed protocol intends to verify the integrity of the encrypted file.
- In the integrity verification process, we only assume a scenario in which the server is compromised. That is, we do not consider a scenario in which malicious users forge auditing results to claim unreasonable indemnity.

3.2 Proposed Protocol

Our protocol consists of two phases: Tag generation and Verification. The CSP generates and deploys a blockchain smart contract SC to support the protocol before initiating the phase. Every t period, the SC calculates $\text{Root}_t = \text{Hash}(\text{Root}_{t,1} \parallel \dots \parallel \text{Root}_{t,i})$ from the root of the Merkle Hash Tree (MHT) $\text{Root}_{t,i}$ submitted by the DOs and records these values on the blockchain.

Tag Generation. In the first step, DO_i uploads CCP data $\mathcal{D}_{t,i} = \{d_{t,i,1}, \dots, d_{t,i,n_i}\}$ collected during the production process to the CSP periodically (where t denotes the period within which the CCP data was collected.). In the second step, DO_i generates an MHT to verify the integrity of the CCP data uploaded in period t , as shown in Fig. 2. In the third step, all DOs in the smart HACCP system submit their root of MHT $\text{Root}_{t,i}$ to the SC. Finally, the SC keeps a record of the MHT root Root_t at period t along with the root of each DO's hash tree $\text{Root}_{t,i}$ (i.e., $\text{record}_t = (t, \text{Root}_t, \text{Root}_{t,1}, \dots, \text{Root}_{t,i})$).

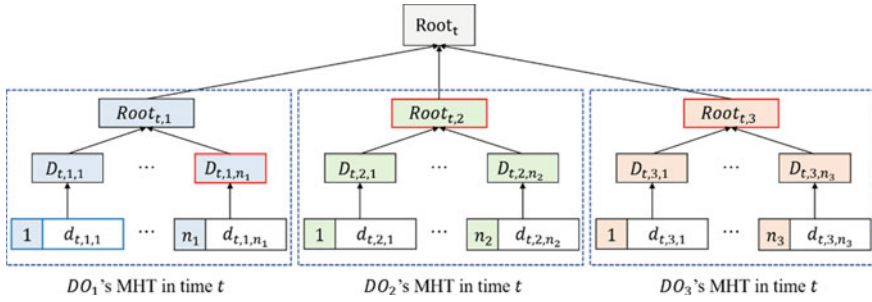


Fig. 2 Structure of Merkle Hash Tree

Verification. Before downloading CCP data $\mathcal{D}_{t,i}$ on the period t , the DU requests the CSP to verify its integrity as follows. In the first step, the DU randomly selects a leaf node k of the CCP data $\mathcal{D}_{t,i}$ to verify and sends the challenge number to the server. In the second step, to prove the challenge k , the server generates the audit path $path_{t,i,k}$ and calls the SC's verification function $verifyPath(k, path_{t,i,k})$. Figure 2 shows the audit path $path_{t,1,1}$ provided by the CSP when the DU selects $k = 1$ to verify CCP data $\mathcal{D}_{t,1}$ (where the blue border represents the challenge, and the red border represents the audit path). In the third step, SC calculates a new root value $Root'_t$ from the path and compares it with $Root_t$ that was recorded in the blockchain. If the two value $Root'_t$ and $Root_t$ are equal, the integrity of the CCP data $\mathcal{D}_{t,i}$ is guaranteed. Otherwise, the data's integrity has been violated. Finally, the DU checks the auditing results from the SC.

4 Discussion

In this section, we discuss how our protocol achieves data integrity auditing. In the proposed protocol, we make use of MHT and audit path to verify the integrity of CCP data stored on the semi-trusted CSP. The proposed protocol requires that each DO construct its own MHT, and SC keeps the hash of these MHT's roots to improve the efficiency of verification. Depending on the characteristics of the CCP data, the collecting period of each CCP data may vary. Therefore, in the worst case, the number of CCP data collected during the period in the factory may not be sufficient to construct the hash tree. That is, there is a possibility that the CSP may try to attempt a replay attack in this situation by repeating the previously submitted path. However, since the hash tree used for verification in the proposed protocol contains the MHT of other DOs, it can be easily detected if the CSP reuses the path. This strategy, however, may raise the computational burden on the CSP because it requires the reconstruction of MHTs for other DOs in addition to the MHT from the DO requested during the integrity verification process. We solve this problem by having SC store the MHT roots of all DOs $\{Root_{t,1}, \dots, Root_{t,i}\}$ collected in period t together. The number of

Table 1 Computational cost

Phase		
Entity	Tag Generation (every t period)	Verification (each request)
DO_i	$(n_i + 1)$ hash	–
SC	1hash	2hash
CSP	–	n_i hash
DU	–	–

hash operations required for entities in each phase of the proposed protocol is shown in Table 1.

As shown in Table 1, in the proposed method, we minimized the number of hash operations in the smart contract SC. The size of input data in the tag generation phase is approximately $256 \times i$ bits, and the size of the input data in the verification phase is $256 \times n_i$ bits, $256 \times i$ bits, respectively. Considering that the cost of hash operation on the Ethereum platform is determined by the size of the input data, the cost of computation in our protocol can be considered reasonable.

5 Conclusion

In this paper, we proposed a blockchain-based data integrity auditing protocol tailored to South Korea's smart HACCP system. We used an MHT to verify the integrity of the data stored on the cloud server without downloading the entire original data. Furthermore, we design the protocol to improve computational efficiency in smart HACCP system environments. The only data recorded in the blockchain in the proposed protocol is the root of MHTs, and the required number of hash operations is the minimum number of times, so the cost aspect of the protocol is also reasonable.

Acknowledgements This research was supported by the Republic of Korea's MSIT (Ministry of Science and ICT), under the ICT Convergence Industry Innovation Technology Development Project(2022-0-00614) supervised by the IITP and the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2022-2020-0-01797) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

References

1. Korea Agency of HACCP Accreditation and Services, Smart HACCP support portal. <https://fresh.haccp.or.kr/>. Accessed 20 Oct 2022
2. Sun Y, Zhang J, Xiong Y, Zhu G (2014) Data security and privacy in cloud computing 10(7). <https://doi.org/10.1155/2014/190903>

3. Deswarte Y, Quisquater J-J, Saïdane A, Remote integrity checking. Springer US, Boston, MA, pp 1–11
4. Ateniese G, Burns R, Curtmola R, Herring J, Kissner L, Peterson Z, Song D (2007) Provable data possession at untrusted stores. IACR ePrint 2007-202
5. Chavan NS, Sharma D (2018) Secure proof of retrievability system in cloud for data integrity, pp 1–5. <https://doi.org/10.1109/ICCUBEA.2018.8697842>
6. Li J, Wu J, Jiang G, Srikanthan T (2020) Blockchain-based public auditing for big data in cloud storage 57(6):102382. <https://doi.org/10.1016/j.ipm.2020.102382>
7. Yue D, Li R, Zhang Y, Tian W, Huang Y (2020) Blockchain-based verification framework for data integrity in edge-cloud storage 146:1–14. <https://doi.org/10.1016/j.jpdc.2020.06.007>
8. El Ghazouani M, El Kiram MA, Er-Rajy L (2019) Blockchain & multi-agent system: a new promising approach for cloud data integrity auditing with deduplication. Int J Commun Netw Inf Secur 11(1):174–184
9. Yu H, Yang Z, Sinnott RO (2019) Decentralized Big Data auditing for smart city environments leveraging blockchain technology, vol 7. Institute of Electrical and Electronics Engineers (IEEE), p 6288. <https://doi.org/10.1109/access.2018.2888940>
10. Han H, Fei S, Yan Z, Zhou X (2022) A survey on blockchain-based integrity auditing for cloud data. <https://doi.org/10.1016/j.dcan.2022.04.036>

BPFL: Blockchain-Enabled Distributed Edge Cluster for Personalized Federated Learning



Muhammad Firdaus, Siwan Noh , Zhuohao Qian ,
and Kyung-Hyune Rhee 

Abstract Due to the fact that the client's data distribution in federated learning (FL) is highly heterogeneous and leads to poor convergence, the concept of personalized federated learning (PFL) is on the rise. PFL aims to tackle the effect of non-IID data and statistical heterogeneity problems along with achieving rapid model convergence and personalized models. Moreover, in the context of clustering-based PFL, it allows a multi-model method with group-level client relationships to accomplish personalization. However, since the current method still relies on the centralized method, where the server orchestrates all processes, this paper introduces a blockchain-enabled distributed edge cluster for PFL (BPFL) that exploits the benefits of blockchain and edge computing to address these shortcomings. Blockchain can be employed to enhance client privacy and security by recording all transactions in immutable distributed ledger networks as well as improving efficient client selection and clustering. Furthermore, an edge computing system offers reliable storage and computation, where computational processing is locally performed in the edge infrastructure to be closer to clients. Thus, it also improves real-time services and low-latency communication of PFL.

Keywords Personalized federated learning · Non-IID data · Blockchain · Edge computing · Client clustering

M. Firdaus
Department of Artificial Intelligence Convergence, Pukyong National University, Busan 48513,
Republic of Korea
e-mail: mfirdaus@pukyong.ac.kr

S. Noh · Z. Qian
Department of Information Security, Pukyong National University, Busan 48513, Republic of
Korea
e-mail: nosiwan@pukyong.ac.kr

K.-H. Rhee (✉)
Division of Computer Engineering, Pukyong National University, Busan 48513, Republic of
Korea
e-mail: khrhee@pknu.ac.kr

1 Introduction

With the recent establishment of data privacy preservation regulations, such as the general data protection regulation (GDPR) [1] and the health insurance portability and accountability act (HIPAA) [2], the need for privacy-preserving AI has been continuously growing. Hence, federated learning (FL), which allows clients to train their data collaboratively without exposing private data to each other, is gaining more widespread lately and has arisen as a promising approach to tackle the privacy-preserving AI issues. The introduction of FL firstly has been used by Google for Gboard's next-word prediction [3]. Then, following the success of Gboard, several works empower FL utilization in myriad applications, such as health care [4], Industrial Internet of Things (IIoT) [5], vehicular networks [6], finance [7], and so forth.

Although FL offers some benefits, the general FL approach, i.e., federated averaging (FedAvg), resulting poor convergence since the client's data distribution is highly heterogeneous and typically a non-independent identically distributed (non-IID). Thus, the idea of personalized federated learning (PFL) was developed to address the effects of non-IID data and statistical heterogeneity issues [8] and converge models and create customized models quickly. Furthermore, in the context of clustering-based PFL [9], it enables personalization using a multi-model approach with group-level client interactions. In this sense, by mapping client associations, this approach allows each client learns a personalized model with their affiliated clients in a similar cluster.

However, the existing cluster-based PFL method still depends on the centralized strategy where the server orchestrates all processes and requires high communication and computation cost. In order to address these shortcomings, this paper introduces a blockchain-enabled distributed edge cluster for PFL (BPFL) that exploits the benefits of blockchain and edge computing. Blockchain can enrich client privacy and security by recording all transactions in immutable distributed ledger networks and enhancing efficient client selection and clustering. Likewise, an edge computing system offers appropriate storage and computation, where computational processing is locally performed in the edge infrastructure to be nearer to clients. Thus, it provides a proper computation capability and improves real-time services and low-latency communication of PFL.

The structure of this paper is arranged as follows: We explain the current works related to personalized federated learning in Sect. 2. Next, Sect. 3 presents our proposed model, a distributed edge cluster for PFL. We discuss several remarks and concerns in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 Related Work

Lately, PFL research is becoming prevalent since it proposed to overcome one of the primary challenges of current FL, i.e., statistical heterogeneity due to the effect of non-IID data distribution among clients. In 2017, one of the initial works on PFL was started by proposing MOCHA with a multi-task learning approach [8]. MOCHA is an algorithm that simultaneously learns client task settings and a similarity matrix that investigates the issue of federated multi-task learning. The distributed multi-task MOCHA handles learning problems in several ways, including fault tolerance, stragglers, and communication limitations. Moreover, the authors in [9] considered the context-based client clustering approach to facilitate multiple global models and be flexible enough to manage inherent client populations over time. In this approach, the FL model is trained for each homogeneous group of clients, where the clients are divided into clusters using an ideal partitioning approach based on the cosine similarity of the gradient updates from the clients. Additionally, there are several methods for realizing the PFL framework, such as data augmentation [10], meta-learning [11], transfer learning [12], fine-tuning [13], and so on. However, context-based client clustering with a multi-task approach is the focus of this paper.

3 Distributed Edge Cluster for PFL

In this section, we present the BPFL approach that leverages the concept of distributed edge clusters in realizing personalized federated learning by empowering the merits of edge computing and blockchain technology. Due to the misassumption that a single global model can accommodate all clients, this approach allows a multi-model method with group-level client associations to achieve personalization. By employing blockchain, these models can be stored in a distributed manner with privacy-preserving protection. Moreover, blockchain with smart contracts can be used to evaluate client relationships to improve efficient client selection and clustering. Meanwhile, in order to reduce the communication and computation cost in the client clustering processes, edge computing servers are counted to provide suitable storage, communication, and computation capabilities. In this sense, the computational processing is locally conducted in the edge infrastructure to be closer to clients as data providers.

Figure 1 shows our proposed model composed of several procedures, i.e., system initialization, collaborative edge cluster establishment, personalized local training models, and personalized global model aggregation. The detailed procedures are described as follows.

System initialization. The initial model parameters of the global model (θ^{in}) are stored in distributed ledger blockchain (which can be integrated with off-chain storage, e.g., IPFS) as the initial learning model process. It is maintained by edge servers that are placed throughout clusters. Then, the number of K clients (C_k) need

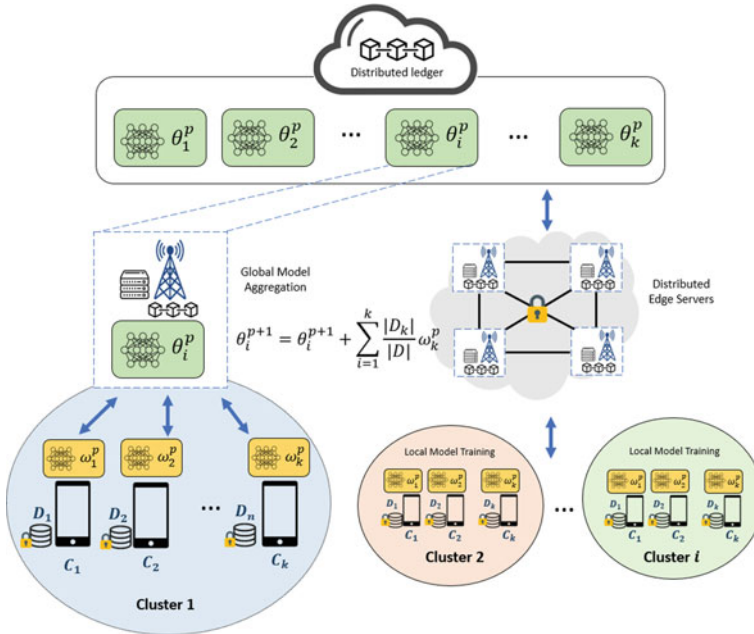


Fig. 1 Distributed edge cluster for personalized federated learning

to register to the BPFL system by sending asset statement (refers to [14]) validation with their pseudo-public key address. In sort, only validated clients can access the system.

Collaborative edge cluster establishment. In this step, BPFL can distinguish incongruent clients after converged to a stationary point [9] as well as generates multi-model method (θ_k) that maintained by distributed edge servers. Hence, according to the client’s asset statement, the collaborative edge clusters (ECL_i) are established based on client’s data distribution (D_k) similarity, where their datasets are non-IID. In this sense, the group-level client similarity can be counted using specific partition techniques, such as Euclidean distance, Cosine similarity, and Gaussian mixture.

Personalized local training models. In the p -th iteration, C_k in each ECL_i downloads global parameter θ_i^p from blockchain and performs local training using their datasets D_k to obtain personalized model (w_k^p) owned by client C_k by following formula [15]:

$$w_k^p = \arg \min_{w \in \mathbb{R}^d} F_i(w) + \frac{\mu}{2\alpha_p} \|w - \theta_k^p\| \tag{1}$$

where F_i is a loss function associated with i -th data point on C_k , $w \in \mathbb{R}^d$ encodes the parameters of local model w , μ represents regularization parameter, and α_p is the step size of gradient descent in iteration p .

Personalized global model aggregation. After the clients collaboratively upload their w_k^p to distributed edge cluster, the cluster global model aggregation is executed as follows:

$$\theta_i^{p+1} = \theta_i^p + \sum_{k=1}^k \frac{|D_k|}{|D|} w_k^p \quad (2)$$

where θ_i^{p+1} is a new personalized global model that obtained for the next iteration ($p + 1$). In this step, all participants in the BPFL system can download θ_i^{p+1} through distributed ledger blockchain. As a result, iteration continues until the model achieves precise accuracy or reaches the maximum number of iterations.

4 Remarks and Discussion

The study of PFL is earning vogue since it addresses statistical heterogeneity from non-IID data distribution as one of the primary challenges of the existing FL approach. Even though the PFL approach is advantageous compared with the general FL framework (i.e., FedAvg), there are some significant challenges, especially in the client’s data privacy risk during the model training process. Therefore, to enhance security and privacy in PFL, we introduce BPFL, which exploits the concept of distributed edge clusters that leverage the merits of edge computing and blockchain technology. Table 1 summarizes the advantages of BPFL compared to current PFL clustering techniques theoretically.

Nevertheless, to strengthen and achieve a robust PFL technique, further work is required to investigate its potential attacks and defenses due to the more complex protocols and structures of PFL. Also, developing a reliable incentive mechanism to maintain fairness and motivate clients’ contributions is feasible to study. On the other hand, representative datasets are essential for developing the PFL field. Datasets with more modalities (e.g., sensor signals, video, and audio) and involving a more comprehensive assortment of machine learning tasks from practical applications are

Table 1 Summary of BPFL advantages compared to current PFL techniques

Key parameters	Smith et al. [8]	Sattler et al. [9]	This work
Computation and communication cost	High	High	Low
Works on non-convex setting	No	Yes	Yes
Multiple global model setting	Not investigated	Not investigated	Yes
Distributed edge cluster	Not applied	Not applied	Yes
Privacy guarantees	Not investigated	Yes	Yes
Asset statement	Not applied	Not applied	Yes
Incentive scheme	Not applied	Not applied	Applied

required to feed PFL research. Additionally, performance benchmarking is another critical aspect for the long-term expansion of the PFL research domain.

5 Conclusion

This paper introduced a blockchain-enabled distributed edge cluster approach for PFL (BPFL) that exploits the benefits of blockchain and edge computing. We consider utilizing blockchain to protect client privacy and security by recording all transactions in immutable distributed ledger networks and enhancing efficient client selection and clustering. Likewise, an edge computing system offers appropriate storage and computation, where computational processing is locally performed in the edge infrastructure to be nearer to clients. Thus, it provides a proper computation capability and improves real-time services and low-latency communication of PFL. Additionally, we remarked privacy-preserving, trustworthy PFL, as well as representative datasets and benchmarks, as several challenges for future study directions of the PFL field.

Acknowledgements This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01797) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and partially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1I1A3046590).

References

1. Voigt P, Von dem Bussche A (2017) The EU general data protection regulation (gdpr). In: A practical guide, 1st edn, Springer International Publishing, Cham, 10.3152676, 10-5555
2. Annas GJ (2003) HIPAA regulations: a new era of medical-record privacy? *N Engl J Med* 348:1486
3. Hard A et al (2018) Federated learning for mobile keyboard prediction. arXiv preprint [arXiv:1811.03604](https://arxiv.org/abs/1811.03604)
4. Rieke N et al (2020) The future of digital health with federated learning. *NPJ Dig. Med.* 3(1):1–7
5. Lu Y et al (2019) Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Trans Ind Inf* 16(6):4177–4186
6. Samarakoon S et al (2019) Distributed federated learning for ultra-reliable low-latency vehicular communications. *IEEE Trans Commun* 68(2):1146–1159
7. Long G et al (2020) Federated learning for open banking. In: *Federated learning*. Springer, Cham, pp 240–254
8. Smith V et al (2017) Federated multi-task learning. *Adv Neural Inf Process Syst* 30
9. Sattler F, Müller K-R, Samek W (2020) Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints. *IEEE Trans Neural Netw Learn Syst* 32(8):3710–3722
10. Zhao Y et al (2018) Federated learning with non-iid data. arXiv preprint [arXiv:1806.00582](https://arxiv.org/abs/1806.00582)

11. Jiang Y et al (2019) Improving federated learning personalization via model agnostic meta learning. arXiv preprint [arXiv:1909.12488](https://arxiv.org/abs/1909.12488)
12. Chen Y et al (2020) Fedhealth: a federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems* 35(4):83–93
13. Li T et al (2020) Federated optimization in heterogeneous networks. *Proc Mach Learn Syst* 2:429–450
14. Weng J et al (2019) Deepchain: auditable and privacy-preserving deep learning with blockchain-based incentive. *IEEE Trans Dependable and Secure Comput* 18(5): 2438–2455
15. Huang Y et al (2021) Personalized cross-silo federated learning on non-IID data. In: *AAAI*, 2021

Healthcare Big Data Analysis Using Representative National Community Health Survey of 2021: Is Income Change Due to Covid-19 Pandemic Associated with Unmet Healthcare Needs in S. Korea



Yeong Jun Lee, Se Hyeon Myeong, and Sun Jung Kim 

Abstract This paper was to explore how income changes due to Covid-19 associated with unmet healthcare needs among Korean population. We used representative National Community Health Survey 2021, and 209,202 respondents were included as final study subjects. We examined the sample characteristics and the association between income change due to Covid-19 and unmet healthcare needs. The multivariate survey logistic regression model was used to identify the association. Descriptive statistics showed that population with decreased income due to Covid-19 had more unmet healthcare needs than not changed group. The survey logistic regression results also showed that odds of having unmet healthcare needs were $OR = 1.301$ (95% CI: 1.229–1.377) than not changed group. We assume that the main reason for the unmet healthcare needs disparity is economic reason. Further study should be done to find out the association. Result of this study suggested that we need understanding where at risk populations are and identifying underserved healthcare needs. Our research provides evidence of promoting targeted financial support, preventive programs, and resource allocations to those underserved population groups.

Keywords Unmet healthcare needs · Covid-19 · Income changes · Disparity

1 Introduction

Easy access and utilization of healthcare is not only a fundamental solution to patients' health problems, but also an essential element for ensuring their quality of

Y. J. Lee · S. H. Myeong · S. J. Kim (✉)

Department of Health Administration and Management, College of Medical Science, Soonchunhyang University, Asan, Republic of Korea
e-mail: sunjkim0623@sch.ac.kr

S. J. Kim

Center for Healthcare Management Science, Soonchunhyang University, Asan, Republic of Korea

Department of Software Convergence, Soonchunhyang University, Asan, Republic of Korea

life [1]. Accordingly, S. Korea has prepared several institutional arrangements, such as improving accessibility to medical utilization and expanding medical coverage to bridge the gap [2]. Since the introduction of health insurance in 1977, Korea has established national health insurance in a short period of 12 years, and through this, it has promoted the expansion of health insurance coverage so that all citizens can receive necessary medical services regardless of health condition or ability [3, 4]. However, health insurance's low coverage issues such as increased non-benefit and high out-of-pocket expenses have been raised [5]. Especially those with low income status, the elderly, the disabled, and chronically ill, experiencing unmet healthcare needs due to a number of factors, although adequate healthcare needs to be provided [6, 7].

Unmet healthcare needs refer to a state in which medical needs are not met in a timely manner, and the subject feels necessary, or the medical professional does not receive services even though the medical professional has determined the needs [8, 9]. The problem of equity related to the utilization of healthcare services appears in various forms. Inequality has been reported several times, such as the inverse care law, in which new health services are applied in the order of higher socioeconomic levels, the inverse care law, in which low-income areas lack medical resources, and the inverse prevention law, in which higher socioeconomic status use preventive care such as health checks [10]. This pattern has also been reported in the UK, Canada, New Zealand, and Australia, where health care is operated based on tax-based systems [11]. In addition, the factors of occurrence of unmet healthcare needs can be largely divided into availability, accessibility, and acceptability. Availability occurs because the necessary service or available service is not provided at the desired time or place. Accessibility refers to factors resulting from economic burdens or transportation problems, including medical costs, and acceptability refers to low acceptance due to the patient's lack of time to receive services, lack of service, or lack of information about service providers [12].

As follows, the causes of unmet healthcare needs appear in various aspects, but the biggest cause is considered an economic burden [13]. Specifically, looking at previous domestic studies related to unmet healthcare needs, socioeconomic status such as income, education status, and employment status is pointed out as the decisive causes of unmet healthcare needs. In particular, the occurrence of unmet healthcare needs was different depending on the income status of individuals or households, and it is known that the group with low income status has a high experience of unmet healthcare needs [14–16]. As such, the relationship between income status and unmet healthcare needs has been reported through many domestic studies, but no studies have been reported on how the income status affected by the special situation of COVID-19 is related to unmet healthcare needs. Therefore, this study aims to establish basic data to check the level of unmet healthcare needs by using raw data from the 2021 Community Health Survey to identify which subjects are mainly unmet healthcare needs according to differences in socioeconomic characteristics and health behavior characteristics.

2 Method and Materials

2.1 Data and Study Subject

This study analyzes the relationship between changes in total income caused by COVID-19 and the occurrence of unmet healthcare needs using raw data from the 2021 Community Health Survey published in August 2022. The Community Health Survey is a statistical survey approved by the National Health Center every year since 2008 to establish and evaluate local healthcare plans based on Article 4 of the Local Health Act and Article 2 of the Enforcement Decree of the same Act.

The subjects of the 2021 Community Health Survey were adults aged 19 or older living in sample households at the time of the survey, and 229,242 subjects were surveyed through the process of analysis and stratification of the population, distribution, and extraction of samples. In this study, out of 229,242 survey subjects, 209,202 were the final study subjects, excluding those who did not respond to research-related items.

2.2 Variables

The survey items and contents of the variables used in the study are specifically presented in the guidelines for using raw data for community health surveys. The variables used in this study are as follows.

(1) Dependent variable

The dependent variable of this study is whether unmet healthcare needs occurs annually, and each was classified into “occurrence” and “not occurrence” according to the response of the data source. When asked, “Have you ever needed a hospital (excluding dentistry) treatment (examination or treatment) in the past year?” but have you not received it?” If you answered “Yes,” the annual occurrence of unmet healthcare needs was “occurring,” if you answered “no,” if you answered “no,” the annual occurrence of unmet healthcare needs was classified as “unoccurring.”

(2) Independent variable

The independent variable of this study is the change in total income related to COVID-19, and each was classified according to the response of the data source.

When asked, Has there been a change in the total income of households compared to before the COVID-19 epidemic (January 2020), the response variables on the three scales, “decreased,” “unchanged,” and “increased,” were used as independent variables.

(3) Control variables

The control variables of this study are variables that related to demographic characteristics, socioeconomic characteristics, and health behavior that can affect the occurrence of unmet healthcare needs. The demographic and socioeconomic variables of the survey subjects included age, gender, residential area, education level, economic activity, marital status, and income level based on previous study [17]. Variables related to the health behavior of the survey subjects included subjective health level and subjective stress level. Both subjective health level and subjective stress level were classified according to the raw data response variable.

2.3 Statistical Analysis

Descriptive statistics were conducted for each variable to indicate the frequency (n) and percentage (%) of variables related to demographic, socioeconomic, and health behavior of the study subjects, and a survey chi-square test was conducted to determine whether the frequency difference between each variable and the annual unmet healthcare needs was statistically significant. In addition, a univariate and multivariate survey logistic regression analysis was conducted to determine the association between the change in total income due to COVID-19 and the occurrence of annual unmet healthcare needs. And to determine the cause of the high annual unmet healthcare needs in households with increased total income, the causes of healthcare needs were classified into six subgroups: “lack of time,” “weak symptoms,” “economic reason,” “inconvenient traffic,” “Covid-19,” and “other,” and a survey chi-square test and multivariate survey logistic regression analysis were conducted. All the analysis shown in the study was conducted through the statistical program SAS ver. 9.4 (SAS Institute Inc., Cary, NC, USA).

3 Results

3.1 Descriptive Statistics

The frequency (n) and percentage (%) for each variable by annual unmet healthcare needs are as follows: Table 1. Among the households whose total income decreased due to COVID-19, 880,456 (6.17%) had unmet healthcare needs annually. Among households with the same total income, 1,019,364 (4.40%) had unmet healthcare needs annually. Among the households whose total income increased due to COVID-19, 65,357 (5.91%) had unmet healthcare needs annually. As a result of examining the annual unmet healthcare needs according to the characteristics of the survey subjects, it was found that the lower the education level, the lower the subjective

Table 1 Characteristics of survey subjects according to unmet healthcare needs

Variables	Unmet healthcare needs (sample $n = 206,321$ /Weighted $n = 38,544,331$)				
	Yes		No		P-value
	N/MEAN	%/SD	N/MEAN	%/SD	
<i>Income changes</i>					
Decreased	880,456	6.17	13,382,382	93.83	< 0.0001
Unchanged	1,019,364	4.40	22,155,830	95.60	
Increased	65,357	5.91	1,040,943	94.09	
<i>Age (years)</i>					
≤ 44	804,074	.36	14,184,319	94.64	< 0.0001
45–64	798,201	5.30	14,254,416	94.70	
65–74	204,099	4.12	4,747,768	95.88	
75+	158,802	4.47	3,392,652	95.53	
<i>Sex</i>					
Male	807,114	4.31	17,919,197	95.69	< 0.0001
Female	1,158,062	5.84	18,659,958	94.16	
<i>Region</i>					
Metropolitan	786,727	4.86	15,388,630	95.14	< 0.0001
City	760,929	5.05	14,316,766	94.95	
Region	417,521	5.73	6,873,759	94.27	
<i>Education level</i>					
Univ. & Higher	627,447	4.34	13,828,418	95.66	< 0.0001
2/3 years college	285,262	5.47	4,933,974	94.53	
Belo High school	1,052,467	5.58	17,816,763	94.42	
<i>Subjective health</i>					
Very good	79,965	3.05	2,542,076	96.95	< 0.0001
Good	482,029	3.39	13,728,791	96.61	
Normal	914,683	5.50	15,713,711	94.50	
Bad	383,238	8.94	3,902,237	91.06	
Very bad	105,262	13.20	692,340	86.80	
<i>Economic activity</i>					
No	721,141	5.09	13,459,200	94.91	0.8750
Yes	1,244,036	5.11	23,119,955	94.89	
<i>Marital status</i>					
Unmarried	420,810	4.87	8,222,094	95.13	< 0.0001
Married	1,158,959	4.83	22,842,159	95.17	
Divorce/separated	385,407	6.53	5,514,902	93.47	

(continued)

Table 1 (continued)

Variables	Unmet healthcare needs (sample $n = 206,321$ /Weighted $n = 38,544,331$)				
	Yes		No		P-value
	N/MEAN	%/SD	N/MEAN	%/SD	
<i>Income level</i>					
Q1 (lowest)	399,539	6.66	5,596,999	93.34	< 0.0001
2	581,401	5.25	10,491,557	94.75	
3	553,775	4.93	10,676,547	95.07	
Q4 (highest)	430,460	4.20	9,814,051	95.80	
<i>Subjective stress level</i>					
Very much	207,176	14.45	1,227,010	85.55	< 0.0001
Much	669,199	8.12	7,570,242	91.88	
Little bit	871,249	4.10	20,376,505	95.90	
Hardly	217,552	2.85	7,405,398	97.15	

health level, the lower the income level, and the higher the subjective stress level has the higher the frequency of annual unmet healthcare needs.

3.2 Multivariate Analysis Using Survey Logistic Analysis

The association between each variable and the annual unmet healthcare needs is as follows Table 2. The odds ratio of annual unmet healthcare needs was 1.301 for households with “decreased” changes in total income due to COVID-19 compared to “Unchanged” households, and 1.277 for “increased” households. Compared to the case where the age was “75 years old or older,” the odds ratio of “65–74 years old” was 1.106 times, “45–64 years old” was 1.727 times, and “44 years old or younger” was 2.155 times, and the lower the age, the more unmet healthcare needs occurred annually. In the case of gender, the odds ratio of annual unmet healthcare needs in “female” was 1.248 times more than in “male,” and in the case of residential areas, the odds ratio of annual unmet healthcare needs in “region” was 1.141 times more than in Metropolitan. Compared to the case of “Higher than 4 years university” in the level of education, the odds ratio of “2/3 years University” was 1.105 times, and the odds ratio of “Less than high school” was 1.187 times. The lower the level of education, the more unmet healthcare needs occurred annually. Compared to “very good” for subjective health level, the odds ratio of “good” annual unmet healthcare needs was not statistically significant, but “normal” was 1.636 times; “bad” was 2.261 times, and “very bad” was 3.813 times, and the lower the subjective health level, the more unmet healthcare needs occurred annually. In addition, compared to the income level “Q4” (highest), the odds ratio of “Q3” was 1.108 times; “Q2”

Table 2 Multivariate analysis of unmet healthcare needs

Variables	Annual unmet healthcare needs	
	Odds ratio	95% CI
<i>Income change</i>		
Decreased	1.301	1.229–1.377
Unchanged	1.000	–
Increased	1.277	1.101–1.481
<i>Age (years)</i>		
≤ 44	2.155	1.914–2.427
45–64	1.727	1.564–1.906
65–74	1.106	1.000–1.222
75+	1.000	–
<i>Sex</i>		
Male	1.000	–
Female	1.248	1.181–1.319
<i>Region</i>		
Metropolitan	1.000	–
City	1.020	0.956–1.089
Region	1.141	1.069–1.218
<i>Education level</i>		
Univ. & higher	1.000	–
2/3 years college	1.105	1.014–1.203
Below high school	1.187	1.103–1.277
<i>Subjective health level</i>		
Very good	1.000	–
Good	1.094	0.955–1.254
Normal	1.636	1.432–1.870
Bad	2.621	2.270–3.027
Very bad	3.813	3.205–4.537
<i>Economic activity</i>		
No	1.000	–
Yes	1.119	1.053–1.190
<i>Marital status</i>		
Unmarried	0.868	0.799–0.943
Married	1.000	–
Divorce/Separated	1.213	1.132–1.299
<i>Income level</i>		
Q1 (lowest)	1.520	1.378–1.676
2	1.181	1.086–1.284

(continued)

Table 2 (continued)

Variables	Annual unmet healthcare needs	
	Odds ratio	95% CI
3	1.108	1.020–1.205
Q4 (highest)	1.000	–
<i>Subjective stress level</i>		
Very much	3.954	3.521–4.440
Much	2.364	2.174–2.572
Little bit	1.344	1.241–1.456
Hardly	1.000	–

was 1.181 times; and “Q1” (lowest) was 1.52 times, and the lower the income level, the more unmet healthcare needs occurred annually. Finally, at the subjective stress level, compared to “Hardly,” the odds ratio of “Little bit” was 1.3444 times; “Much” was 2.364 times, and “Very much” was 3.954 times. The higher the subjective stress level, the more unmet healthcare needs occurred annually.

4 Future Research Direction

We will explore the association between households whose income has decreased due to COVID-19 and the occurrence of unmet healthcare needs due to economic reasons, and the cause of the high annual incidence of unmet healthcare needs for households whose income has increased due to COVID-19 in the future research.

References

1. Shin J (2013) Determinants of unmet healthcare needs among the community dwelling Korean adults with disabilities: a case of the National Health Insurance beneficiaries. *Health Soc Welf Rev.* 33(2):549–577
2. Kim WJ et al (2020) Analysis of dental unmet needs medical research trends and influence factors: using structural literature review. *J Korean Soc Dental Hygiene* 20(3):243–256
3. Hwang BD, Choi R (2015) The prevalence and association factors of unmet medical needs by age group in the elderly. *The Korean journal of health service management.* 9(1):81–93
4. Lee SA, Park EC (2017) Main indicators of National Health Insurance during 40 years. *Health Policy and Management* 27(3):267–271
5. Shin SS (2017) Election pledge and policy tasks of president Moon Jae-in in healthcare sector. *Health Policy and Management* 27(2):97–102
6. Song HY, Choi JW, Park EC (2015) The effect of economic participatory change on unmet needs of health care among Korean adults. *Health Policy and Management* 25(1):11–21
7. Choi R, Hwang BD (2016) Analysis of unmet medical needs according to mental health. *The Korean Journal of Health Service Management.* 10(1):117–129
8. Aday LA, Andersen R (1974) A framework for the study of access to medical care. *Health Serv Res* 9(3):208

9. Heo SI, Kim MG, Lee SH, Kim SJ. A study for unmet health care need and policy implications. Korea Institute for Health and Social Affairs. 2009.
10. Victora CG, Vaughan JP, Barros FC, Silva AC, Tomasi E (2000) Explaining trends in inequities: evidence from Brazilian child health studies. *The Lancet* 356(9235):1093–1098
11. Hwang J, Rudnisky C, Bowen S, Johnson JA (2017) Measuring socioeconomic inequalities in eye care services among patients with diabetes in Alberta, Canada, 1995–2009. *Diabetes Res Clin Pract* 127:205–211
12. Chen J, Hou F (2002) Unmet needs for health care. *Health Rep* 13(2):23–34
13. Kim SA, Seo Y, Woo KS, Shin YJ (2019) A systematic review of studies on current status and influencing factors of unmet medical needs in Korea. *J. Crit. Soc. Welf.* 62:53–92
14. Huh SI, Kim SJ (2007) Unmet needs for health care among Korean adults: differences across age groups. *Korean J Health Econ Policy.* 13(2):1–16
15. Shin YJ, Shon JI (2009) The prevalence and association factors of unmet medical need-using the 1st and 2nd Korea welfare panel data. *Health Soc Welf Rev.* 29(1):111–142
16. Song KS, Lee JH, Rhim KH (2011) Factors associated with unmet needs for health care. *Korean Public Health Res.* 37(1):131–140
17. Han KT, Park EC, Kim SJ (2016) Unmet healthcare needs and community health center utilization among the low-income population based on a nationwide community health survey. *Health Policy* 120(6):630–637

River Health Assessment Based on an Artificial Neural Network



Zonglin Pang, Yufeng Liu, Zitong Liu, and Chao Liu

Abstract River ecological health evaluation is a complex problem, and although there are many existing river evaluation methods, it is still a challenge to synthesize them and better express the nonlinear relationships among multiple influencing factors of river ecosystems. In this paper, we try to evaluate the ecological health of Dagu River using artificial neural network, and train and validate the designed network to establish the corresponding river health evaluation model of Dagu River. The analysis of the results shows that the mean absolute error (MAE) of the test data is 4.98, and as a comparison, the MAE is 8.125 when multiple linear regression is chosen as the baseline algorithm, indicating that the artificial neural network can accurately evaluate the river health condition of Dagu River. By using artificial neural network to evaluate the river health, the method has wide practicability and provides the basis for the credibility of river health.

Keywords River eco-system · Health assessment · Knowledge elicitation · Artificial neural network

1 Introduction

With the extensive expansion of urban scale, the natural ecological environment of rivers has been greatly disturbed by the intervention of human activities, which has changed the original function of rivers and led to the degeneration of water system, black and smelly water, and other water ecological environment problems. River health assessment has both ecological and economic significance for the development of society, yet rivers are a fuzzy [1], nonlinear, and complex ecosystem [2]. The traditional method could linearly express the relationship between indicators and evaluation results, while the evaluation of river ecosystem health is extremely

Z. Pang · Y. Liu · Z. Liu · C. Liu (✉)
Qingdao University of Technology, Qingdao 266520, China
e-mail: Liuchao@qut.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_59

449

complex and does not reflect the real situation of the river well [3, 4], Such as the multi-step grey correlation evaluation method.

Artificial neural network (ANN) is an advanced river health evaluation method with good adaptability, self-organization, and strong learning ability, which can be used to describe the nonlinear relationships contained in data samples and evaluate the health of river ecosystems [5]. The artificial neural networks have a strong ability to synthesize information, which can handle both quantitative and qualitative information and coordinate various input information well [6]. The artificial neural network is suitable for handling complex nonlinear and uncertain objects. Anmala et al. [7] studied the application results of the nonlinear regression model and artificial neural network model in the upper Green River Basin in Kentucky, USA. The results show that the neural network model is more accurate than the nonlinear regression model in predicting water quality parameters. This confirms the superiority of neural networks in dealing with nonlinear problems.

The Dagu River is located in the Jiaodong Peninsula of China and flows through Qingdao. It plays a crucial role in the socio-economic development of Qingdao city. Since the river flows through both urban and rural areas, the river water is vulnerable to pollution, so a study to evaluate the river health of Dagu River is necessary and important.

2 Materials and Methods

2.1 Study Area

In the study, Dagu River in the Jiaodong Peninsula of China is chosen as research object, which flows from the northeast to the southwest through the Qingdao. It is located between 36.2° to 37.1°N and 120.1° to 120.5°E, with 137 km long and an overall area of 4781 km². The north is mountainous and shallow hills, and the south is plain and depression. The terrain slope gradually becomes slow from north to south. The mean annual temperature of the Dagu River is 12.7 °C, and the mean annual rainfall is 684.0 mm. Rainfall is mainly concentrated in the summer, with heavy rains. The location of the Dagu River is shown in Fig. 1.

2.2 Data Collection and Analytical Methods

Hydrological, morphological, chemical, biological, and social services sustainability survey data were collected in 2017 as objective data for this study. The collected information was provided to ten experts for assessment in order to gain knowledge. Ten experts in the field of river hydrology and water pollution control were selected

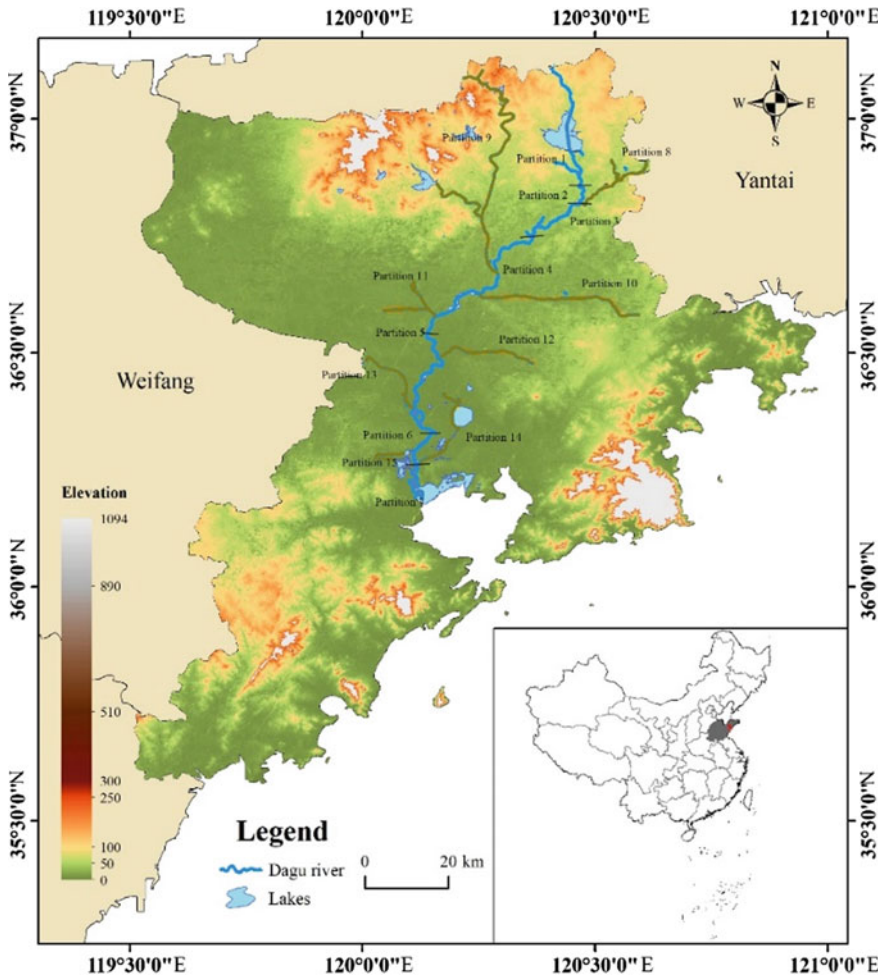


Fig. 1 Location of Dagu River

to score the target and indicator layers of river health in the Dagu River as the base data for the neural network model.

2.3 River Health Indicator System and Score Criteria

River health is a comprehensive problem involving many subjects. At present, the research on river health is still in the process of exploration and in-depth. In order to investigate the health of the Dagu river ecosystem, it is necessary to screen and

evaluate the indicators of river evaluation. The index system of Dagu river health assessment is shown in Table 1.

The river health was evaluated from the natural property and social service function of the river, and the evaluation grade was divided into five grades: “very healthy”, “healthy”, “sub-healthy”, “unhealthy”, and “bad condition” according to the “River Health Evaluation Guide (Trial)” and related literatures. The score intervals corresponding to the five grades are shown in Table 2.

Table 1 River health assessment index system

Objective layer	Rule layer	Index layer
Comprehensive state of river health (A)	Hydrological integrity (B1)	Variation degree of flow process (C1)
		Degree of ecological flow satisfaction (C2)
		River sediment transport capacity (C3)
	Morphological and structural integrity (B2)	Riparian vegetation zone width (C4)
		Longitudinal continuity of riparian vegetation (C5)
		Fluvial connectivity (C6)
		Channel solidification rate (C7)
	Chemical integrity (B3)	Dissolved oxygen content in rivers (C8)
		Degree of organic pollution in rivers (C9)
		Heavy metal pollution in rivers (C10)
	Biological integrity (B4)	Diversity of river benthic macroinvertebrates (C11)
		Zooplankton diversity in rivers (C12)
		Phytoplankton diversity in rivers (C13)
		River fish biodiversity (C14)
	Sustainability of social service functions (B5)	The standard of river water function areas (C15)
		Development and utilization of river water resource (C16)
		River flood control (C17)

Table 2 Grading of evaluation criteria

Estimation scale	Very healthy	Healthy	Sub-healthy	Unhealthy	Bad condition
Health level	[80, 100]	[60, 80]	[40, 60]	[20, 40]	[0, 20]

2.4 Artificial Neural Network Modeling

In this study, artificial neural network (ANN) and error back propagation (BP) algorithms were used to model and calculate the total scores of indicators such as natural conditions and social service functions. The input of the artificial neural network model is the score of the indicator layer. A three-layer artificial neural network with hidden layers is used to build the evaluation model. Based on previous experience and data magnitude, the two hidden layers are set up with 128 and 64 neurons, respectively. According to the related research and the actual situation of rivers, the river health evaluation results are divided into five categories, which can visually express the evaluation results of river health conditions.

The loss function is mean square error (MSE), which is expressed as follows:

$$MSE = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2 \tag{1}$$

Assume that the number of training samples is N , the set of true values are $y = \{y_1, y_2, y_n, \dots, y_N\}$, and the set of model predicted values are $\hat{y} = \{\hat{y}_1, \hat{y}_2, \hat{y}_n, \dots, \hat{y}_N\}$. y is the ground truth in the training data set, and \hat{y} is the predicted value of the current model. In order to introduce nonlinear factors into artificial neural network, ReLU function is selected as the activation function.

$$f(x) = \max(0, \omega^t + b) \tag{2}$$

3 Results and Discussion

3.1 Modeling and Result

The data in this study came from expert rating and objective measured data, with a total of 300 pieces of data. In this dataset, 192 data instances were randomly selected as the training data set of the neural network model, and 48 data were selected as the verification set. During the training, adam optimizer was used to adjust the learning rate adaptively, with batch size of 24. Due to the large number of parameters in the

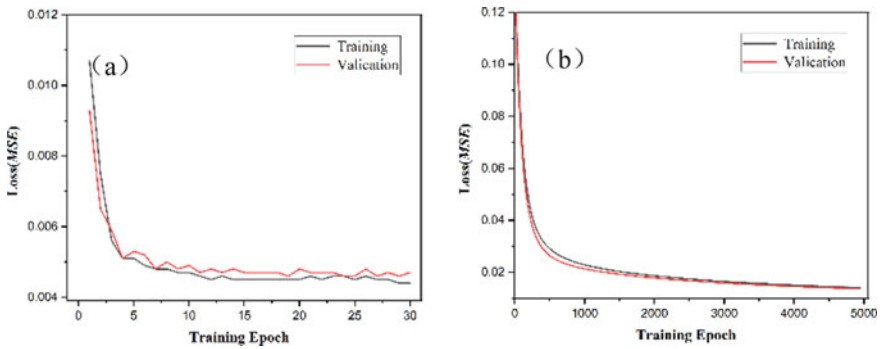


Fig. 2 a, b The MSE variance of training in ANN (a) and MLR (b)

ANN model, 60 sample data were randomly selected to verify the fitting results of the model.

This research employed Python and TensorFlow to implement artificial neural network modeling. We employed a computational resource with Inter (R) Core (TM) i7-11700 k, CPU@3.60GHZ, 64 GB RAM, and Geforce GTX 1050Ti GPU. According to the training results of the training set and validation set, the model converges and no over-fitting phenomenon occurs. The MAE result of the trained model in the test set is 4.98 (Fig. 2).

To further validate the performance of the model, we use Multiple linear regression (MLR) as a baseline against which to compare the model. Multiple linear regression is used to establish the correlation between multiple factors and variables, which is expressed as follows:

$$y = x_1\omega_1 + x_2\omega_2 + \dots + x_n\omega_n + b \tag{3}$$

Using SGD to find the best weights and using the same dataset as the neural network, the final MAE was calculated to be 8.125. Compared with the baseline method, ANN reduces MAE by 41.56% and MSE by 38.75%, proving its superior performance to the baseline method.

3.2 River Health Assessment

Based on the established evaluation model, the Dagou River basin was evaluated for river health, and the index layers assigned to 15 river zones were successively input into the trained evaluation model. At this point, the scoring results of the model for 15 river zones are 57, 57.4, 54.8, 59, 63, 42.6, 43.6, 42.7, 56, 45, 43.7, 55.7, 40.8, 55, 57 in sequence.

According to the scoring results and evaluation criteria, Sect. 4 of Dagu River is in a healthy state, while the rest of the river are in a sub-healthy state. The scoring results are consistent with the actual situation of Dagu River.

4 Conclusions

By applying artificial neural network to the river ecosystem health evaluation, and the corresponding network model between objective data and expert scoring was constructed. The results related to the river ecosystem health condition of Dagu River were obtained with the help of the actual monitoring objective data of Dagu River. The use of artificial neural network can reduce the errors caused by the perceived subjective factors in the evaluation process and improve the reliability of the evaluation of the river health status of Dagu River.

References

1. Shan C, Dong Z, Lu D et al (2021) Study on river health assessment based on a fuzzy matter-element extension model. *Ecol Ind* 127:107742
2. Zhou Y, Yue D, Li S et al (2022) Ecosystem health assessment in debris flow-prone areas: a case study of Bailong River Basin in China. *J Clean Prod* 357:131887
3. Petesse ML, Siqueira-Souza FK, de Carvalho Freitas CE et al (2016) Selection of reference lakes and adaptation of a fish multimetric index of biotic integrity to six amazon floodplain lakes. *Ecol Eng* 97:535–544
4. Shan C, Yang J, Dong Z et al (2020) Study on river health assessment weight calculation. *Pol J Environ Stud* 29(2):1839–1848
5. Fallahizadeh S, Kermani M, Esrafil A et al (2021) The effects of meteorological parameters on PM10: Health impacts assessment using AirQ+ model and prediction by an artificial neural network (ANN)—ScienceDirect. *Urban Clim* 38
6. Bansal S, Ganesan G (2019) Advanced evaluation methodology for water quality assessment using artificial neural network approach. *Water Resour Manage* 33(9):3127–3141
7. Anmala J, Venkateshwarlu T (2019) Statistical assessment and neural network modeling of stream water quality observations of Green River watershed, KY, USA. *Water Supply* 19(5–6):1831–1840

Optimization of Low Impact Development Scenarios Based on Computer-Aided Design



Fanglin Zhu, Ruolan Mu, Han Li, and Xuefei Li

Abstract With the increasing impact of surface runoff on hydrological cycle and environmental capacity, low-impact development (LID) facilities are widely used in stormwater management due to their good control effects on runoff and pollution. In order to quickly quantify the control effects of LID, efficient and intuitive modeling, with computer-aided design is essential. In this paper, we explore the optimal LID scenario in the study area by considering runoff control, pollutant control, life cycle cost, carbon emission, and other indicators. eCognition is used for landuse classification based on object-based method. The drainage model is established using InfoWorks ICM to analyze surface runoff and runoff pollution under the current condition through numerical simulation. LID combination scenarios are set to analyze control effects of runoff and pollutants. The life-cycle cost (LCC) and carbon emission of each scenario are calculated and evaluated. The entropy-TOPSIS method is used to comprehensively evaluate LID combination scenarios and explore the optimal LID scenario.

Keywords Computer-aided design · Low impact development · InfoWorks ICM · eCognition · Entropy-TOPSIS

1 Introduction

Continuous urbanization and climate change caused by global warming have made cities vulnerable to problems such waterlogging and water pollution. LID facilities can effectively control waterlogging and non-point source pollution caused by runoff [1], but LID facilities will produce carbon emissions in the construction, maintenance and recycling stages, which will cause damage to the atmospheric environment [2]. Therefore, it is significantly crucial to comprehensively evaluate different LID scenarios with appropriate indicators. With the development of computer technology,

F. Zhu · R. Mu · H. Li · X. Li (✉)
Qingdao University of Technology, Qingdao 266520, China
e-mail: 17860428205@163.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_60

457

computer-aided design has become an indispensable tool because of its accuracy and efficiency. The method of using computer-aided design to simulate the control effects of LID and selecting the optimal scenarios according to the simulation feedback is widely recognized.

At present, there are many simulation software. SWMM, MIKE Urban, InfoWorks ICM, and other stormwater management models can simulate different scenarios under different conditions. Fan et al. [3] established the hydraulic and water quality model to simulate the stormwater control effect of the bio-retention, rain garden, and vegetative swale combination scenario. Compared with the traditional development mode, LID measures can reduce the overflow by 9.75% and pollution load by 36.46%.

The landuse information in the study area plays a fundamental role in the model construction. With the improvement of image resolution, the pixel-based classification usually only exploit spectral information, resulting in low accuracy [4]. Object-based classification method that can simultaneously utilize spectral information, texture, and geometric features has emerged and developed. Ulloa-Torrealba et al. [5] applied eCognition software to conduct landuse analysis in Germany based on object-based classification method. The study achieved an overall classification accuracy of 82%.

In order to maximize the effectiveness of LID facilities and the synergistic effect, it is necessary to comprehensively evaluate different combination scenarios of LID facilities with multiple objectives. Mei et al. [6] applied the evaluation framework based on SWMM model and life cycle cost analysis to comprehensively evaluate the flood control development of watershed from the two aspects of runoff control effect and cost-effectiveness.

2 Method

In order to make the constructed model more accurate, eCognition is applied for landuse classification. After determining the main landuse types and proportions in the study area, the hydraulic and water quality model of the study area is established by Infoworks ICM. The model is utilized to simulate the hydrological and water quality indicators of before and after the implementation of different LID scenarios. The LCC and carbon emission of each scenario are calculated according to the formula. Finally, the entropy-TOPSIS method is used to comprehensively evaluate different LID scenarios and select the optimal scenario. The specific process is shown in Fig. 1.

2.1 Landuse Classification

Date preprocessing. The remote sensing image of Gaofen-2 satellite is selected in this study. The image is acquired on October 17, 2021, and the longitude of the

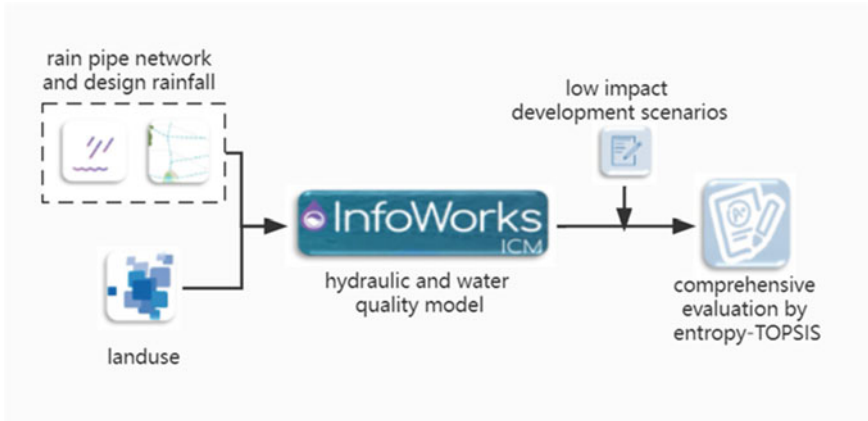


Fig. 1 Flowchart of study method

image center is 120.3°E and the latitude of the image center is 36.1°N. ENVI5.3 is applied to preprocess high resolution remote sensing data. Firstly, orthorectification is performed for multispectral and panchromatic images. Then, the processed panchromatic and multispectral images are fused and processed with atmospheric correction. Finally, the image is clipped using the shp file of the study area.

Multiresolution segmentation. In this study, we choose the Estimation of Scale Parameter (ESP) tool to determine the optimal segmentation scale. Firstly, the segmentation scale and compactness are set to 50 and 0.5 respectively, and the image is sequentially segmented with the shape factor from 0.1 to 0.9. The value of the shape factor is determined by observing the segmentation results. Then, we repeat the above steps to determine the appropriate compactness parameter. Finally, according to the rate of change of local variance (ROC-LV) graph generated by ESP, the optimal segmentation scale of different levels of the image is determined.

Landuse classification. Considering the actual landuse situation, the landuse types in the study area are divided into three categories: building, road, and green space. In this study, we apply eCognition to classify the high resolution remote sensing image based on object-based classification method after the optimal scale segmentation. By using the feature information such as spectrum, shape, and texture of ground object classes at each level, the buildings, roads, and green spaces are classified step by step.

Precision evaluation. Random points are generated by ArcGIS, and these data points are divided into buildings, roads, and green spaces based on visual interpretation. We assess classification accuracies applying Error Matrix based on TTA mask, the accuracy evaluation tool of eCognition. The classification accuracy is evaluated by the overall accuracy and Kappa value obtained.

2.2 Hydraulic and Water Quality Model Establishment

Storm design. In this study, five rainfall conditions (i.e. 1, 2, 5, 10, and 20 years return periods) are used to simulate design storm event in the study area. Combined with the coefficient of rain peak (the value is 0.4), the rainfall process of 120 min is generated from Chicago rain pattern. The rainfall intensity in this study is calculated as follows:

$$i = \frac{11.45 \times (1 + 0.997 \lg P)}{(t + 10.74)^{0.738}} \quad (1)$$

where i is the rainfall intensity, mm/min; P is the return period, year; and t is the rainfall duration, min.

Model setting. According to the data of rainwater drainage network and landuse types in the study area, the inspection wells, pipelines, and water outlets are generalized. We divide the subcatchment area by the Tyson polygon method.

When selecting the runoff yield model, we separately consider the permeable surface and the impermeable surface. For the impervious surface, the fixed percentage model is adopted. For the pervious surface, the Horton model is used to determine the initial loss, initial permeability rate, stable permeability rate, and attenuation coefficient through literature.

The SWMM model is applied for the routing model, and the kinetic wave theory is used to calculate the slope flow, which is the nonlinear reservoir method. The nonlinear reservoir method is also used to calculate the pipeline flow.

In this study, the accumulation model of surface pollutants adopts saturation model, and the scour model adopts exponential scour model. According to the landuse types in the study area, the parameter values of SS, COD, TN, and TP are determined.

Model calibration and validation. In order to ensure the accuracy and applicability of the model, two rainfall events of 20,220,702 and 20,220,917 are selected in this study for model calibration and validation. The 20,220,702 is used for model calibration, while the 20,220,917 is used for model validation. The Nash–Sutcliffe efficiency (NSE) is used to evaluate the goodness of fit.

2.3 Low Impact Development Scenarios

In this study, three kinds of LID facilities are implemented for three main landuse types: green roof for building rooftop, permeable pavement for road, and sunken lawn for green space. Considering landuse types and building distribution area, the area collocation of each LID facility is respectively: 3% + 3% + 3%, 3% + 3% + 6%, 3% + 3% + 9%, 3% + 6% + 3%, 3% + 6% + 6%, 3% + 9% + 3%, 6% + 3% + 3%, 6% + 3% + 6% and 6% + 6% + 3%.

2.4 Comprehensive Evaluation

Evaluation indicators. According to the low impact development control objectives and the evaluation content in the “Sponge City Construction Evaluation Standard”, this study considers indicators from three aspects: environmental benefits, economic benefits, and social benefits. Finally, nine evaluation indicators such as runoff control rate, peak flow control rate, overflow node control rate, SS total load control rate, COD total load control rate, TN total load control rate, TP total load control rate, life cycle cost, and carbon emission are determined.

Control rate calculation. Under the condition that the return period is 20 years, the simulation values of the model before and after the implementation of LID facilities are taken as the calculation objects, and the calculation method is shown in Eq. (2).

$$R_i = \frac{W_{i\text{-withoutLID}} - W_{i\text{-LID}}}{W_{i\text{-withoutLID}}} \times 100\% \tag{2}$$

where R_i is the control rate of indicator i (runoff, peak flow, overflow node, non-point pollution loading), %; $W_{i\text{-without LID}}$ is the value of indicator i before LID construction; $W_{i\text{-LID}}$ is the value of indicator i after LID construction.

Life-cycle cost calculation. The life-cycle cost of LID facilities can be divided into initial construction cost, operation and maintenance cost, and salvage value. In this study, we adopt the present value (PVC) to calculate the life cycle cost of LID facilities, and the calculation method is shown in Eq. (3).

$$PVC = IC + \sum_{t=0}^T \left(OMC \times \frac{1}{(1+r)^t} \right) - \frac{1}{(1+r)^T} \times SV \tag{3}$$

where PVC is the present value of life cycle cost of LID, RMB; IC is the initial construction costs of LID, RMB; OMC is the annual LID operation and maintenance costs, RMB; SV is the salvage value of LID, which is 0; t is the design service life of LID, which is calculated as 20 years; r is the discount rate, which is 5%.

Carbon emission calculation. In this study, the emission factor accounting method provided by the Intergovernmental Panel on Climate Change and the United Nations Framework Convention on Climate Change is adopted to conduct carbon emission accounting for the construction process of LID [7]. The carbon sink function in LID facilities is mainly assumed by plants, but the carbon sink capacity of plants has great uncertainty, so carbon sink is not considered in this study.

Comprehensive evaluation. Firstly, the initial data of evaluation indicators are standardized to form a standardized judgment matrix, and the entropy value and entropy weight of each indicator are calculated. Then, the weighted decision matrix is formed by the weight of the indicators and the standardized judgment matrix. The optimal solution and the worst solution of the matrix are calculated, and the Euclidean distances between different LID schemes and the best solution are calculated.

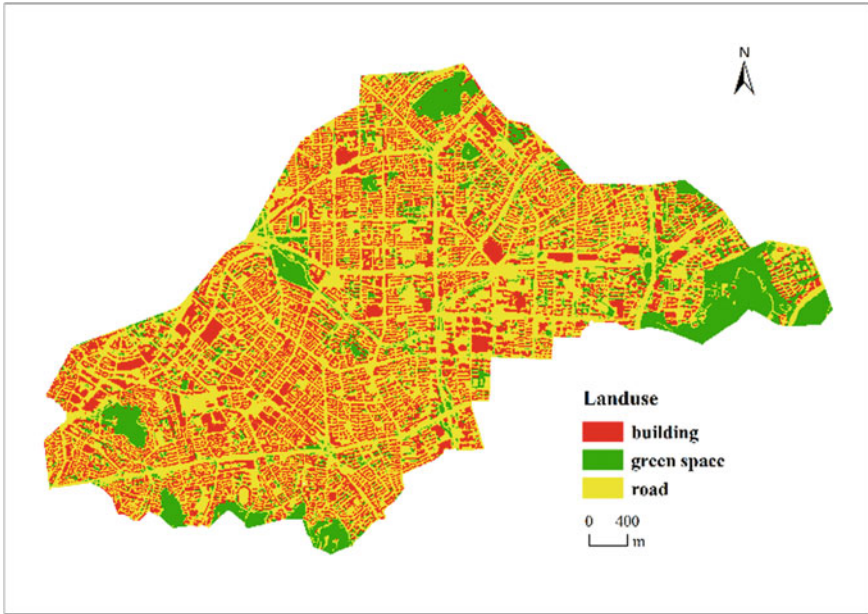


Fig. 2 Landuse map of study area

3 Preliminary Result

The overall accuracy and kappa coefficient of the landuse classification are 93.1% and 0.885, which is a relatively satisfactory classification result. The landuse classification of the study area is shown in Fig. 2.

4 Conclusion

In this study, we simulate and evaluate the performance of different LID scenarios based on InfoWorks ICM and entropy-TOPSIS method. The current hydraulic performance is evaluated based on InfoWorks ICM. Different LID combination scenarios are constructed, and the scenarios are compared and selected from three aspects: environmental benefits, economic benefits, and social benefits. Preliminary results show that this method is feasible and effective.

References

1. Hou X, Qin L, Xue X et al (2021) A city-scale fully controlled system for stormwater management: Consideration of flooding, non-point source pollution and sewer overflow pollution. *J Hydrol* 603:127155
2. Moore TLC, Hunt WF (2013) Predicting the carbon footprint of urban stormwater infrastructure. *Ecol Eng* 58:44–51
3. Fan G, Lin R, Wei Z et al (2022) Effects of low impact development on the stormwater runoff and pollution control. *Sci Total Environ* 805:150404
4. Wu YJ, Zhang PF, Wu J et al (2021) Object-Oriented and Deep-Learning-Based High-Resolution Mapping from Large Remote Sensing Imagery. *Can J Remote Sens* 47(3):396–412
5. Ulloa-Torrealba Y, Stahlmann R, Wegmann M et al (2020) Over 150 Years of Change: Object-Oriented Analysis of Historical Land Cover in the Main River Catchment. Bavaria/Germany. *Remote Sensing* 12(24):4048
6. Mei C, Liu J, Wang H et al (2018) Integrated assessments of green infrastructure for flood mitigation to support robust decision-making for sponge city construction in an urbanized watershed. *Sci Total Environ* 639:1394–1407
7. Lin X, Ren J, Xu J et al (2018) Prediction of Life Cycle Carbon Emissions of Sponge City Projects: A Case Study in Shanghai. China. *Sustainability* 10(11):3978

The Development of Expressive Vocabulary of Korean Learners as a Heritage Language



Sin Hye Nam , Chaerin Jang , and Sunyoung Kim 

Abstract The purpose of this study is to analyze and describe the development of expressive vocabulary of Korean learners as a heritage language. Therefore, this study will check whether the vocabulary used by Korean learners as a heritage language actually expands from individual to social and professional fields in light of Vygotsky's theory based on the learners' corpus data. In this study, the Korean language learners' corpus was used as research data. Keywords were selected using the TextRank algorithm. As a result, keywords in the individual domain were prominent in beginner, keywords in the social domain in intermediate, and keywords in the social and professional domain in advanced proficiency levels. Through the analysis, it was confirmed that the cognitive development area of heritage learners was expanded from the individual scope to the social, national, and professional domain, and it was possible to confirm the pattern of negotiation between areas. It can be seen that this is in line with the learning aspect of Vygotsky's social-constructionism.

Keywords Korean learners as a heritage language · Expressive vocabulary · Vocabulary development · Corpus linguistics · Keywords analysis

1 Introduction

The purpose of this study is to analyze and describe the development of expressive vocabulary of Korean learners as a heritage language. Korean learners have different

S. H. Nam (✉)

Kyung Hee University, 26, Kyungheedaero, Dongdaemun-gu, Seoul, Republic of Korea
e-mail: namsh@khu.ac.kr

C. Jang

Myongji University, 34 Geobukgol-ro, Seodaemun-gu, Seoul, Republic of Korea
e-mail: jangchaerin@mju.ac.kr

S. Kim

Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea
e-mail: answh33@gmail.com

characteristics from the group of learners who want to learn Korean as a foreign language or a second language in that they are mainly raised in an environment in which their mother tongue and national language are inconsistent and have not acquired Korean as a first language. In particular, Korean learners as a heritage language may have distinct characteristics from groups of learners who do not learn it as a heritage language, in terms of the purpose of Korean learning, the degree of background knowledge they had before starting Korean learning, or the degree of familiarity with linguistic characteristics such as Korean vocabulary and word order [1–5]. Therefore, it is worthwhile to analyze their interlanguage, distinguishing it from a general group of Korean learners. Moreover, there has been no research in the field of Korean language education to examine the specific aspects of how their interlanguage develops according to proficiency. Therefore, this study will check whether the vocabulary used by Korean learners as a heritage language actually expands from individual to social and professional fields in light of Vygotsky's theory [6] based on the learners' corpus data.

2 Methods

2.1 Data

In this study, the Korean language learners' corpus established by the National Institute of Korean Language was used as research data. The Korean learners' corpus consists of language data produced by Korean learners. In this study, only the data produced by Korean learners as a heritage language among the total data were analyzed. In order to control the variables according to the language characteristics of each register, only written language data were included, and as a result, 589 texts consisting of a total of 72,802 words were analyzed. Since such data is an interlanguage corpus, errors produced by learners remained intact, and all errors at the vocabulary level were corrected for this study. However, corrections were not made if it was not a form error and the learner misused vocabulary that did not fit the context. This is to analyze the vocabulary itself produced by the learner with a certain intention.

2.2 Keywords Analysis Method

In this study, keywords were selected using the TextRank algorithm, one of the machine learning techniques used to extract keywords for document summarization. Text Rank has been widely used in lexical extraction studies, text summary studies, word recommendation studies, and educational lexical list selection studies since it was first proposed in Mihalcea and Tarau (2004). TextRank is designed to weigh

words or sentences by considering the frequency of words or sentences that make up text and the connection between them. In this study, keywords of major content words were selected for each proficiency level of learners using this algorithm.

3 Result

We will discuss what keywords came out by level while looking at the order shown in the TextRank index. The WordCloud image below shows notable words among keywords from level 1 to 6 (Fig. 1).

The meaning categories of first level keywords appeared in the order of “concept, eating, life, human, and economic life”. What is characteristic of the first level is the appearance of many eating-related vocabulary. Since “eating” is related to an aspect most frequently encountered in daily life, it seems that “eating” categories frequently appeared in the first level. The meaning category of the second level keywords frequently appeared in the order of “concept-human-life-social life-education-state-eating”. Concept, human, and life categories appeared many times, which was an observation also made in the first level’s interlanguage. Meanwhile, categories such as “social life” and “education” were witnessed relatively more often compared to the first level. Unlike the prominent presence of eating-related vocabulary in the first level, it can be found that the topics have expanded away from survival to cover topics such as “travel, class, and hometown” (Fig. 2).

In the next proficiency levels, the “culture” category was prominent compared to the beginner level. Keywords in the first and second levels were mainly vocabulary related to “daily life”, while the third level is the starting point of the intermediate level and further reflects the cultural aspect that can be seen as a more expanded topic. In level 4, abstract concepts are included in the upper keyword list compared to key words in level 1 to 3, such as “experience, difference, life”, etc. Vocabulary in the social domain has already appeared in keywords of level 3, but vocabulary in the abstract domain appears distinctively in level 4 (Fig. 3).



Fig. 1 Keywords for level 1, level 2



Fig. 2 Keywords for level 3, level 4



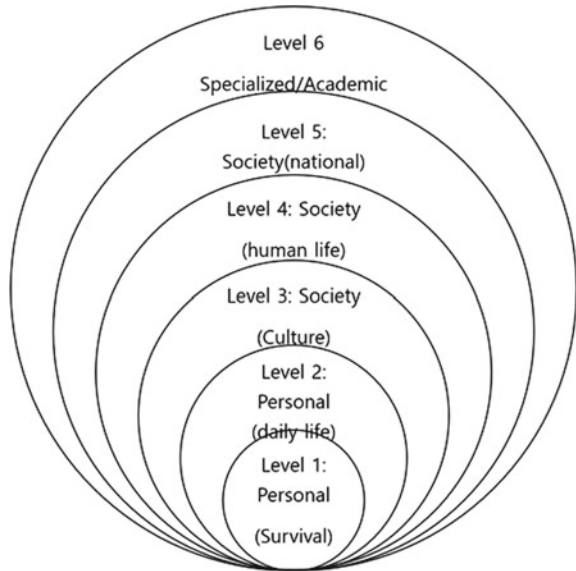
Fig. 3 Keywords for level 5, level 6

In level 5, the meaning category of keywords was found to be more frequent in the order of “concept, human, social life, and life”. Many keywords of abstract concepts were included, and the meaning category began to appear at the national level beyond the personal and social domains. Compared to the keywords in proficiency level 1 to 5, the prominent meaning category in level 6 was “politics and administration”. It is worth noting that the vocabulary related to “politics and administration” was not considered a keyword at all in level 5, but it was considered in the sixth level. “Politics and administration” are relatively difficult and can be seen as professional areas. Level 5 is also an advanced course, and keywords in the social and national domains appear, but it can be seen that level 6 extends beyond them to detailed professional and academic fields.

So far, keywords by level 1 to 6 have been analyzed and examined. The expanded pattern of the keywords of the heritage language learners discussed above is shown as follows (Fig. 4).

Keywords in the individual domain were prominent in beginner (level 1 and 2), keywords in the social domain in intermediate (level 3 and 4), and keywords in the social and professional domain in advanced (level 5 and 6) proficiency levels. Through the analysis discussed above, it was confirmed that the cognitive development area of heritage learners was expanded from the individual scope to the social,

Fig. 4 Extension of keywords by proficiency



national, and professional domain, and it was possible to confirm the pattern of negotiation between areas. It can be seen that this is in line with the learning aspect of Vygotsky’s social-constructionism.

4 Conclusion

In this study, the development of expressive vocabulary of Korean learners as a heritage language in the Korean learner corpus was analyzed, and the analysis results were examined by applying Vygotsky’s theory of social constructivism. As a result of analyzing keywords from learners’ writing materials from levels 1 to 6, it was found that vocabulary related to eating was the most common in level 1, and the topic was expanded to “travel” or “class” in level 2. In the third level, more cultural-related vocabulary such as “concert” and “performance” began to appear compared to beginners. In level 4, vocabulary starts to reflect abstract experiences such as “experience, difference, and life”, which were not widely observed in the previous three classes. In advanced level 5, meaning categories that reflect social or national dimensions began to appear rather than vocabulary at the individual level. It can be said that vocabulary that appeared in level 6 belong to a similar category to level 5, but is characterized by expanding to more professional and academic areas. In conclusion, vocabulary belonging to the individual meaning category at the beginner level, the social and abstract meaning category at the intermediate level, and the social and professional meaning category at the advanced level showed that the vocabulary used by Korean learners as a heritage language expands from individual to social and

professional fields, as in Vygotsky's theory of social constructivism. Going forward, by comparing the data and analysis results of Korean learners for general purposes, it will be possible to more clearly reveal the characteristics of Korean learners' use of vocabulary as a heritage language.

References

1. Kim CE, Pyun DO (2014) Heritage language literacy maintenance: a study of Korean-American heritage learners. *Lang, Cult Curriculum* 27(3):294–315
2. Kim H (2017) Syntactic complexity in the writing of Korean heritage learners in the United States. *Korean Lang Am* 21(2):186–217
3. Lee D (2008) Revisiting advanced Korean learners' pragmatic competence: focusing on the discourse analysis of heritage and non-heritage learners in the US. *J Korean Lang Educ* 19(3):295–320
4. Lee J (2014) A longitudinal study of Korean vocabulary development as a heritage language for Korean-American preschool children. *J Korean Lang Educ* 25(3):259–280
5. Lee H (2012) A discourse analysis of oral presentations in Korean language classes: focusing on the style characteristics of heritage and non-heritage learners. *J Korean Lang Educ* 23(4):283–306
6. Vygotskii LS (2008) *The development of higher psychological processes*. Harvard University Press, Boston

Applying Process Mining Techniques in Sewage Treatment Plant Management



Bin Shao, Guoqing Ni, Jingbo Zhao, and Sheng Miao

Abstract Due to the operational process of sewage treatment plant is complex and changeable. As a result, the sewage treatment plant may have operational errors of the staff. The measure is taken in this paper by asking an expert to debug equipment in the sewage plant for a period of time, who is proficient in the whole operation process. First, the Python library PM4Py is called. Then the expert event log is used as input for process mining. Finally, conformance checking is the most critical technology, which contrasts the expert Petri net with the event logs of workers, to detect, locate, and explain the deviations between them. The method that is used in this study can investigate or evaluate the operation of sewage treatment plant workers. This paper can assist sewage plant managers to identify employees who are not operating properly. The method is convenient and fast with high accuracy. In addition, some accidents caused by abnormal staff handling can be effectively avoided.

Keywords Process mining · PM4Py · Petri net · Conformance checking

1 Introduction

Due to the high time delay and complex reaction of the sewage treatment process, so the debugging operations often take several hours before the result is returned. Nowadays, there are many studies that relate to the application of data mining to

B. Shao · G. Ni · J. Zhao · S. Miao (✉)
Qingdao University of Technology, Qingdao 266520, China
e-mail: smiao@qut.edu.cn

B. Shao
e-mail: shaobin77@outlook.com

G. Ni
e-mail: n1633703426@outlook.com

J. Zhao
e-mail: zhaojingbo6666@163.com

wastewater treatment plants. But for this issues, traditional data mining has some limitations [1]. Because data mining is operated in a black box mode. It gives direct conclusions which are difficult to trace back to the cause [2]. Therefore, the new concept of process mining is introduced in this paper. Different perspectives on process paths are provided by process mining. By displaying deviations in the actual process, bottlenecks and the influence of inconsistency can be identified.

Process mining is widely used in areas such as enterprise process analysis, health-care process management, and manufacturing process management. Data is extracted from the actual process by process mining. This can reveal the actual implementation of the business. Process mining techniques are useful to provide insight and help to improve (work) processes. It identifies ineffective or inefficient work that has a significant impact on company operations without affecting existing systems. In summary, process mining plays an important role in the management of sewage treatment plants.

With the continuous improvement of the informatization degree and working efficiency in sewage treatment plants. At the same time, it also brings management problems to the sewage treatment plant. For example, it is difficult for director to assess the professional ability of the sewage plant workers. This paper proposes an approach that uses process mining to deal with this problem. This method has three steps. Above all, the PM4Py library is called in the Anaconda environment. Then, the expert event log is used as input for process discovery technology. An actual process model can be automatically built by analyzing the event log [3]. In the end, model and worker's event logs are used as input for conformance checking. The token-based replay is an essential technology in this paper. Fitness is used to analysis in this paper to determine the deviation between the logs and the model. If some traces in the log can't be replayed by the Petri net model, the corresponding workers will be found. Because the operation of sewage treatment plant workers may be wrong. The research of this paper demonstrates that PM4Py library implemented well in this type of analysis.

2 Background

2.1 *Current Status of Sewage Treatment Plant*

Sewage treatment processes are risky due to the different level of sewage plant workers. It is essential that manager evaluates the professional ability of sewage treatment plant employees [4].

Nowadays, two serious problems are faced by enterprises. To begin with, the continuous rise of labor cost which makes the cost of sewage treatment plant increase; in the second place, with the rapid development of sewage treatment plant. Processes within the business become complex. Taking all these factors into consideration, this

research naturally come to the conclusion that sewage treatment plant should take an effective approach to deal with these problems.

2.2 *Current Status of Process Mining*

Process mining is regarded as a cross-field research direction. It can be viewed as a measure to bridge the gap between data science and process science [5]. On the one hand, it lies between machine learning and data mining. On the other hand, it lies between process modeling and analysis [6]. Process mining is a technique that can be used to find process models from event logs. By replaying the data, bottlenecks and the influence of inconsistency can be identified [7]. Event logs are easily obtained in today's information system.

With the development of information systems, the various forms of event logs are stored in the information system [8]. It provides the sewage plant manager with historical information about process execution. Process mining can assist plant manager to observe the actual operation of the sewage plant. In this way, the relevant personnel of the sewage plant can be assisted to detect, locate, and explain deviations, and to measure the severity of these deviations [9]. Process mining becomes a popular area of research. However, there are only a handful of projects that have applied process mining to specific scenarios. This paper applies process mining to wastewater treatment plant, which can effectively solve the relevant problems.

Process mining techniques have three main parts. The first technology is process discovery. Event logs are used as input to process discovery. It has alpha algorithm, inductive algorithm, and heuristic algorithm [10]. Then the actual process model (Petri net) can be obtained. The expert event log is used as input to this paper. The PM4Py library is called in the Anaconda environment. A Petri net is generated by inductive algorithm. The model explains the behavior that is recorded in the expert event log.

The second technology is conformance checking. Here, an existing Petri net model is compared with an event log. Therefore, conformance checking can be used to check the deviation and bottleneck of the actual process [11]. Process mining has four quality dimensions: fitness, accuracy, simplicity, and generalization [12]. For conformance checking, the main concern is fitness. The third technology is enhancement. The purpose is to extend or improve an existing process.

As an open-source Python library. A lot of algorithms are provided by PM4Py. Process mining has some open source software such as ProM. And it also has some commercial software such as Disco and Celonis [13]. However, commercial software has certain limitations. For example, it only supports a discovery algorithm, and it has limitations in terms of conformance checking [14, 15]. A wide range of functions are offered by ProM. At the same time, it is a software with extensible feature (e.g., the ability for users to add plug-ins). However, it can only be used through the graphical user interface (gui), which makes experimental research limited. PM4Py attempts to

fill the gap of the software mentioned above, and try to apply it to a wide range of scientific or industrial experimental environments.

3 Conformance Checking of Sewage Treatment Plant

3.1 Data and Preprocessing

Event logs are used as a starting point for process mining. In this study, an expert in the sewage treatment field is invited to operate equipment for a specific period of time. The expert is proficient in every process segment of the sewage plant. First of all, this paper needs to collect the event log from the sewage plant expert. Secondly, process discovery technique is applied to this event log. In the end, the actual petri net model is obtained. After that, the event logs of 50 sewage plant operators are collected.

The various process datas from the sewage plant are mainly represented by event logs. Due to the complexity of the data that is generated by the process. The event logs are extracted from the information system, which cannot be directly used for process mining. Event logs need to be filtered in this research. This study mainly carried out three operations on the logs of sewage plant expert and operation staff: data cleaning, data integration, and data transformation.

The first step is to delete irrelevant or repetitive records. For example, the operation to regulate the flow of water is performed twice by sewage plant worker. Then, the process logs of sewage plant expert and sewage plant operator are filtered and saved as CSV files. The event log format for the process mining is XES format. By calling Python's PM4Py library. CSV files can be imported directly in XES format. As shown in Fig. 1, the main research of this paper can be divided into three parts.

Fig. 1 Block diagram of the role of the conformance checking technique

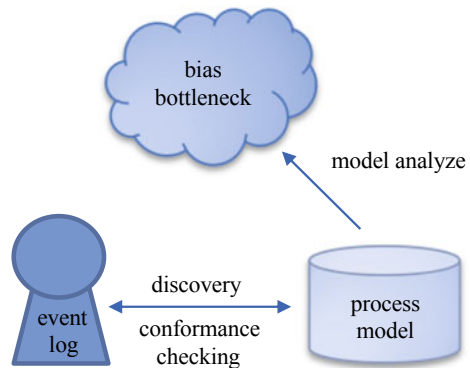
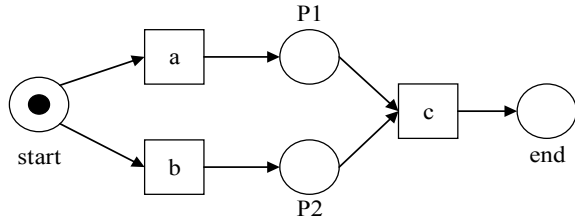


Fig. 2 Petri net



3.2 Process Modeling and Conformance Checking

Petri net is a process modeling language that supports modeling concurrency semantics. A directed net is a triplet $N = (P, T, F)$. This directed network needs to satisfy certain conditions.

$$P \cap T = \emptyset \tag{1}$$

$$P \cup T \neq \emptyset \tag{2}$$

$$F(P \times T) \cup (T \times P) \tag{3}$$

P is the set of place. T is the set of transition. F is the set of connection. Place and transition are connected by F , which represents a flow relationship. The triple N is a Petri net. P elements are represented by circles, T elements are represented by boxes, and F is represented by directed arrows. Token is the dynamic object in a place. It can be moved from one place to another place. The visual representation of Petri net is shown in Fig. 2.

The event log of the sewage plant expert is first used as input for process discovery. The PM4Py library is called by the Anaconda environment. The actual process model of sewage plant is discovered by the inductive algorithm. The model is a Petri net. The significant part of this section puts emphasis on conformance checking between expert model and the worker event logs. By calling the function of token-based replay in Pm4py library, the deviations between the worker logs and the model are found respectively. The token-based replay technology is used in this research. It calculates the fitness based on the number of all tokens. The formula is as follows:

$$\text{fitness}(L, N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times P_{N,\sigma}} \right) \tag{4}$$

The degree of fitness between event log L and Petri net N is expressed by the formula. Four counters are displayed for each stage: $P_{N,\sigma}$ (produced tokens), $c_{N,\sigma}$ (consumed tokens), $m_{N,\sigma}$ (missing tokens), and $r_{N,\sigma}$ (remaining tokens). By analyzing

the size of fitness, it can help the manager of the sewage treatment plant to check whether the worker operation is too different from the expert operation. If the behavior of the event log of the workers in the sewage plant deviates too much from the expert process model. The workers can be evaluated or trained by the director of the sewage treatment plant. This is beneficial to reduce the cost of sewage treatment plant and some accidents caused by abnormal staff handling can be effectively avoided.

4 Conclusion and Discussion

This paper mainly demonstrates the usability of process mining in the use of PM4Py to treat sewage plant management problems.

Three main research components are mentioned in the theoretical part: the data requirements of process discovery, the basic representation of process model and conformance checking. The main algorithm is used by the open-source platform PM4Py. In the practice part. A method is then proposed in this research to solve such problems by using process mining. The assessment shows that the inductive algorithm is capable of identifying process scenarios from event logs in the domain of sewage treatment. The expert model is used to check the conformance between the expert model and the worker event logs. Finally, the manager of the sewage plant can be assisted to analyze the deviation and bottleneck. This method is simple, and it can be easily applied.

In the future, this study plans to implement process mining into the whole sewage treatment process of the sewage plant. Through the use of process mining technology, the deviation or bottleneck can be found in the sewage treatment. It helps the sewage plant director to better manage the sewage treatment plant.

References

1. Trávníček P, Junga P, Kotek L et al (2022) Analysis of accidents at municipal wastewater treatment plants in Europe. *J Loss Prev Process Ind* 74:104634
2. Pechenizkiy M, Trcka N, Vasilyeva E et al (2009) Process mining online assessment data. In: International working group on educational data mining
3. Van der Aalst W, Adriansyah A, van Dongen B (2012) Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisc Rev: Data Min Knowl Discov* 2(2):182–192
4. Zhang Z, Johnson C, Venkatasubramanian N et al (2022) Process scenario discovery from event logs based on activity and timing information. *J Syst Architect* 125:102435
5. Mannhardt F (2018) Multi-perspective. *Process Min*, pp 41–45
6. Rodríguez-Quintero J, Sánchez-Díaz A, Iriarte-Navarro L et al (2021) Fraud audit based on visual analysis: a process mining approach. *Appl Sci* 11(11):4751
7. Marin-Castro HM, Tello-Leal E (2021) Event log preprocessing for process mining: a review. *Appl Sci* 11(22):10556
8. Maeyens J, Vorstermans A, Verbeke M (2020) Process mining on machine event logs for profiling abnormal behaviour and root cause analysis. *Ann Telecommun* 75(9):563–572

9. Leno V, Polyvyanyy A, Dumas M et al (2021) Robotic process mining: vision and challenges. *Bus Inf Syst Eng* 63(3):301–314
10. Weijters A, van Der Aalst WM, De Medeiros AA (2006) Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP 166(July 2017):1–34
11. Munoz-Gama J, Martin N, Fernandez-Llatas C et al (2022) Process mining for healthcare: characteristics and challenges. *J Biomed Inform* 127:103994
12. Berti A, Van Zelst SJ, van der Aalst W (2019) Process mining for Python (PM4Py): bridging the gap between process-and data science. arXiv preprint [arXiv:1905.06169](https://arxiv.org/abs/1905.06169)
13. Ghahfarokhi AF, van der Aalst WM (2022) A Python tool for object-centric process mining comparison. arXiv preprint [arXiv:2202.05709](https://arxiv.org/abs/2202.05709)
14. Nagy Z, Werner-Stark A (2022) An alignment-based multi-perspective online conformance checking technique. *Acta Polytech Hung* 19(4)
15. Dreher S, Reimann P, Gröger C (2021) Application fields and research gaps of process mining in manufacturing companies. *INFORMATIK 2020*

Intelligent Sewage Treatment Control System Based on Digital Twin



Hao Jiang, Zikun He, Shuqi Liu, Yang Hai, Chao Liu, and Sheng Miao

Abstract With the development of economy, people pay more and more attention to urban sewage treatment. At present, the operation and management of sewage treatment plants are gradually changing from digital to intelligent. As an emerging technology, digital twin can make an important contribution to the development of intelligent water. Based on digital twin, machine learning, Internet of Things, and other technologies, this study constructed a smart control system applied to sewage treatment plants. Water quality information such as COD, TN, and TP collected on site is driven into the 3D model of the system. At the same time, based on the historical data, the effluent quality is predicted. Assist staff in dosing and equipment maintenance according to forecast results. Preliminary experimental results show that the system realizes the data-driven 3D model, completes the data prediction and remote interaction, and runs stably. It can save a lot of manpower and material resources and has feasibility.

Keywords Sewage treatment · Digital twin · Machine learning · Water quality prediction · IoT

1 Introduction

With the continuous improvement of residents' quality of life, urban sewage production is increasing year by year, and urban water pollution is becoming increasingly serious. The urban sewage treatment plant, as an important facility in the city, is the

H. Jiang · Z. He · S. Liu · C. Liu (✉) · S. Miao
Qingdao University of Technology, Qingdao 266520, China
e-mail: liuchao@qut.edu.cn

S. Miao
e-mail: smiao@qut.edu.cn

Y. Hai
Hidrostal (Qingdao) Pumps Co., Ltd, Qingdao 266200, China

focus of the sewage treatment industry in recent years. In the construction of smart city, the intelligent management of water resources has become an indispensable part. Intelligent sewage treatment is of great significance to promote the long-term healthy development of the city.

At present, most plants adopt BIM-based 3D visualization technology as a single static model display. But it will not change with the update of sewage plant information. Collected data is often only available for query and is difficult to display on the model in real time. As an important part of smart city, the application of digital twin of plant has been paid more and more attention. Digital twin technology will break the technical barrier of the industry and realize the fusion of 3D model and data driven. Sewage treatment plant is an important urban infrastructure. If the sewage treatment equipment encounters failure or loss, the function of the plant will be interrupted in the long term or short term, which will cause serious impact on the surrounding environment and population [1]. In addition, in order to effectively remove pollutants such as nitrogen and phosphorus, additional chemicals need to be added to the facility. Due to the different experience levels of staffs, the dosage will often exceed the standard, resulting in unnecessary expenses.

In this paper, a model of intelligent sewage treatment control system is built based on digital twin. The machine learning algorithm is used to model and predict the effluent quality. The data collected in the field and predicted by the model are respectively driven into the system. The virtual and real interaction of digital twin is finally realized.

2 Related Works

The idea of digital twin was first proposed by Professor M. Grieves of the University of Michigan in the United States, who defined it as information mirror model, and then evolved into the term “digital twin” [2]. In recent years, digital twin technology has been widely used in industrial fields such as virtual prototype, digital twin workshop, and digital twin satellite [3], but it has not made a high breakthrough in the field of sewage treatment.

Tao et al. [4] proposed a new method of product design, manufacturing, and service based on digital twin drive, and illustrated the application prospect of digital twin in three stages of product through three cases. Liu et al. [5] proposed a rapid personalized design method for automated flow shop manufacturing system based on digital twin, which solved the personalized design problem faced in large-scale personalized manufacturing mode. Park et al. [6] used the digital twin technology based on the Industrial Internet of Things to solve the cost and performance barriers of Factory-as-a-Service personalized production and overcome the difficulties of integrated management. Guo et al. [7] systematically introduced the application of digital twin in industry, medicine, smart city, aerospace, commerce, and other fields, and put forward the prospect of digital twin.

With the rapid growth and accumulation of data in all walks of life, applications related to big data begin to attract attention [8]. In recent years, in the field of sewage treatment, many researches have applied machine learning methods to solve complex practical engineering problems [9], and made a series of progress in solving the low accuracy of water quality prediction, difficulty of fault diagnosis and lag of regulation in the process of sewage treatment.

Zhang et al. [10] used machine learning method to study the energy consumption of sewage treatment plants and established a new non-numerical variable energy consumption model by using random forest algorithm. Torregrossa et al. [11] used machine learning technology to generate a high-performance energy cost model for plants. Chen et al. [12] optimized the content of DO and dosage of agents in plants at the same time through multi-agent deep reinforcement learning.

3 System Architecture

The architecture of smart sewage treatment control system is shown in Fig. 1. Internet of Things (IoT) devices collect data from the plant. Data is transmitted through the Data Transfer Unit (DTU) and stored in the database. Through the database, the system realizes 3D visualization, remote interaction, and data prediction.

3.1 Data Collection

Through the IoT technology, managers can realize the information management of sewage treatment process. Data acquisition is the foundation of IoT, and sensor technology is an indispensable part of data acquisition technology. The system adopts a variety of IoT intelligent devices that support Modbus protocol, such as flowmeter, water quality on-line detector, and intelligent temperature sensor. Through a series

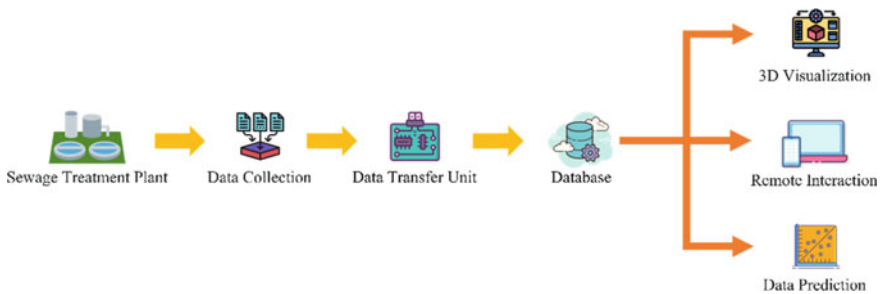


Fig. 1 Architecture of sewage treatment system

of sensors, the system can obtain flow rate, velocity, temperature, dissolved oxygen, water quality, operating frequency, current, voltage, and other relevant data.

3.2 Data Transmission

As one of the general standards in the industrial field, Modbus communication protocol is of great importance and universality in industrial control systems. Modbus protocol not only supports the real-time capture of sensor data, but also supports the operation command of the device. Real-time monitoring and remote interaction of wastewater treatment plants are possible through IoT sensors and Modbus protocol.

Through the Modbus protocol, the data collected by IoT sensors are aggregated into the DTU. Finally, the data collected in the field is transmitted to a server on the Internet through a DTU. All data is stored in the database.

3.3 3D Visualization

According to the design drawing, build a 3D model of the sewage treatment plant. The Revit software is used to construct the model of the structure, pipeline, and wall of the plant. Export the 3D model in “FBX” format to Unity 3D software. 3D scenes are rendered in the software. In addition, by connecting Unity 3D with the database, the real-time data collected by the sensor and the future data predicted by the algorithm can be driven into the 3D model. Through the Web-GL function of the software, the visual model is implemented to the web page. So far, the system has realized the 3D visualization of the plant. The 3D model building process is shown in Fig. 2.

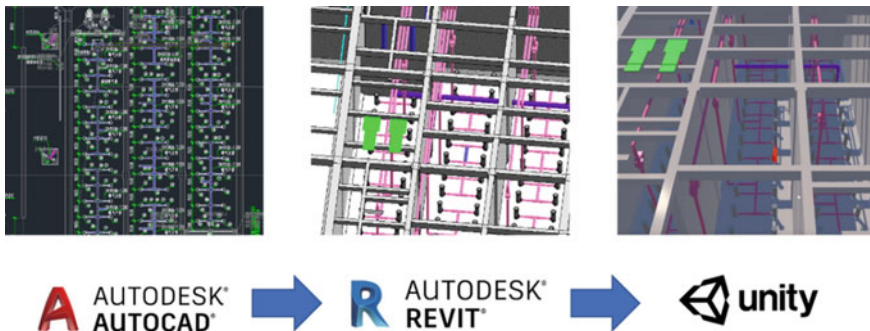


Fig. 2 Process of building the 3D model

3.4 Data Prediction

According to the temporal characteristics of the collected data, the system uses long short-term memory (LSTM) neural network to construct a water quality prediction model. LSTM neural network is a kind of time recurrent neural network, which can effectively keep the memory for a long time. Due to its unique design structure, the model has better performance for sequence data with longer time intervals.

3.5 Remote Interaction

The remote interaction module of this system is mainly carried out on the web page. You can view real-time data and forecast data of IoT devices. At the same time, you can generate historical data trend in the system, and download the data to the local. Through the Internet of Things device, the monitoring camera at the sewage treatment plant is connected to obtain the scene capture picture. When the system predicts that the water quality is not up to standard in the future, it will alarm in time. Workers are alerted and immediately adjust equipment or add chemicals. When the equipment fails, the system will remind the management personnel for maintenance in the first time.

In order to facilitate users to check information at any time, a mobile platform is developed. Users can use their mobile phones to view real-time data, the scene capture picture, and other simple information. The remote interaction is shown in Fig. 3.

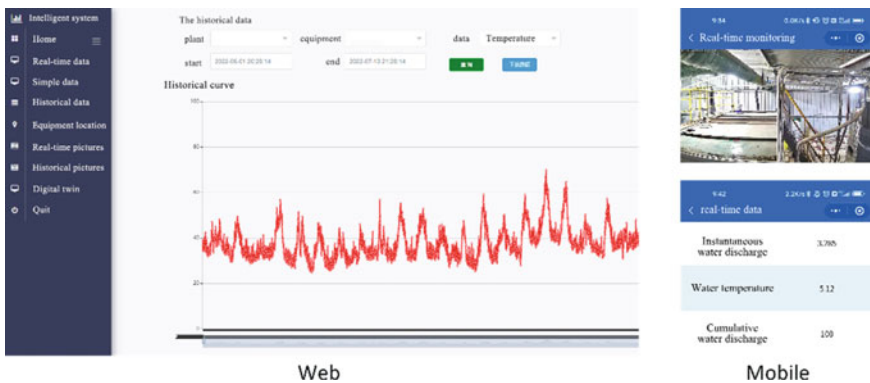


Fig. 3 Remote interactive interface

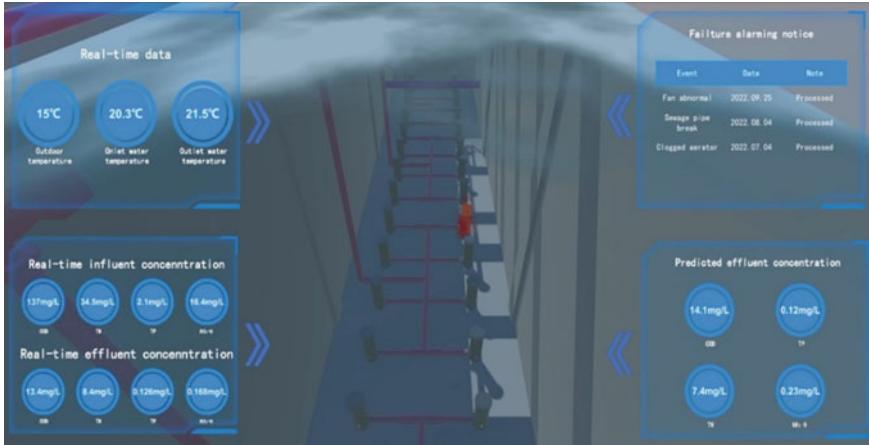


Fig. 4 Sewage management system interface (Laboratory scenario)

4 Results

The running data is driven into the 3D model in real time. The construction of sewage treatment intelligent management system is realized on the web page. The system can present the field information in real time. The system interface mainly displays four parts, which are: real-time temperature, real-time water quality, alarm notification, and water quality prediction. The screen shows the 3D model of the sewage plant, and the information collected by the sensor can be displayed at the location of each equipment in real time. When the system detects that the equipment is about to fail, the equipment will emit red light alarm. The system interface is shown in Fig. 3 and Fig. 4.

5 Conclusion

In this paper, a smart sewage treatment management system is constructed based on digital twin, IoT, machine learning, and other technologies. It can solve the pain point of backward management and service means existing in most sewage management systems, provide important ideas for the wisdom of sewage treatment plants, and meet the management needs of water plants in smart cities. Preliminary experimental results show that the system realizes intelligent operation and predictive maintenance of sewage treatment plant. It saves a IoT of manpower and material resources, has the feasibility, and runs stably, has the good application prospect. With the continuous advancement of smart city, the maintenance and control of plants will realize a leap from digital to intelligent.

References

1. Tušer I, Oulehlová A (2021) Risk assessment and sustainability of wastewater treatment plant operation. *Sustainability-Basel* 13(9):5120
2. Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F (2018) Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol* 94
3. Tao F, Zhang H, Liu A, Nee AYC (2019) Digital twin in industry: state-of-the-art. *IEEE T Ind Inform* 15(4):2405–2415
4. Tao F, Sui F, Liu A, Qi Q, Zhang M, Song B et al (2019) Digital twin-driven product design framework. *Int J Prod Res* 57(12):3935–3953
5. Liu Q, Zhang H, Leng J, Chen X (2019) Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system. *Int J Prod Res* 57(12):3903–3919
6. Park KT, Nam YW, Lee HS, Im SJ, Noh SD, Son JY et al (2019) Design and implementation of a digital twin application for a connected micro smart factory. *Int J Comput Integ M* 32(6):596–614
7. Guo J, Lv Z (2022) Application of digital twins in multiple fields. *Multimed Tools Appl* 81(19):26941–26967
8. Chen M, Mao S, Liu Y (2014) Big Data: a survey. *Mobile Netw Appl* 19(2):171–209
9. Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B et al (2021) Machine learning: new ideas and tools in environmental science and engineering. *Environ Sci Technol* 55(19):12741–12754
10. Zhang S, Wang H, Keller AA (2021) UC Santa Barbara previously published works. *ACS ES&T WATER* 1(12):2531–2540
11. Torregrossa D, Leopold U, Hernández-Sancho F, Hansen J (2018) Machine learning for energy cost modelling in wastewater treatment plants. *J Environ Manage* 223:1061–1067
12. Chen K, Wang H, Valverde-Pérez B, Zhai S, Vezzano L, Wang A (2021) Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning. *Chemosphere* 279:130498

Visual Simulation and Demonstration of Building Quality Improvement in Existing Residential Areas Driven by Integrated Information System



Yue Fan, Li Zhang, Qiong Zhang, Guangxun Cui, Yueer He, and Yu Zhao

Abstract Visual simulation and demonstration supported by digital simulation technology become an important way to promote and demonstrate scientific achievements. At present, many scientific research projects are trying to promote various types of scientific research achievements, such as popular science websites and visualization software. However, most of the projects remain in the scientific popularization of the final scientific research result and related derivative technologies. There is little systematic presentation of the scientific methods and principles behind them, which has affected the promotion effect of scientific research achievements. The project integrates research principles and processes and, under the support of multi-source information integration platform, attempts to build a scientific principle-driven full process visual simulation and demonstration integration platform for the quality improvement of existing residential areas in North China, providing direct feedback for the reconstruction design practice of existing residential areas, and providing scientific data support and theoretical basis for the design practice and standard guidelines.

Keywords Visual simulation and demonstration · Scientific research projects · Building quality improvement

1 Introduction

In northern China, various forms, and scales of renovation of old residential areas have been carried out in the past decade, but most of them are government-led special energy-saving renovation or environmental management projects, with many problems such as lack of basic data, unscientific standards in decision-making, and lack of renovation evaluation. The lack of long-term renovation plans and systematic

Y. Fan · L. Zhang (✉) · Q. Zhang · G. Cui · Y. He · Y. Zhao
Shenzhen University, 3688 Nanhai Ave, Shenzhen, China
e-mail: zhangliszu@szu.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_64

487

scientific methods has restricted the improvement of the overall quality of residential areas. In the scientific system of architecture and planning design, no systematic theory has been established for environmental quality improvement and low-carbon retrofitting of existing residential buildings and settlements, and the existing standards also lack effective methods to deal with such problems. Although theoretical research on low-carbon and energy-saving renovation has certain achievements, the theoretical system and scientific methods in architecture have yet to be established and improved [1]. The key project of the National Natural Science Foundation of China “The basic theory and optimization methods toward quality improvement and low-carbon renovation of residential buildings in Northern China” was carried out by the project team in 2017–2021. Finally, a basic database under the “hierarchical” scientific model was built, the scientific system of diagnosis and evaluation knowledge base based on the correlation analysis of “part-disease” and “disease-diagnosis-repair” was constructed [2], the method of target hierarchical classification and element strategy subdivision was established [3], and the optimization method of transformation strategy based on multi-dimensional simulation and comprehensive evaluation was proposed [4]. Relevant achievements have strong social promotion value and practical guiding significance.

With the continuous progress of digital simulation and demonstration technology, many scientific research projects are trying to promote various types of scientific research results such as popular science websites and visualization software; however, most of the results stay in the popularization of the final results of scientific research and related derivative technologies, but there are few systematic presentations of the scientific methods and principles behind them, which affects the promotion effect of scientific research results to a certain extent. Here, we developed a system which integrated the research principles and processes and attempts to build a scientific principle-driven “data-diagnosis-strategy-evaluation” visualization simulation and demonstration platform for the whole process of quality improvement of existing residential areas in northern China with the support of a multi-source information integration platform, providing direct feedback for the design practice of existing residential area renovation, and providing scientific data support for the design practice and standard. It will provide scientific data support and theoretical basis for design practice and standard guideline development.

2 Method

The system is built on the JavaScript platform and is divided into four main modules, namely current state research and basic database construction; diagnostic assessment and element performance simulation; hierarchical target and renovation strategy generation; and efficiency simulation and strategy policy demonstration.

Firstly, the process of collecting and integrating multiple sources of data includes 54 existing residential settlements built between the 1980s and 1990s in 11 cities, all of which are facing severe Quality degradation issues in northern China, indoor

and outdoor temperature and humidity simulation data of typical old settlements, a digital model of typical settlements, and the content of the hierarchical menu of settlement quality improvement strategies.

After that, the simulation demonstration modules are constructed, including the demonstration module of existing settlement quality improvement principles and workflow, the simulation module of indoor and outdoor temperature and humidity in winter and summer, and the simulation module of settlement quality improvement strategy.

Finally, the simulation demonstration paths of each module are designed according to the quality improvement working principle, as well as the related adjustment and testing work, and the final software architecture is as follows.

3 Result

The final construction of the visual simulation and demonstration platform is mainly composed of the following four parts.

3.1 The Existing Survey of the Live Area and the Construction of Basic Databases

This part will be visually demonstrated from the regional geographical environment, group transportation and layout, residential enclosure system, flat layout, structural structure, residential model, and laws and regulations. At the same time, the quality of the pathological phenomenon and degradation mechanism of the quality of the building and the surrounding environment of the residential area and its surrounding environment can be visually presented. Based on the above work, the information obtained in the early stage is classified and integrated to form a database, which includes the information on the residential area, the texture of the residential area, the standardization and statistical data, the pathological collection and analysis data, the research data of drawing and statistical data, the hot environment survey data, interview survey data, carbon emissions data, case collection, and analysis of the nine sectors of data, which is convenient for all kinds of research data, drawings, documents, and related cases to archive, retrieve, and visualize.

3.2 Demonstration of Quality Degradation Problems and Improvement Target Generation

This part is a demonstration of the research method in combination with the actual case optional information collection (including basic information in the residential area, the basic information of the group, the design drawings of the residential building, the maintenance resume, etc.), the analysis of the deterioration of the problem (including the degradation phenomenon collection, the comparison of the drawing, the correlation between the degradation and the countermeasures), the performance simulation analysis (including the evolution of the performance of the enclosure system, the survey of the current status of the enclosure system, the actual measurement of performance before the transformation) [5, 6], the research purposes, research content, research methods, and research results of the four major sections of quality improvement targets (see Fig. 1).

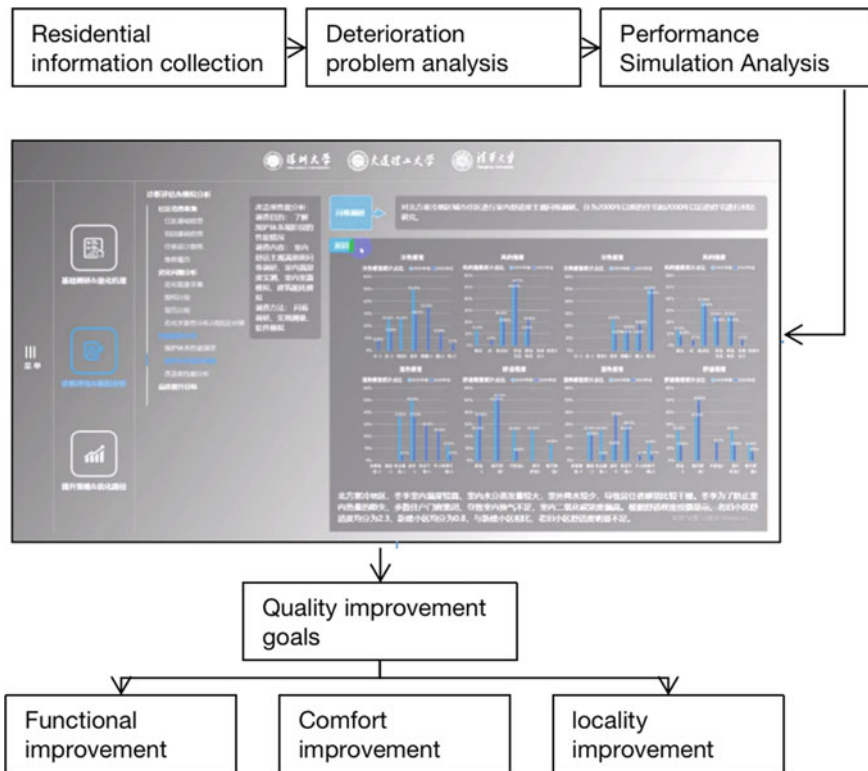


Fig. 1 Demonstration of quality degradation problems and improvement target generation process

3.3 H & Level Target and Quality Improvement Strategy Integration Generation

This part will show the processing process of hierarchical targets (basic type, upgraded type, complete type), as well as the generating process of quality improvement elements (including transformation element extraction, reconstruction method collection, peripheral care tool integration generation, the collection of hierarchical quality improvement elements), based on this, presents quality improvement element sets and visually analyzes energy-saving performance, structural practices, and material styles of various quality improvement elements. Specific content includes the following: (1) increasing the target according to the multi-quality improvement, and based on the results of the problem of infuriating problems based on the relevant standards of the State Council, the hierarchical goals and specific connotations are proposed for residents' actual needs; organizing the corresponding transformation method corresponding to the corresponding elements, building a quality improvement toolset on this basis, and building a menu-based quality improvement tool set according to the specific connotation of hierarchical improvement targets.

3.4 Quality Improvement Strategy Simulation and Efficiency Simulation

This section will rely on the typical residence of the Taishan Residential District to simulate the implementation effect of the existing residential area quality improvement strategy, showing the actual effect of the combination of element combinations under all levels (basic type, perfect, more than perfect type) quality improvement. The performance of each strategy scheme is compared horizontally, and it also involves the introduction of the structure and structure of the key parts. Specific functions include the following: (1) demonstration of the performance optimization effect brought by the quality improvement strategies, such as the annual hot and cold load change value brought by the preservation layer update [7–9]; (2) the slope settings and related detail points of the obstacle slope [10]; and (3) presenting the simulation effect of the quality improvement strategy, such as the visual effects of each elevator and the visual effects of the flat sloping strategy.

4 Conclusion

This system integrates basic data collection and integration, building elements and quality-related research, target decision-making reconstruction methods, and efficiency evaluation optimization method. Relying on the increasingly mature visual

simulation technology, the system has demonstrated the principle of quality improvement research on the typical residential area of the northern China as an example. The four major sections demonstrated the entire process of quality improvement research and their scientific principles and have the functions of digital sharing of scientific research data and renovation strategy simulation demonstration. The design practice in the residential area provides direct feedback to provide scientific data support and a theoretical basis for design practice and standard guidelines.

References

1. Fan Y, Li Z, Zhang Q (2020) Research on scientific framework and knowledge base system of maintenance regeneration on existing residential buildings. *Time Archit* 2020(1):4
2. Fan Y, Li Z, Zhang Q (2020) Research method and quantitative simulations of building quality improvement in existing residential areas based on data analysis. *Contemp Archit* 2020(5):4
3. Li G, Li Z, Li Y (2019) Open building as a design approach for adaptability in Chinese Public Housing. *World J Eng Technol* 2019(07):598–611
4. Zhang N, Cao B, Wang Z, Zhu Y, Lin B (2017) A comparison of winter indoor thermal environment and thermal comfort between regions in Europe, North America, and Asia. *Build Environ* 117:208–217
5. Liu M, Sun C, Zhang B, Wang J, Duan Q (2018) Optimized operation of an existing public building chilled station using TRNSYS. *Buildings* 8(7):87
6. Liu M, Zhang B, Ren J (2018) Sustainability evaluation of the ecological footprint of rural residential house with difference martials. *Marter Sci Eng* 394(3)
7. Ji W, Cao B, Geng Y, Zhu Y, Lin B (2017) Study on human skin temperature and thermal evaluation in stepchange conditions: from non-neutrality to neutrality. *Energy Build* 156:29–39
8. Hong B, Lin B, Qin H (2017) Numerical investigation on the coupled effects of building-tree arrangements on fine particulate matter (PM2.5) dispersion in housing blocks. *Sustain Cities Soc* 34(2017) 358–370
9. Wang Z, Cao B, Lin B, Zhu Y (2020) Investigation of thermal comfort and behavioral adjustments of older people in residential environments in Beijing. *Energy Build* 6(217):110001
10. Zhang L, Fan Y, Yang X et al (2021) Promoting green real estate development by increasing residents' satisfaction. *Sustainability* 13

A Study on the Integrated Database of Sino-Korean Words in <Goryeo sa(高麗史)>



Joo-hee Lee, Qianhang Tang, and Ryang-jin Kim

Abstract In this paper, we introduce a Sino-Korean words Database based on <Goryeo sa(高麗史)> for future research on Sino-Korean words. Based on the original text and translation data of <Goryeo sa(高麗史)>, we will extract a list of Sino-Korean words and converting it into a Database using computational statistical processing methods such as N-gram analysis. The extracted list will be checked for agreement through comparison and contrast with the head word of the existing dictionary, and the meaning will be checked through interpretation and analysis of the original data, and finally Sino-Korean words will be classified by theme. For the purpose of this paper, we plan to complete the list of Sino-Korean words in <Goryeo sa(高麗史)>.

Keywords <Goryeo sa(高麗史)> · Sino-Korean word · Database

1 Necessity for This Study

1.1 A Subsection Sample

It is well known that nearly 70% of Korean vocabulary are Sino-Korean words. Although the main source of Sino-Korean words is China, Korea itself has created and used Korean-style Sino-Korean words through long experience in the Chinese character group. After the modern era the time of enlightenment, as Japanese-style Sino-Japanese words were introduced in large numbers, the lexical of modern Korean Sino-Korean words took on a very complex aspect.

Therefore, in order to clarify the historical origin of Sino-Korean words, it is necessary to accurately organize the appearance patterns of Sino-Korean words by period. Fortunately, relatively recently, a list and interpretation of Sino-Korean words from

J. Lee (✉) · Q. Tang · R. Kim
Kyung Hee University, Seoul 02447, Republic of Korea
e-mail: jfleur@khu.ac.kr

the creation of Hunmin jeongeum(訓民正音) in the fifteenth century to the modern period, edited by Park Jaeyeon (2002, Chinese Korean dictionary of the Chosun Dynasty(중조대사전)), and a significant number of Japanese Sino-Japanese words introduced into Korean were identified by Lee Hanseop (2014, Korean Dictionary from Japanese(일본어에서 온 우리말 사전)). Nevertheless, we are still unable to properly describe the lexical history of Sino-Korean words, which is partly due to the limitations of academia dominated by the lexical history research methodology centered on Korean native words. Above all, it is also because the initial form of the Sino-Korean words formation period before the Middle Ages and its development process cannot be pinpointed.

It is not clear when Sino-Korean words were introduced into Korean, but it is estimated that many Sino-Korean words were established in the period of the Three States.

2 The Subject of This Study

We extract these assumptions from historical documents written by our ancestors, such as <Samguk sagi(三國史記)>, <Samguk Yusa(三國遺事)>, <Goryeo sa(高麗史)>, and <the Annals of the Joseon Dynasty(朝鮮王朝實錄)>, and trace the Sino-Korean words of modern Korean that can be met by going back in time. In this paper, among them, we want to explore ways to convert the Sino-Korean words of <Goryeo sa(高麗史)> into a database (The acronym DB will be used later).

Kim Yangjin (2015) emphasized the existence of Sino-Korean words from the Goryeo period that were used as their own words, which were written by Chinese characters among the < Goryeo sa(高麗史)>. In this thesis, it was confirmed that there are the following three types of Sino-Korean in <Goryeo sa(高麗史)>, focusing on their relevance to modern Korean (Table 1).

Among these, ‘holchi(홀치), sunma(순마), achachi(아차치), bansang(반상), gusa(구사), paochi(파오치), aema(애마), gujong(구종), yeogja(역자), ingjib(잉집), sadaebu(사대부), and hwacheog(화척)’ are the words of the < Goryeo sa(高麗史)>, but whose appearance was limited to the late period in <Goryeo sa(高麗史)>. For these words, it is needed to explain their introduction from Chinese.

The following are words of Sino-Mongolian words (which belong to Han’er language) appearing after King Chungryeol(충렬왕) in <Goryeo sa(高麗史)> (Table 2).

Many of these words continued to be used until the early Joseon Dynasty, but with the exception of some, they are just historical words and not used today. From a historical perspective, these Sino-Korean words have sufficient value as a target for DB construction to be pursued in our project.

Table 1 Three types of Sino-Korean in <Goryeo sa(高麗史)>

<p>A type: Words processed as modern Korean words in <Standard Korean Language Dictionary(표준국어대사전)> : gwihyang(귀향), muyeog(무역), bongwan(본관), haegwan(해관), geojib(거집), eogmae(mae)(억매(매)), gyeong-gi(경기), wonsu(원수), gujong(구종), ingjib(잉집), gyeogwon(계권), gogwa(고과), jonjang(존장), gajang(가장), inli(인리), suyang(수양), hwamae(화매), yagmae(약매), yangin(양인), gaega(개가), yangja(양자), chimgong(침공), gwasil(과실), jeolsang(결상)</p>
<p>B type: Words processed as <history> terminology in <Standard Korean Language Dictionary(표준국어대사전)> : chamsanggwang(참상관), gongjang(공장), jeongli(정리), eumbyeong(은병), hajeon(하전), nongjang(농장), holchi(홀치), sunma(순마), sapae(사패), jaechu(재추), mogli(목리), jeonli(전리), gusa(구사), gwan-si(관·시), aema(애마), naeseung(내승), neumgeub(능급), gongho(공호), yeogja(역사), jablyu(잡류), sadaebu(사대부), ju(주), injeong(인정), obo(오보), jabcheog(잡척), janyeoan(자녀안), hwacheog(화척), myeongbu(명부)</p>
<p>C type: Words that are not included in the <Standard Korean Language Dictionary(표준국어대사전)> but should be included in the category of ‘B’: mijang(미장), chamoegwan(참외관), achachi(아차치), cho(초), bansang(반상), paochi(파오치), bugogin(부곡인), gongchin(공친), danjeong(단정), joyeog(조역), chinhyeongja(친형자), danghyeongja(당형자), jingdong(징동)</p>

Table 2 Sino-Mongolian words (which belong to Han'er language)

<p>ganghwagyeong(江華京/강화경), gyejeomsa(計點使/계점사), gwabucheonyeochugodogam(寡婦處女推考別監/과부처녀추고도감), gwangungwan(管軍官/관군관), gwanja(關字/관자), gyrocho(交鈔/교초), gugu(驅口/구구), manja(蠻子/만자), manjagun(蠻子軍/만자군), gwibugun(歸附軍/귀부군), dongjeong-gun(東征軍/동정군), dumog(頭目/두목), myeonsin(免身/면신), sahwa(賜花/사화), sangdo(上都/상도), saenggwongun(生券軍/생권군), seonhwaagal(宣花呵喝/선화아갈), sunmaso(巡馬所/순마소), sinbugun(新附軍/신부군), ssangseong(雙城/성), aeducham(厓頭站/애두참), yeongyeong(燕京/연경), wangjibyeolgam(王旨別監/왕지별감), eungbang(鷹坊/응방), uigwanjaje(衣冠子弟/의관자제)</p>
--

3 Research Promotion Strategy

Existing research on Sino-Korean words has been mainly focused on the borrowed character vocabulary, and research on the acceptance of actual Sino-Korean words used in Korean (foreign Chinese character vocabulary), the formation of new vocabulary, and the change of these Sino-Korean words is not very active.

The planning intention of this research project is to organize the Sino-Korean words of <Goryeo sa(高麗史)> as a starting point to check the characteristics of the historical development from the origin, development process, and change of the historical development of <Goryeo sa(高麗史)>. It can be said that this is the part that distinguishes with other studies. This study is conducted through the following steps.

In order to proceed with this study, the task that must be preceded is to complete the list of Sino-Korean words which can be used for easy research and arrangement.

(1) Establishment of Sino-Korean words Extraction Processing for Research

For this purpose, using computerized text data, ① by using computational statistical processing methods such as N-gram analysis, the list of Sino-Korean words is primarily extracted and converted into a DB. ② The extracted list is checked for agreement by comparing and contrasting with the head words in the existing dictionary. ③ Then, the meaning is confirmed through interpretation and analysis of the original data. ④ Finally, the Sino-Korean words is classified by the theme of meaning. ⑤ It will be possible to complete the DB of Sino-Korean words in the literature suitable for each research topic to be carried out on a yearly basis. From the construction of the <Yongbieocheonga(龍飛御天歌)> vocabulary dictionary in the 1st year to the 6th year, according to the flowchart below, the database of the Sino-Korean words in the literature for each period will be completed and compiled in the 7th year, which is the ultimate goal of this project.

(2) A study by DB conversion of Sino-Korean words data

In order to extract the Sino-Korean words from the research target data of this study, it is planned to use the 'N-gram statistics' technique to create a DB. The Sino-Korean words calculated through 'comparison and contrast with the existing dictionary headings' from this extracted list will also be converted into a DB, and after classifying the Sino-Korean words by theme, the results can be organized into individual DBs to facilitate research.

(3) Comparison and contrast with headwords in existing dictionaries

In order to secure meaningful results from the list of Sino-Korean words extracted by the 'N-gram statistics' technique, comparison and contrast work is performed with the head words of the dictionaries that are closely related to this study among the previously published dictionaries. The result is checked with the original data to check whether it is a Sino-Korean word or not, and if it is determined to be a Sino-Korean word, it will be organized in a separate DB.

① Sino-Korean words related to Buddhism/Confucianism/Taoism/Traditional Thoughts.

② Sino-Korean words related to food, clothing and shelter, agriculture, fisheries economy, legal system, medical care, and science and technology.

③ Sino-Korean words for person/place/office name.

④ Sino-Korean words that are borrowed character.

⑤ Sino-Korean words that used by Khitan/Jurchen/Mongolian.

4 Conclusion

As discussed above, in this paper, based on the original text and translation data of <Goryeo sa(高麗史)>, ① extracting a list of Sino-Korean words and converting it into a DB using computational statistical processing methods such as N-gram analysis, ② the extracted list is checked for agreement through comparison and contrast with the head word of the existing dictionary, ③ the meaning is checked through interpretation and analysis of the original data, and ④ finally, Sino-Korean words will be classified by theme ⑤ for the purpose of this paper we want to complete the list of Sino-Korean words in <Goryeo sa(高麗史)>.



This paper focuses on the value of <Goryeo sa(高麗史)> as a starting point for Sino-Korean words, and by extracting Sino-Korean words from this document and turning it into a DB, it is expected that a systematic description of the history of Sino-Korean words will be possible, and the plan for DB creation is explored. If the Sino-Korean words in <Goryeo sa(高麗史)> are organized according to the plan of this paper, it will be an important reference point to reveal the origin of the lexical entry and exit of Sino-Korean words in Korean in the future.

References

1. Pelio P (1930) 「Les mots mongols dans le 高麗史 Koryesă». *J Asiat* 217:253–265
2. Kang S (1975) A study on the Kye-Rim-Ryu-Sa. *J. Eastern Stud*, Daedong Institute for Korean Studies, pp 1–92
3. Kim R (2008) A study on Goryeo language in <Goryeo sa(高麗史)>. *J Korean Linguist.* 52, The Society of Korean Linguistics), pp 189–219
4. Kim R (2010) About some place name in Korye-sa(高麗史)(Sikhwaji(食貨志)) ‘Joun(漕運)’ article. *J Place Name Soc Korea* 16, The Place Name Society of Korea, pp 193–226
5. Kim, R.: Preliminary Study for the Synthetical Study of the Place Names in Koryeo-sa[高麗史], *The Study of Korean Language and literature* 46, The Study of Korean Language and literature, 49–83 (2013)
6. Kim R (2015) A Sino-Korean in Koryeo-sa(『the history of Koryeo』), *Korean Historical Linguistics* 21, The Society of Korean Historical Linguistics, pp 239–285

A Case Study on Data Mining for Korean Painting Poetry



Haeyoung Park  and Younghoon An 

Abstract This study introduced the work of analyzing the meaning by extracting the painting poetry of Maehwado (plum blossom paintings) in the early Joseon Dynasty contained in *the Hanguk Munjip Chonggan* (Korean Literary Collections in Classical Chinese) in Hanguk Gojeon Jonghap DB (Korea Classics DB) using the computer information processing method. Through the data mining method, we divided the types by extracting and cataloging the painting poetry from the Korean classical literature. Then, the scope was narrowed down to painting poetry of Maehwado in the early Joseon Dynasty, and the text was analyzed in units of syllable corpus.

Keywords Data mining · Painting poetry · Extracting · Cataloguing · Type classification · Corpus

1 Introduction

This study introduced the work of analyzing the meaning by extracting the painting poetry of Maehwado (plum blossom paintings) in the early Joseon Dynasty contained in *the Hanguk Munjip Chonggan* (Korean Literary Collections in Classical Chinese) in Hanguk Gojeon Jonghap DB (Korea Classics DB) using the computer information processing method. This is one of the basic works of research on Hansi (poems in the Chinese style), which has subdivided types of Hansi according to material. In addition, since this work is based on the data of computer program DB, it is possible to objectively present the expression techniques and meanings of Hansi by type by converting them into numerical values. In the existing traditional research on Hansi, qualitative evaluation based mainly on the researcher's intuition was emphasized. Digital data analysis is a quantitative evaluation method that can supplement this.

H. Park · Y. An (✉)

Kyung Hee University, 26, Kyungheedaero, Dongdaemun-gu, Seoul, Republic of Korea
e-mail: yhnahn@khu.ac.kr

H. Park

e-mail: hy000p@khu.ac.kr

2 Research Methods

2.1 Data

The data for this study, *Hanguk Munjip Chonggan*, is “Hanguk Gojeon Beonyeokwon (Institute for the Translation of Korean Classics) DB”. It is called “Hanguk Gojeon Jonghap DB”, and Hanguk Gojeon Beonyeokwon organizes and translates Korean classics and then digitally converts them to open them to public data. This data is provided in an XML format in the form of OpenAPI [1]. Here, basic data of Korean classics such as author, book name, style, title, original text, and publication year is provided. Through this, we extracted and analyzed the painting poetry data from the *Munjip* of the early Joseon Dynasty.

2.2 Data Mining Techniques

Data mining is a technology that extracts useful information by analyzing statistical patterns, rules, and relationships from large amounts of data [2]. This technology is used in various fields such as statistics, computerization, and management today. I also tried to apply it to the research on Korean classical literature. Currently, there are 142 anthologies in the early Joseon Dynasty included in the *Hanguk Munjip Chonggan*. First, we selected painting poetry based on this data and classified its type. This is the primary mining of painting poetry data. Next, we analyzed the text by narrowing the scope [3] to painting poetry of Maehwado in the early Joseon Dynasty. This is the secondary mining of painting poetry data.

3 Result

3.1 Extracting and Cataloguing

We chose poetry for style at Hanguk Gojeon Jonghap DB and selected and cataloged poems with characters such as “je (mention)”, “do (picture)”, “hwa (painting)”, “hoe (drawing)”, “sa (drawing)”, “byeong (folding screen)”, “jok (hanging scroll)”, “cheop (album)”, “chuk (scroll)”, and “muk (ink)” in the title. After that, we checked the title and content of each poem to determine whether the poem was painting poetry or not. Through this, it was confirmed that there are a total of 842 painting poetry currently included in *Hanguk Munjip Chonggan*.

3.2 Type Classification

We classified painting poetry of the early Joseon Dynasty, which was extracted earlier, by type based on the objects in the painting [4]. This is possible because the material of the picture is presented as an individual object name before and after terms such as “do” and “hwa”, which usually means a picture. For example, most painting poetry titles of Maehwado appear in expressions such as “Maehwado” or “Jemae (reciting a plum blossom)”. The criteria divided by referring to the types of paintings in the early Joseon Dynasty are as follows (Fig. 1).

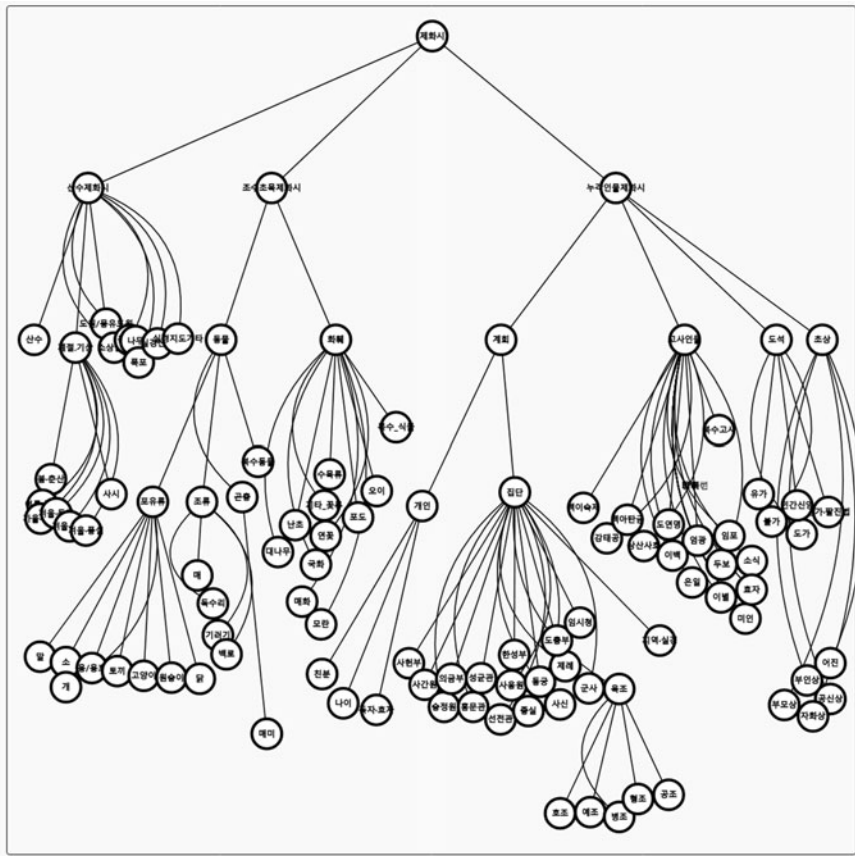


Fig. 1 Type classification of painting poetry of the early Joseon Dynasty

Table 1 Corpus unit of o’eon and chil’eon poem

If 1 line of o’eon poem is assumed to “1 2 3 4 5”, the corpus unit	If 1 line of chil’eon poem is assumed to “1 2 3 4 5 6 7”, the corpus unit
One-syllable corpus (5): 1/2/3/4/5	One-syllable corpus (7): 1/2/3/4/5/6/7
Two-syllable corpus (4): 12/23/34/45	Two-syllable corpus (6): 12/23/34/45/56/67
Three-syllable corpus (3): 123/234/345	Three-syllable corpus (5): 123/234/345/456/567
Four-syllable corpus (2): 1234/2345	Four-syllable corpus (4): 1234/2345/3456/4567
Five-syllable corpus (1): 12345	Five-syllable corpus (3): 12345/23456/34567

3.3 Syllable Unit Segments and Statistics

Next, we analyzed the text data of painting poetry of Maehwado in the early Joseon Dynasty. In order to interpret a poem [5], it is convenient to divide it into morphemes, the smallest unit of meaning. Since Hansi uses Chinese characters, it can be divided into morphemes up to one-syllable unit. In addition, the unit of this syllable is basically cut into rows to form a layer. Therefore, we first used one line as the basic unit, and among the original text of painting poetry, o’eon (five-character) was divided into 1–5, and chil’eon (seven-character) was divided into 1–7. Then, per line, o’eon generated five corpus of one syllable, two syllables, three syllables, four syllables, and five syllables, and chil’eon generated seven corpus of one syllable, two syllables, three syllables, four syllables, five syllables, six syllables, and seven syllables (Table 1).

The following is the result of dividing the original data of painting poetry by syllables and calculating the frequency. This is the actual figure measured using the JAVA program to see how many syllable corpus units are repeated in the actual painting poetry text. The one–five syllable data value is the sum of the data for each syllable of o’eon and chil’eon, and six–seven syllables are data values for chil’eon only. In these values, only words from one to three syllables were actually valid, and all four–seven syllables were meaningless because the word combination was once. Here, we present only one-syllable corpus usage frequency data (Fig. 2).

3.4 Analysis of the Meaning of Corpus

The most commonly used one-syllable word in painting poetry of Maehwado is “hyang (scent: 38 times)”. “Hyang” is a combination of words in the order of “mu hyang (no scent: 6 times)”, “cheong hyang (a clear scent: 5 times)”, “yu hyang (a mellow scent: 5 times)”, “hyang ip (The scent permeates: 5 times)”, “hyang dong (The scent spreads: 5 times)”, and “am hyang (a delicate scent: 3 times)” in two syllables. Next is “bul (no: 37 times)”, which belongs to an adverb that represents negativity in Chinese characters and is not very meaningful in single syllables, but

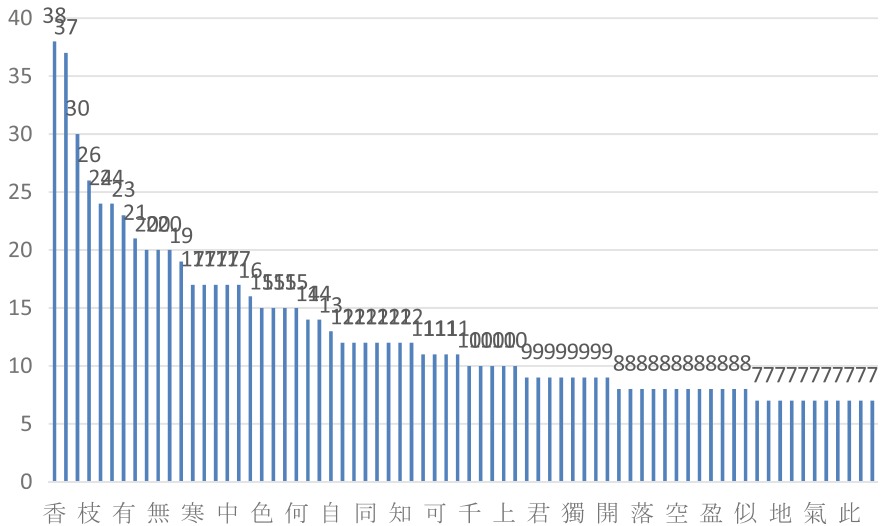


Fig. 2 One-syllable corpus of poetic diction

further emphasizes its meaning when combined with other words. The third is “seol (snow: 30 times)”, which is connected in two syllables: “yeom seol (a fine snow: 3 times)”, “bing seol (ice and snow: 3 times)”, and “jeok seol (snow piled up: 3 times)”. The keywords repeated in painting poetry include “hyang” and “seol”, which are consistent with the conventional symbols of plum blossoms, “amhyang budong (The delicate scent doesn’t spread)” and “flowers blooming through the cold in the snow”.

4 Conclusion

The painting poetry is a product of social communication. In the schema connected with “object-picture-poetry”, the original meaning of an object is transformed into a form or language. This meaning is a symbol, and the meaning of this symbol was shared among the literati at the time when they saw the poem and the painting. The symbolic meaning of objects in painting poetry is the result of communication between painters, poets, literati, society, and times.

The more people share a symbol like this, the greater its influence. In the same type of painting poetry, a series of identical corpus are repeated. The meaning of this corpus is the symbol of the picture, and it is the symbolic meaning of the object in the picture. Through this, the symbolic meaning of objects in the picture can be analyzed and derived more clearly.

References

1. Lee B, Min K (2021) A study on visualization of the analysis between the collections of Korean literature in Korea classic DB: Hanguk Gojeon Beonyeokwon: national culture, vol 57, pp 5–32
2. [ko.wikipedia.org/wiki/Data Mining](https://ko.wikipedia.org/wiki/Data_Mining). Last accessed 2021/10/15
3. Kwon K, Choi W, Kim D (2022) A lengthwise comparative study of different versions of Yadam-based on “Ok So-seon”. Korean classical studies, vol 57. Korean Classical Research Society, pp 87–120
4. Park H (2021) A study on the painting poetry of the former part of the Joseon Dynasty. Kyunghee University Graduate School, pp 19–34
5. Lee B (2019) A study on the construction and utilization of Korean Chinese poetry corpus: geun-yeoghanmunhaghoe: hanmunhagnonjib, vol 53, pp 153–177

A Study on Digital Content Using Yun Dong-Ju's Works



Minji Choi, Hwa-Young Jeong, Young-Hoon An, and Peerapat Luanmongkol

Abstract Since modern literature is a discipline that has no choice but to reflect reality, it has infinite possibilities for content conversion. In addition, this is even more so if you estimate the vast amount of the original data. Therefore, when the current social atmosphere is said to be progressing through convergence between various studies, it is very important to examine the contents of literature that exist as the basis. Therefore, this study attempted to examine the case of poet Yoon Dong-Ju, who has already successfully achieved contentization in various fields, and to present the criteria for it. In addition, the possibility was also examined by presenting the case of other writers showing proximity to the preceding model.

Keywords Modern literature · Content · Storytelling · Yoon Dong-Ju · Sky, wind, stars, and poetry

1 Introduction

Until now, modern literature has reflected various times under the theme of “literature = reality.” In particular, such trends were more pronounced in the era including historical events such as the Japanese colonial period, the Korean War period, and the military dictatorship. For example, the works of poets representing the Japanese colonial period, such as Han Yong-woon, Lee Yuk-sa, and Yoon Dong-Ju, contain the

M. Choi · Y.-H. An · P. Luanmongkol
Department of Korean Language and Literature, Kyung Hee University, 26, Kyungheedae-ro,
Dongdaemun-gu, Seoul 02447, Republic of Korea
e-mail: yhnahn@khu.ac.kr

P. Luanmongkol
e-mail: ryusehyun96@naver.com

H.-Y. Jeong (✉)
Humanitas College, Kyung Hee University, 26, Kyungheedae-ro, Dongdaemun-gu, Seoul 02447,
Republic of Korea
e-mail: hyjeong@khu.ac.kr

common emotions of Koreans at the time, and accordingly, their works are used as various educational and research materials, and are still considered a favorite poem of many people. This fact reminds us that the more strongly a work that reflects the reality, the higher the possibility that it can be used as various contents.

Among them, Yoon Dong-Ju is a representative case. Yoon Dong-Ju, who has been considered one of the most beloved poets of Koreans so far, has appeared in numerous aspects in various categories [1]. Even in the case of the primary data, *Sky, Wind, Stars, and Poetry*, as copyright protection was lifted in 2016, it was published in various forms such as reproductions, manuscript poetry, and calendar poetry. In addition, it was the same in the fields of movies, musicals, plays, and music. In short, it is proof that literature is no longer limited to the center of narratives, and that cultural discourse is already formed in a more multi-layered manner through “storytelling,” the basis of cultural content forms of popular consumption culture [2].

On the other hand, the current status of domestic technology construction has not progressed to the stage of complete “telling” and remains only in the “story” itself [3]. These limitations are not considered a problem of the technology itself. Rather, due to the vast amount of primary data, it is understood as a practical limitation resulting from the lack of sufficient reflection of narrative characteristics and requirements when it is converted into content. Therefore, systematically examining the preceding models that have successfully achieved contentization in various sectors is expected to be a major data to examine other possibilities. In addition, as the current social atmosphere is already proceeding with active convergence between various disciplines beyond the convergence stage of art and engineering, we would like to review the correct recognition and use of literature, which is the basis, to suggest the direction of related fields.

2 Related Works

2.1 *Yoon Dong-Ju’s Life and Works*

The poet Yoon Dong-Ju (1917–1945) has been considered one of the most beloved poets of Koreans. He was born in Bukgando Island on December 30, 1917, graduated from Yeonhui College in Seoul, and entered Kyoto Doshisha University in 1942, but was arrested a year later on charges of independence movement. After that, on February 16, 1945, he died in prison without achieving his dream of becoming a poet, and it was a short life of about 27 years. However, his desire for literature could be revealed to the world in 1948 under the name of *Sky, Wind, Stars, and Poetry*. Since then, the collection of poems was edited and published twice in 1955 and 1976, and in particular, in 1999, the collection of poems written by Yoon Dong-Ju was published under the auspices of the bereaved families. In this book, there are

many traces of retort that have been rewritten several times over one work, which gives an indication of how great Yoon Dong-Ju's passion and anguish for poetry was.

It is noteworthy that Yoon Dong-Ju has a "short writing period" for about a year. From September 1939 to December 1940, he chose silence at the beginning, and this is quite unusual to recall an anecdote in which he did not give up the will of literary books even though he had a conflict with his father when he graduated from middle school. However, the fact that "Hospital," "Counting the Stars," and "Seoshi" were all mentioned as his representative works, and that he planned to publish a collection of poems after that period suggests that "Julpilgi" had a great influence on his poetic transformation. First, it can be recalled that he faced an inflection point in the poetic worldview by entering the liberal arts department of Yeonhui College and devouring the works of foreign artists such as Rilke, Valéry, and Gide. Second, it is a case where his religious worldview, which was born with Christian ideas as a connection to the Christian worldview, has passed on to literary imagination through a series of processes.

As described above, Yoon Dong-Ju was a person who chose to complete himself to the end for the aesthetic and ethical completion of life even in the face of the grim reality of the Japanese colonial period after going through the process of exhaustion several times [4]. That would be the "moral purity" that represents him. In addition, the constant love of his literature containing that spirit or the use of various contents can be seen as proof that it is not limited to a specific time or space. Therefore, access and utilization from a more continuous and multifaceted perspective are required.

2.2 Contents Related to Yoon Dong-Ju

Yoon Dong-Ju's work and life have been used as various contents so far. Representatively, there is the most well-known movie "DongJu." The movie "DongJu" is director Lee Joon-ik's third film, which is known for its popularity, cinematic quality, and box office success. Actor Kang Ha-neul appeared in the role of Yoon Dong-ju, and actor Park Jung-min appeared in the role of Song Mong-gyu, his cousin of King Gojong and lifelong friend. In addition, actor Moon Sung-geun appeared in the role of Jeong Ji-yong, the best poet of the time, and the meeting between Yoon Dong-Ju, a young literary man who was only passed down as a retrospective statement, and Jeong Ji-yong, a literary man, met his cinematic imagination and actually took place. The film, which was released on February 17, 2016, attracted 1.17 million viewers, recording \$7.8 million at the World Box Office, despite its low-budget production.

Next, there are musicals and plays. First, there is a creative musical, "Yoon Dong-Ju, a man who becomes more and more transparent." Since its premiere in 2014 under the title of Yoon Dong-Ju's representative work *Seosi*, the work has faced the audience under the title "Yoon Dong-Ju, a Man Who Becomes More and More Transparent" to mark the 100th anniversary of Yoon Dong-Ju's birth in 2017. Although it has not been officially revealed, a scene related to a biological experiment believed to have been Yoon Dong-ju's decisive cause of death added imagination to the stage. In

addition, the play “Yoon Dong-ju, Shoot the Moon” has Yoon Dong-Ju’s few prose titles as the title of the work, and it was reenacted in 2013 after its premiere in 2012, and attracted so many people that the third performance took place in 2016.

Yoon Dong-Ju’s work was also transformed into a song, not a video. It is the existence of the band “Snowing Map,” which was formed with the intention of promoting Yoon Dong-Ju’s poem with a song. Likewise, they also decided on the band’s name with Yoon Dong-Ju’s poem “Snowy Map.” They are making songs based on Yoon Dong-Ju’s work as song lyrics, and they also translate it into English. As they have a representative work featuring Yoon Dong-Ju’s love poem “Soon!” in their name, they are not limited to the Japanese colonial period but are receiving favorable responses from many people with universal values that can still be used until now, and they are continuing to perform until relatively.

3 Contemporary Literature and Content

In order for modern literature to be actively contented, it is necessary to first approach what kind of literary historical meaning network can be reached. The following Table 1 summarizes the examples according to the object’s work for this purpose, and Table 2 divides them into life history elements.

In addition to the above-mentioned cases, it can be seen that various content conversion is underway according to the work. In particular, the exhibition <Walk and Self-Portrait> is an exhibition content using media and attracts attention as it even leads to the experience of the audience. This is because the media panel installed on the wall has the effect of reflecting the face on the actual well when the work is screened. Along with the work, you can see your face reflected in the water.

The following is the case of writers during the Japanese colonial period who share lifelong characteristics. Among them, if it shows a closer meaning, such as Christianity/poet/Rilke, it is marked with an underline, and if it shows proximity in duplicate, it is marked in bold. On the surface alone, it can be seen that there are many writers who show proximity to share the times. This suggests that more modern literature content can be made based on the preceding model.

Table 1 『Sky, Wind, Stars, and Poetry』 which is used as various contents

『Sky, Wind, Stars, and Poetry』		
『Epigraph』	『Self-portrait』	『Star-counting night』
Video content: Movie <Innocent>	Performance contents: Play <Self-Portrait>	Video content (TV drama): Mini series <Counting the Stars>
Exhibition contents: an exhibition <Epigraph—A poem that sings the wind and stars that depicts the sky>	Exhibition contents: Exhibition <Walk and Self-Portrait>	Video content (TV program): Variety show <Counting the stars with you>

Table 2 Writers during the Japanese colonial period who share Yoon Dong-Ju’s lifelong characteristics

	Christianity/Poet/Lilke
A religious poem	Kim Hyun-seung (1913–1975) Park Doo-jin (1916–1998) Han Yong-woon (1879–1944)
Yeonhui College of Professional Studies	Mo Ki-Yoon (1912–1983) Park Yong-cheol (1904–1938) Seol Jung-sik (1912–1953) Song Mong-gyu (1917–1945) Lee Eun-sang (1903–1982) Lee Hee-seung (1896–1989) Jeong Byung-wook (1922–1982) Jo Poong-yeon (1914–1991)
Overseas literature	Park Yong-cheol (1904–1938) Baekseok (1912–1996) Seo Jeong-ju (1915–2000) Lee Won-jo (1909–1955) Lee Yuk-sa (1904–1944)

4 Conclusion

Modern literature has infinite possibilities for contentization. It is true that due to the vast amount of originality, it is not able to move on to the stage of complete “telling” only focused on “story,” but if the vast amount is successfully contentized, it will fully satisfy the desire for convergence pursued by the current academic atmosphere. Therefore, in this paper, we tried to present the basis by looking at the case of Yoon Dong-Ju, who has already successfully achieved contentization. Based on this, if a topic map is constructed in the future or text mining techniques are used, the gap between modern literature and content will be easily supplemented. Therefore, the point will be left as a future task of this study.

References

1. Kim E (2017) Memory of Yoon Dong-ju—March 2017: Yoon Dong-ju’s perception of Yoon Dong-ju: Yoon Dong-ju research, vol 10. In: Research of Korean literature, vol 62. Korean Literature Research Society, pp 320–323
2. Ryu E (2009) Narrative and storytelling: from literature to cultural content. In: Humanities content, vol 14. Humanities Contents Society, p 231
3. Han H (2013) Academic expansion and convergence of literature. In: Korean language research, vol 46. Korean Language Literature Society, pp 424–426
4. Kim W (1977) The situation of writers under Japanese imperialism. In: Poet of a poor age. Minumsa, p 55
5. Yoon D (2004) In: Jeong H et al. The complete collection of Yoon Dong-ju in the original contrast, sky, wind, stars, and poetry. Yonsei University Press, p 325
6. Jeong H, Ko I (2012) Application and design of contemporary literature content ontology using topic maps, vol 10, no 6. Korea Digital Policy Society, pp 215–216

Digital Transformation of Simulation Process to Support Decision-Making on Green Remodeling



Woo-Seung Yun , Wontaek Ryu , Ye-Bon Jo , and Hyuncheol Seo 

Abstract In this study, we propose a model that supports green remodeling decision-making through the digital transformation of the simulation process. Energy performance improvement measures are defined through a rule-based filter, which saves computation time and resources in the automation process. The derived combinations of improvement plans are written into a simulation file through an automated process to gain improvement plan options as an outcome, which fit the goal and budget for saving energy. In this study, model tests were performed at the sites where the green remodeling project was implemented. Computing time required for each case study building was approximately 40 min, and various efficient improvement methods were derived compared to the existing ones through experts.

Keywords Green remodeling · Digital transformation · Building energy simulation

1 Introduction

Various carbon reduction policies have been introduced worldwide to achieve net zero by 2050. In order to reduce the energy consumption of existing buildings, the Korean government has carried out the ‘Green Remodeling Project’ in earnest from 2020. Green remodeling is a type of retrofit of old buildings to provide a pleasant and healthy living environment and to reduce greenhouse gas emissions by increasing energy efficiency [1].

In order to proceed with green remodeling, it is necessary to evaluate the energy performance of a building. To this end, in Korea, a program called ECO2-OD produced by the Korea Energy Agency has been used. ECO2-OD, which was built in compliance with ISO-13790 [2], can evaluate the energy consumption and carbon emission of a building. Using the ECO2-OD program, one can enter information

W.-S. Yun · W. Ryu · Y.-B. Jo · H. Seo (✉)
Kyungpook National University, Daegu, South Korea
e-mail: charles@knu.ac.kr

on the thermal transmittance of the building envelope, such as walls, windows, and floors, and information on equipment systems installed in the building.

In order to accurately evaluate the energy efficiency of a building, an expert with a certain level of qualifications evaluates the building's performance. Experts make decisions based on experience to select ways to improve energy performance. However, this process causes deviations in results depending on experts. In addition, trial and error must be undergone before deriving an optimal solution by simulating various combinations of improvement methods, which requires a lot of time and human resources. Considering that green remodeling is frequently carried out in several hundred sites every year (841 sites as of 2021 [1]), automation of such a decision-making process is in high demand. Therefore, in this study, we propose a model that supports green remodeling decision-making through the digital transformation of the simulation process.

2 Methodology

Figure 1 shows the workflow of the model for supporting green remodeling decision-making proposed in this study. The model is largely composed of three steps: preparation, RPA (robotics process automation), and output. In the preparation step, measures for improving energy performance that can be applied for green remodeling and the construction cost DB are built.

In this study, the measure DB provided by Greeny [3] was used. Greeny is a support system for the green remodeling decision-making of public buildings. Greeny provides information on various measures for retrofitting buildings. For example, data such as the type and thickness of insulation materials, heat transmittance, and construction cost for improving the thermal transmittance of exterior walls to a legal level are provided.

In the Greeny DB, information on a total of 39 types of measures for improving the energy performance of a building and construction costs is provided: 16 walls, 8 roofs, 6 windows, 1 lighting, 4 heating and cooling devices, 2 water heaters, 1 renewable energy, and 1 control section. These 39 measures can be applied to a building in various combinations, with approximately 1400 combinations available.

However, it is inefficient to find the optimal solution through all combinations. Therefore, in this study, a rule-based measure filter was introduced. The filter removes factors that are difficult to be introduced in reality in advance or reduces the number of possible combinations of measures by putting frequently applied factors into combinations, thereby saving computation time and resources. In this study, we established a rule-based filter through a case analysis of the 2020 green remodeling project. Examples of established rules are as follows: (1) If there are fluorescent lamps, always replace them with LEDs, and (2) always replace windows instead of additional insulation installation. Improvement measures were filtered through the above rules to reduce the number of cases from 1400 combinations to a total of 96 combinations.

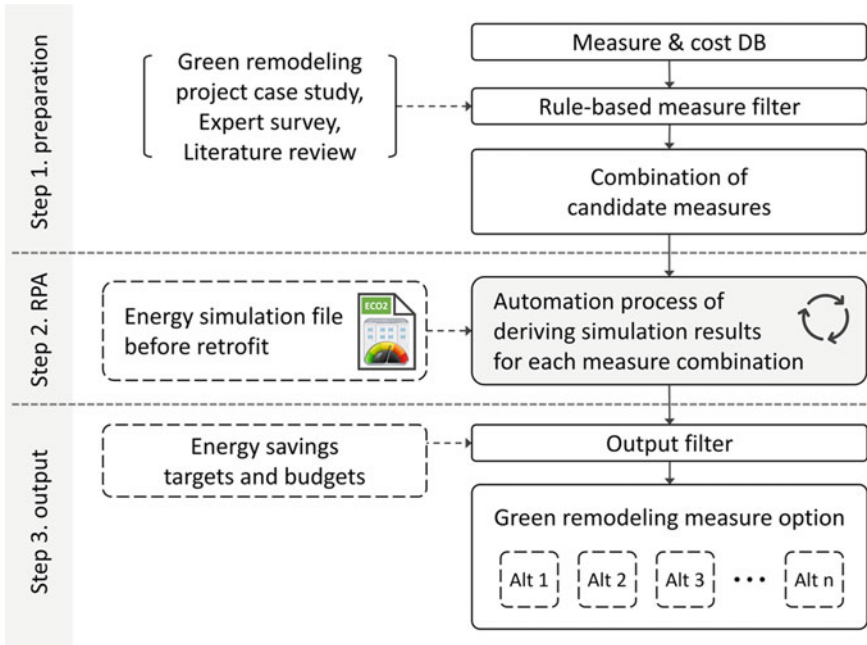


Fig. 1 Flowchart of the proposed model

In the second process, RPA, these 96 improvement methods derived from the above process are applied to the energy simulation file before retrofit (eco2-od file), and the results are obtained for each combination. In this process, Python and the pywinauto library were used. Pywinauto is a library utilized to automate Windows programs through the Python language and is often used for RPA. In the RPA process, the variation of energy simulation input values for each measure is first defined. For example, in the case of wall insulation reinforcement, information such as thermal transmittance and thickness of the wall is changed. Then, depending on the combination of different measures, changes in the corresponding measures are reflected in the eco2-od file. Such a process was repeated automatically for all combinations to obtain a total of 96 energy simulation files and information, such as energy consumption and construction cost after improvement.

The final output step filters items to select those satisfying the project goals and budget among the various options derived in the above process. In this study, the filter was applied with an energy-saving rate of 30% and the construction cost suggested by the consulting company was the upper limit. After filtering, the results for each measure set were finally derived as a model output.

The proposed model was tested through case study buildings where the green remodeling project was implemented. Table 1 shows information on the case study buildings. The buildings are located in Gyeongsangbuk-do, Republic of Korea, and the use is for a public health clinic and a public health unit each.

Table 1 Information of case study buildings

Building id	Story	Gross floor area (m ²)	Primary energy consumption before retrofit (kWh/m ² year)	Year built	Building typology	Construction
1	2	167.58	278.5	2006	Public health clinic	Reinforced concrete
2	2	369.59	169	2004	Public health unit	Reinforced concrete

The target sites are the buildings undergone the green remodeling project in 2021. The consulting company (expert) derived a plan for the energy performance improvement and estimated the construction cost of the buildings. In this study, we compared the improvement methods derived from the model to those suggested by experts for the corresponding buildings.

Through an automated process, the 96 combinations of the improvement plans derived through preparation (step 1) were applied to the energy simulation files before the retrofit of the above sites to obtain the energy-saving rate for each case, estimated construction cost, and an energy-saving coefficient.

The energy-saving coefficient is calculated as the primary energy consumption saving rate [%] \div construction cost [KRW] $\times 10^9$; the higher the coefficient, the greater the energy savings achieved with a small construction cost.

3 Results

Table 2 shows the test results. Tests were performed on a PC with an Intel i7-9700K@3.60 GHz CPU, 32 GB RAM. When executing the calculation of the automated process for each site using only a single core, the computation time was approximately 40–41 min, and 36–52 combinations of improvement measures that satisfy the target for an energy-saving rate of 30% were derived. Considering that the traditional green remodeling decision-making process by experts usually takes more than three days, the computation time of 40 min can be said to be remarkably effective. In addition, it is judged that this model can facilitate a comparative review in the final decision-making process among the experts by providing various options.

Table 2 RPA results

Building id	Required time (min:sec)	Number of options outputted
1	40:17	52
2	41:26	36

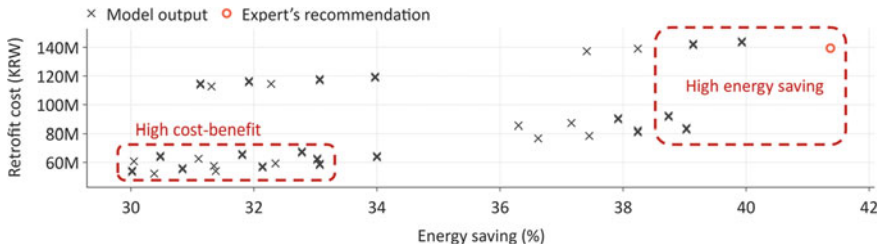


Fig. 2 Comparison of improvement plans for the case study building 1

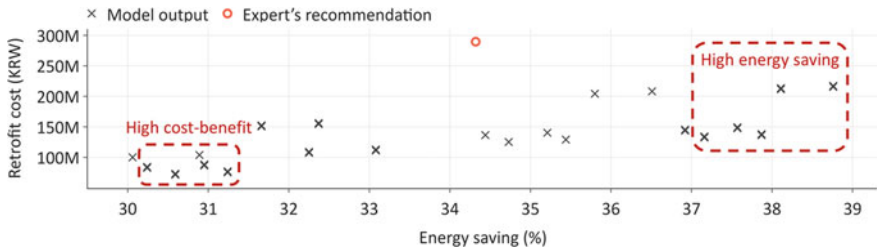


Fig. 3 Comparison of improvement plans for the case study building 2

Figures 2 and 3 show the combination of improvement measures obtained from the model and the expert for each case study building.

In Building 1, the improvement plan by the expert showed the highest energy-saving rate. However, in terms of the energy-saving coefficient, the suggested improvement plan was not an optimal option. The model showed that there is an option to significantly reduce the construction cost by lowering the energy-saving rate to around 39% (lower left of high energy-saving group in Fig. 2), depending on the situation.

In Building 2, the improvement method output by the model showed a higher energy-saving rate than the expert's option. The energy-saving coefficient was also the lowest in the option suggested by the expert. Based on these results, it is suggested that if the green remodeling option is designed by considering the output value of the model, the budget can be used more efficiently.

4 Conclusions

In this study, a model for supporting green remodeling decision-making was proposed. The model can derive various energy performance improvement plans through an automated process in addition to saving computation time and resources by using a rule-based filter. Accordingly, using the model in making decisions on green remodeling can be more efficient than the existing method of consulting

experts, which can facilitate the consulting process itself. Such an automation process is necessary as measures to improve energy performance should be derived for several hundred sites every year due to the nature of the green remodeling project. In some cases, more efficient improvement measures were obtained from the model, which experts could not derive from the test results. Furthermore, the model provides various options to allow decision-makers to consider more diverse options and enables a comparative review of options by experts.

In fact, among the options output from the model, there may be some options that are not applicable due to the actual circumstances of the building. However, such inapplicabilities can be improved by upgrading the filter part of the model. In the future, it is expected to build a DB for energy simulation input–output through this automation process to use it for the development of a surrogate model for energy simulation.

Acknowledgements This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry and Energy, Republic of Korea (No. 20204010600060). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1C1C1007127).

References

1. Green remodeling homepage. <https://www.greenremodeling.or.kr>. Last accessed 2022/10/28
2. ISO (2008) ISO 13790: energy performance of buildings—calculation of energy use for space heating and cooling. International Organization for Standardization
3. Greeny homepage. <https://greeny.website>. Last accessed 2022/10/26

An Application of Deep Neural Network in Facade Planning of Coastal City Buildings



Xue Sun, Yue Wang, Ting Zhang, Yin Wang, Haoyue Fu, Xuechun Li, and Zhen Liu

Abstract Architecture does not exist in a vacuum. Its culture, concepts and aesthetics are constantly influenced by other visual and artistic disciplines, from film, photography, painting and sculpture to fashion, graphic and industrial design. In the field of architecture, design generation work requires architects to collect a lot of data and understand many details before they can design effectively. In the era of big data, design work is becoming more and more complex, and architects are increasingly required to effectively learn and manage the data in front of them, and the process of learning data can actually be done through machine learning. As a generative tool that can effectively learn a large amount of data, the generative deep learning model has played a good role in some architectural design generation work. The main goal of this research is to find inspiration for new aesthetic paradigms outside the architectural discipline. The experimental architectural landmark selection of this paper selects the buildings in the center of Qingdao, China, as our experimental site. Through deep learning and neural network, aiming at the content images of architectural elements of interest, while maintaining the architectural characteristics of the content images, they are stylized with artistic images to provide effective assistance for the facade design of modern urban high-rise buildings.

Keywords Neural style transfer · Facade design · Deep learning

1 Introduction

The concept of style transfer is to imitate the brushstroke style of some artist on an image, so that a photo can become a composite artwork. Originally from the computer vision field, it was studied as a generalization topic for texture synthesis. In 2015,

X. Sun (✉) · Z. Liu
Shandong University of Science and Technology, Qingdao 266590, China
e-mail: zhliu01@126.com

Y. Wang · T. Zhang · Y. Wang · H. Fu · X. Li
Qingdao Agricultural University, Qingdao 266109, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_69

517

Gatys et al. [1] inspired by Convolutional Neural Networks proposed a Neural Style Transfer Algorithm, which uses a pre-trained convolutional neural network to extract artistic images. Style features, extract content features from the content image, and reorganize them on a new image. The ultimate goal is to match the style features of the new image with those of the artistic image, and the content features to match the content features of the content image match. The algorithm does a good job of organically blending a wide variety of artistic styles with images.

The neural style transfer model is also a model that uses neural networks to automatically learn and create. Different from the generative adversarial network, neural style transfer only needs a style image and a content image to complete the generation work, while generative confrontation The network requires a lot of data, and in terms of image resolution, neural style transfer has only a minimum resolution limit, and there is no upper limit on the maximum resolution within the allowable range of computer performance, while the adversarial generation network generally has a higher resolution on the input and output images, strict requirements. Neural style transfer matches style features of a reference image with content features of a target image. It does not explicitly learn to extract features, as pre-trained convolutional neural networks are typically used for this purpose. This is similar to the way of finding the input–output matching function in the structured learning model, except that the style transfer uses the target image to match the content image and style image. Therefore, neural style transfer can actually be regarded as a top-down generation method, and its exhaustive selection is accomplished by self-update of pixels on the target image. Numerous follow-up studies have improved or extended the neural style transfer algorithm based on it, and many successful industrial applications have also been born.

Architecture does not exist in a vacuum, and its culture, concepts and aesthetics are constantly influenced by other visual and artistic disciplines, from film, photography, painting and sculpture to fashion, graphic and industrial design. In the field of architecture, design generation work requires architects to collect a lot of data and understand many details before they can design effectively. In the era of big data, design work is becoming more and more complex, and architects are increasingly required to effectively learn and manage the data in front of them, and the process of learning data can actually be done through machine learning. As a generative tool that can effectively learn a large amount of data, the generative deep learning model has played a good role in some architectural design generation work. The application research of machine learning in the field of architecture started very early, and the research on generative deep learning started a little later.

Because neural style transfer model does not require a lot of data for training and can get good results quickly, it is helpful to inspire design ideas in the early stage of design, so there are also many related studies in the field of architecture. Campo et al. [2] from the University of Michigan used a neural style transfer algorithm to transfer the style of the urban texture of Barcelona to the lunar site in 2020 and argued that humans always classify the urban texture of different cities according to visual characteristics, while neural style transfer algorithm can just transfer the style according to the visual features, so as to demonstrate the significance of artificial

intelligence for architectural design. The final generated result is not particularly good, but as a design reference, the migrated texture can reflect the original texture pattern of Barcelona to a certain extent. Özel and Ennemoser [3] from UCLA used neural style transfer algorithm to transfer some artistic pattern textures to building facades and reconstructed 3D models based on 2D images, so as to carry out certain style transformation design for the facades of urban buildings. And discuss that architectural design is actually often influenced by other visual design disciplines, and neural style transfer algorithm can well integrate the content and elements of these visual art works into the building facade, so as to propose artificial intelligence conceptual design ideas for cross-disciplinary artistic cooperation.

The main goal of this research is to find inspiration for new aesthetic paradigms outside the architectural discipline. The experimental architectural landmark selection of this paper selects the buildings in the center of Qingdao, China, as our experimental site. Through the neural style network, aiming at the content images of architectural elements of interest, while maintaining the architectural characteristics of the content images, they are stylized with artistic images, providing effective assistance for the facade design of modern urban high-rise buildings.

2 Methodology

The architecture of the neural style transfer model is mainly composed of a pre-trained convolutional neural network as shown in Fig. 1. The convolutional neural network is used to extract the content features on the content image, extract the style features on the style image, and recombine and synthesize a target image. The image tends to be the input style image in terms of style and tends to the input content image in terms of content structure. The loss function is to calculate the difference between the style feature of the reconstructed image and the style feature of the style image, plus the difference between the content feature of the reconstructed image and the content feature of the content image, and the style difference and the content difference can be calculated by the weighted sum, way to control the ratio of each other. The parameters of the convolutional neural network in the model are fixed, and the loss function is only continuously iteratively optimized for the pixel values on the reconstructed image through backpropagation.

In the neural style transfer model, content features and style features are the key concepts, and they are also the main factors affecting the effect of model generation. The content feature is the feature extracted by the convolutional neural network from the image, and this feature information can be used to describe the content elements, structure, etc., on the image. The VGG19 (see Fig. 2) used in the general neural style transfer model is a standardized convolutional neural network containing 19 convolutional layers and fully connected layers. Different layers can extract different feature information. The deeper the convolutional layer, the more abstract, high level can be extracted. Characteristics. A convolutional layer can only extract some simple linear edge features. The deeper the layer goes, the richer texture features can be

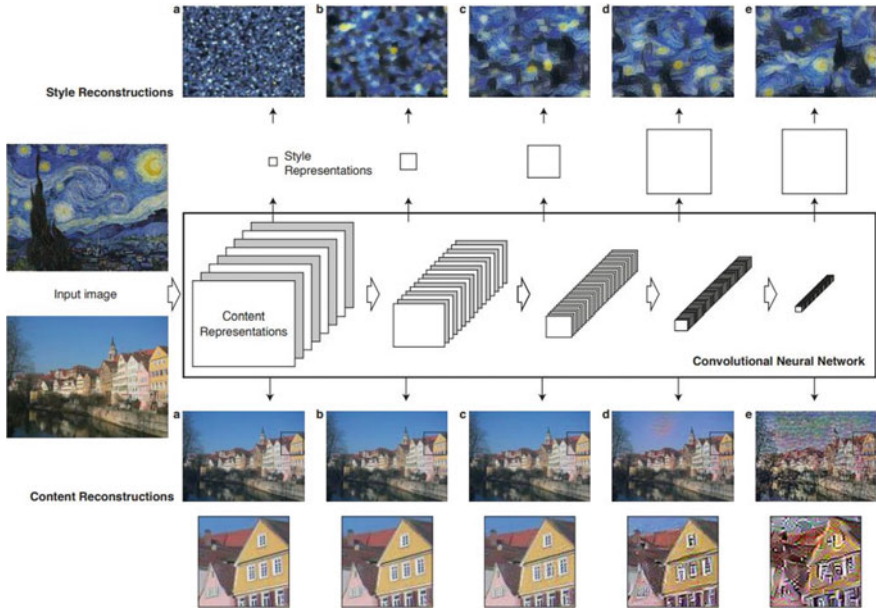


Fig. 1 Content feature reconstruction and style feature reconstruction [1]

extracted, and finally, complex feature information at the individual level can be extracted to distinguish images such as cats and dogs. In the neural style transfer algorithm, special deep features are often not needed as content features. By default, the 4–2 layer output in VGG19 is taken as the content feature value. This layer will have a better effect on image content elements and structure extraction. Style features are more complex than content features. Style is a difficult concept to explain, and it could be strokes, textures, structure, layout, use of color, etc. In the neural style transfer model, the cross-correlation between each feature map of the convolutional layer is used as the style feature of the image.

The neural style transfer model code used in this study comes from an open-source project on GitHub [5], which is an implementation of neural style transfer based on the TensorFlow deep learning library. The paper model has been expanded to a certain extent, and the default parameters can get good results, so it is more convenient to use. During the research process, the code running environment is Python 3.6, and the related libraries installed are TensorFlow, Open-CV, NumPy and SciPy. Among them, TensorFlow requires version 1.0, and NumPy requires version 1.6 and earlier. Throughout the experiment, it was not discovered that any other libraries were present with different versions. Compatibility problems.

The neural style transfer model uses the VGG19 pre-trained model, so it does not need to be trained on large-scale datasets like general deep learning models, and it is easy to use. Prepare a content image and a style image to input to the model for generation. Although there are many adjustable parameters in the model, the

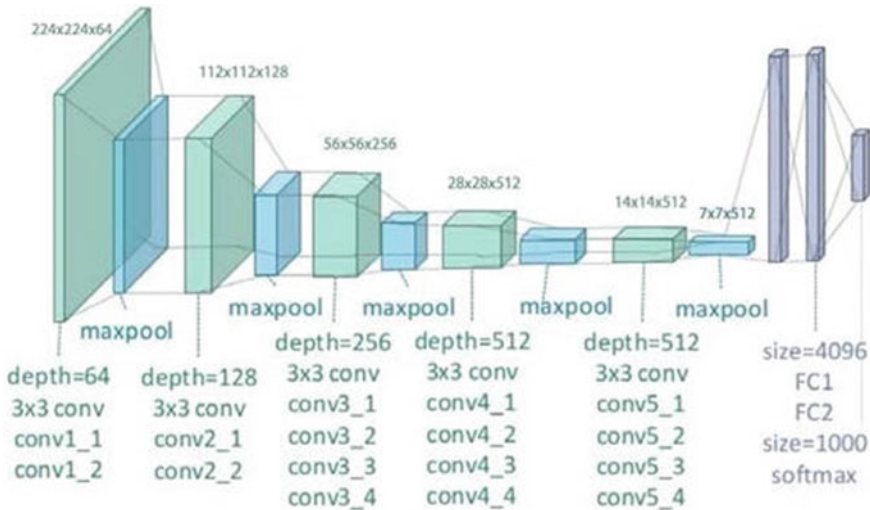


Fig. 2 VGG19 network architecture [4]

main influence on the final effect is the weight ratio of style difference and content difference. This parameter affects the ratio of content and style mixing with each other. Style the image and style will be more intense. In addition, you can increase the number of style images, so that the model has more reference styles to choose from. Since the VGG19 model can perform deep feature extraction on images at different resolutions, it is possible to choose to input images with larger resolutions on the input image.

The model transfers the target style by iteratively optimizing the style and content loss on the target image. The time spent in this process depends on the performance of the computer and the resolution of the target image. The use of GPU can help the model to be generated more quickly. Generally, the operation speed can be accelerated by dozens of times compared with the CPU, and the image resolution is not the same as the final time. In addition, if the configuration is performed on the GPU, the larger the image resolution, the more video memory is required, so the input image resolution cannot increase infinitely. In this study, GPU acceleration was used to speed up the computation, for which CUDA 8.0 and the corresponding cudnn library were installed in the environment.

3 Results and Discussion

Experiment Building Landmark Selection choose a building in the center of Qingdao, China, as our experimental site. Our main goal is to find inspiration for new aesthetic paradigms outside the architectural discipline. By using deep learning and neural

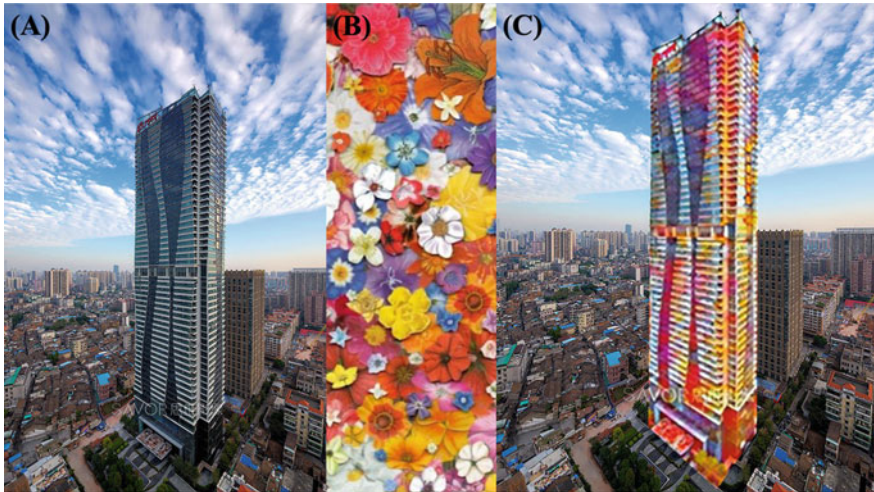


Fig. 3 Architectural feature extraction and style transfer. **a** Content picture; **b** style picture; **c** transferred picture

networks, neural style transfer techniques are applied to the content images of architectural elements of interest to stylize them with artistic images, as this process allows to preserve the architectural features of certain images. As shown in Fig. 3, for example, the system is able to recognize elements such as windows, doors, roofs, and other architectural features and is able to preserve their general morphology while conveying the input style that affects the semantic quality of the image [3]. This technical setup creates a highly accurate architectural definition, thus respecting the initial functional parameters of the building.

4 Conclusion

As with any design development process, there is currently no technical process or algorithmic solution capable of delivering a complete design from start to finish, and the goal of ML integration is not to fully automate the design process. Our technology challenges the traditional workflow of architectural design, in this new workflow, artificial intelligence proposes conceptual designs by considering design priorities determined by human designers, through neural style networks, targeting architectural elements of interest. In the case of maintaining the architectural features of the content images, they are stylized with artistic images to provide effective assistance for the facade design of modern urban high-rise buildings.

References

1. Gatys LA, Ecker AS, Bethge M (2015) A neural algorithm of artistic style. <https://doi.org/10.48550/arXiv.1508.06576>
2. Campo MD, Manninger S, Carlson A (2020) Hallucinating cities: a posthuman design method based on neural networks. In: Proceedings of the 11th annual symposium on simulation for architecture and urban design, pp 1–8
3. Özel G, Ennemoser B (2019) Interdisciplinary AI. Presented at the ACADIA 2019: ubiquity and autonomy, Austin, Texas, USA. <http://doi.org/10.52842/conf.acadia.2019.380>
4. Zheng Y, Yang C, Merkulov A (2018) Breast cancer screening using convolutional neural network and follow-up digital mammography. In: Computational imaging III. SPIE, p 1066905
5. neural-style-tf. <https://github.com/cysmith/neural-style-tf>

A Study of Students Studying Abroad to Tang in the Late Silla's Pass an Imperial Examination and Representation Using Database



Weiyu Chen, Wenxuan Zhao, Shaoshuo She, Young-Hoon An,
and Hwa-Young Jeong

Abstract The status of an international student has existed since ancient times. In particular, the number of international students reached its peak during the Silla period. At that time, many writers were dispatched to the Tang Dynasty to study literature and culture, and international students played a very important role in promoting cultural exchange and literary development between the two countries. This paper aims to look up data on students from the Silla Lower Provincial Party of Korea and China and summarized the list of Silla students in the Tang Dynasty who passed the past examination and the list of executives and employees. In the course of the study, data from both Korea and China were compared and supplemented, and a relatively comprehensive and accurate list of international students in Silla Dodang was summarized, and related information of international students was also attached in the form of a table. Through tables and data, the information of Dodang international students was made simpler and clearer and easier for others to understand.

Keywords Silla · The Tang Dynasty · Pass an imperial examination · Representation

W. Chen · W. Zhao · S. She · Y.-H. An
Department of Korean Language and Literature, Kyung Hee University, 26, Kyungheedaero,
Dongdaemun-gu, Seoul 02447, Republic of Korea
e-mail: cwy0316@naver.com

W. Zhao
e-mail: munseon981011@naver.com

S. She
e-mail: sheshaoshuo@naver.com

Y.-H. An
e-mail: yhnahn@khu.ac.kr

H.-Y. Jeong (✉)
Humanitas College, Kyung Hee University, 26, Kyungheedaero, Dongdaemun-gu, Seoul 02447,
Republic of Korea
e-mail: hyjeong@khu.ac.kr

1 Introduction

During the Tang Dynasty, the national power was strong, the economy was developed, and the culture was prosperous, so many foreign scholars went to the Tang Dynasty to study and became international students in the Tang Dynasty. Silla is no exception. At that time, a large number of international students were dispatched to the Tang Dynasty to study Chinese literature and culture. At the same time, the Tang Dynasty not only provided daily funds to foreign students so that they could study abroad well in the Tang Dynasty, but also established a “Bingonko Corporation” for foreign students to participate in past exams. Those who passed the exam, like Koreans, were awarded government posts and made contributions to the Tang Dynasty. These measures by the Tang Dynasty increased the number of foreign students in the Tang Dynasty and, on the other hand, promoted the spread of Tang literature and culture abroad.

Due to their respect for Tang culture, most of them gained high social status and reputation after returning home, and going to Tang to study abroad was even an important means for Silla writers to rise. Many international students passed past exams in Tang and were awarded government posts and were able to hold government posts after returning home. These international students brought them to Silla from Tang literature and promoted the development of Silla Chinese literature. The Tang Dynasty’s past system had a profound influence on Silla’s literary development. In the cultural exchange between Tang and Silla, Silla’s Tang students can consider an important role. Among them, many international students have excellent grades, have high literary talent, held government posts in the Tang Dynasty, held government posts even after returning home, and became respected writers. Choi Chi-won and Park In-beom created many excellent literary works in their motherland and promoted the prosperity and development of Chinese literature during the Silla period.

In this paper, we survey and provide more intuitive information on Dodang students in the Silla Dynasty by organizing and tabulating the list of international students and executives who passed the past examination in search of data from both Korea and China. However, subsequent paragraphs are indented.

2 Related Works

The main purpose of the text is to organize the list of student studying abroad to Tang during the late Silla period (i.e., during the Tang Dynasty of China). The paragraph below is as follows according to the results of the good research. Researchers Lee and Park [1] attributed the increase to the number of students studying abroad to Tang, including the government’s policy to attract foreign students, Silla’s policy to encourage them to study abroad, and Silla students’ desire to join the company. In other words, the increase in the number of Silla international students in the sense of authority in the lower Silla period was an aspect that appeared in the political and

diplomatic context of both Silla and the party. By closely examining the political background between Silla and the Tang Dynasty and the educational environment of the period, paying attention to the historical aspects, I would like to examine the background of the increase in studying abroad by Silla international students. In a similar vein, the achievements of exploring the relationship with Korean studies by paying attention to the status of students studying abroad to Tang revealed that the Silla education system was linked to the antique system.

Yoon [2] said that during the Silla Dynasty, Gyeongsang international students further raised the cultural level of Silla through a wide understanding of Korean culture, providing the basis for the implementation of the Confucian system and the creation of the imperial examination. In particular, the ideological trends of international students at that time led to the late Silla period, which led to old-fashioned ideas, Taoism, and Zen, and these intellectuals and ideas later became the ideological background of the Goryeo Dynasty. I share these opinions. In summary, it examines the political, social, diplomatic, and cultural influences of the lower Silla period and discusses how Confucian political ideology influences the early Goryeo Dynasty.

Bae [3] discusses how to manage state affairs based on the Confucian order during the Gyeon Hwon period (Later the Three Kingdoms Period, the late Silla monarch) and the great trend of the Later Three Kingdoms period. The combination of the Gyeon Hwon regime and international students, centered on Choi Seung-woo, brought about the departure and resistance of the family members with a reductive influence and military base. Lee [4] paid attention to the places of appointment of figures who were appointed to the local government as a graduate of students studying abroad to Tang. It is discussed focusing on the fact that the appointment of local officials from Dodang International Students was generally found to have been in the west coast area. Jeon [5, 6] presented about 10 municipal reform measures for international students to overcome the national crisis after returning home, centering on Choi Chi-won. In other words, after the peasant uprising in 889, the Silla government's response and the trends of Seongju and generals in the Yeongnam region were examined. In order to overcome the national crisis, Choi Chi-won proposed about 10 municipal reform measures in 894 [7]. In Kang's research [8], the activities of students studying abroad to Tang were organized by period, and the discussion was focused on examining the cause of the increase in students studying abroad to Tang in Silla against the international trend of students studying abroad to Tang decreasing in the ninth century. Cho [9] discusses the purpose of dispatching international students from the Silla Dynasty and the political and social meaning of it. Most of the students studying in the lower Silla Dynasty were government-funded students, and most of their status was nutmeg. The intention of Silla's kings to send them to the party of nutmeg status was to foster figures who could assist and support the royal authority. Kwon [7] said, "The import and dissemination of Chinese culture through the literature of the lower Silla Dynasty has a great influence on Korean history." Through literature, he suggests that he raised Sanra's idea of studying abroad to the next level. Lee [10] discussed the role played by Choi Chi-won and students studying abroad

to Tang, focusing on the significance of East Asian cultural history as well as the history of Korean culture as well as Silla.

In summary, it can be said that research on Silla students studying abroad to Tang has been accumulated since the 1950s. In a pioneering study, a detailed list of students who passed the exam among students studying abroad to Tang in Silla was revealed [11]. Since then, the background of Silla's dispatched status as students studying abroad to Tang, the situation of the times, the existence of students studying abroad to Tang and private students, and the empty space and participation were studied as detailed issues. Based on these studies, more in-depth and systematic research on students studying abroad to Tang was conducted. First, even if the middle and lower ages of Silla were separated, the change in the personality of students studying abroad to Tang was revealed, and the historical significance of students studying abroad to Tang in Silla society was discussed.

The common fact of the study that analyzed the characteristics of students studying abroad to Tang so far is the increase in students studying abroad to Tang in the lower Silla period. As for the reason for the increase in the number of students studying abroad to Tang, most studies pointed out the social confusion of the lower Silla period. In other words, as the competition for employment intensified due to political turmoil and political strife in the late Silla period, it was popular to study abroad to the Tang Dynasty, focusing on six Dufu. In a similar vein, it was suggested that studying abroad to the Tang Dynasty was a more advantageous enterprise in the lower Silla Dynasty compared to the domestic educational system of Gukhak and Reading Sampum [12, 13].

3 Analysis of Pass an Imperial Examination and Representation Charts

According to historical records, the late Silla royal family sent a number of students to the Tang Dynasty, a large number of them, including 90 in Tang, and the total number of historical records of about 38. Table 1 shows that in late Silla, many students went to study in the Tang Dynasty and did well in the imperial examinations showing Tang Dynasty's cultural prosperity and tolerance.

Some of the successful students pass an imperial examination were given official posts in the Tang Dynasty; others returned home to excel and serve in their home countries after successful examinations. At that time, if Silla students passed the imperial examination, they would receive royal commendations and awards and be awarded government posts. Therefore, with the idea of learning advanced culture and becoming a successful Confucian scholar, Silla students got a job in both countries after passing the examination. The following Table 2 gives examples of individual overseas students with historical records.

Table 1 A chart of student studying abroad to Tang in the late Silla's pass an imperial examination time and return to Silla time

Name	Pass an imperial examination time in the Tang Dynasty	Return to Silla time	Source
Kim Woon-kyung	821	841	<Dongsagangmok>
Choi Ri-jung	824-825	825	<Chaekppuwongwi>
Kim Yoon-boo	827-835	826	<Chaekppuwongwi>
Park Geo-mul	872	872	<Hwangnyongsagucheungmoktap chaljubongi>
Choi Chi-won	874	885	<Sindangsoyemunji>
Kim So-yu	Earlier than 879	888	<Dongmunson>
Kim Moo-sun	879-904	888	<Dongmunson>
Choi Won	891	897	<Dongmunson>
Kim Woon-we	894-900	906	<Samgukssagi>

Table 2 A chart of student studying abroad to Tang in the late Silla's official position in Tang and Silla

Name	Official position in Tang	Official position in Silla
Kim Woon-kyung	Sima of the Governor's Office of Kyushu	Deputy ambassador for consolation
Kim Yoon-boo	Unknown	Deputy chief of tributes
Park Geo-mul	Unknown	Servant to the Chief of the Right Army, Governor of the Province
Kim Jun	Unknown	Chusonggun's taesu
Choi Chi-won	Unknown	Busonggun's taesu
Kim So-yu	Unknown	Doctor of taehak
Choi Won	Unknown	Prosecutor's office Takanaka
Kim Woon-we	A construction worker	Purple goldfish bagging minister
Choi Sin-ji	Butler's samurai	Unknown

4 Conclusion

Through ancient international students, the role and status of international students were increasingly important through exchanges between the two countries at that time. Through Sect. 3 of this paper, most of them obtained good government positions or received high positions after returning to Korea as an international student during the Silla period. Although international students currently play a role in the exchange between the two countries, data can be found more easily in the digital era than in ancient international students. Therefore, in this paper, we read various classical sources and then make a more efficient table. According to the table, information

such as the official position of students studying abroad to Tang was more clearly known. Among the studies, we usually make a table by looking at books or papers. From this point of view, we think that information related to students studying abroad to Tang in the lower Silla Dynasty is supported and necessary by database technology.

References

1. Yan G (1955) Silla international students and monks. A collection of Chinese and Korean culture, vol 1. Korea Institute of the Chinese Academy of Sciences
2. Yan G (1961) Journal of historical language research institute, vol 4. Central Research Institute
3. Yan G (1967) Study series on Tang history. New Asia Institute
4. Yoon W (2004) Research on Gyundang international students in the unified Silla period. *Pedagogical Res* 42(4)
5. Kwon D (2005) Western studies in Silla and its historical significance. In: Silla culture, vol 26
6. Lee J (2021) Appointment of local government officials from the lower Silla period of Dodang Confucian School. In: *History of Silla*, vol 53
7. Lee S, Choi C (2016) Dodang students studying abroad. *J Korean Hist* 63
8. Cho B (2009) The dispatch and historical significance of study students in the lower Silla period, vol 25. Sogang Literature and Literature Research Center
9. Bae J, Choi S (2012) A Dodang international student—Choi Seungwoo and later Baekje's Gyeonhwon regime. In: *History of Silla*, vol 26
10. Lee J, Park H (2018) The background of the increase in the number of Silla international students and the economic conditions of studying abroad in the Tang Dynasty of Silla. In: *History education*, vol 46
11. Kang N (2018) The prevalence and background of confucianism in the Late Silla period. In: *Korean historical research on ancient Korean history*, vol 90
12. Jeon D (2021) A study on the trends of Silla government and Seongju and general before and after the King Hyogong of Silla. In: *History of Silla*, vol 51
13. Du Y (2013) A study of student studying abroad to Tang in the late Silla

Study on the Spatial Evolution of Rural Housing's Courtyard Based on the Perspective of Space Syntax



Xingkang Chai and Yu Chen

Abstract Under the background that the focus of rural construction in China has gradually shifted from increasing “quantity” to improving “quality”, this paper takes courtyard space, one of the most important spaces in the rural housing plane, as the research object, analyzing the evolution process of the syntactic characteristics of rural housing courtyards in the five typical rural construction periods in Quzhou area. Through a variety of analysis methods based on spatial syntax theory, this paper quantifies and externalizes the hidden characteristics of courtyard space and combines it with the logic of social and humanistic development. It discusses the evolution process of courtyard space, so as to provide reference for the sustainable development path and design method of courtyard space.

Keywords Rural housing · Courtyard · Space syntax

1 Introduction

The courtyard of a traditional Chinese residence evolves from the earliest stages of the local culture; the void signifies the ancient worship of heaven and earth when the population depends on agriculture, and the courtyard has been one of the most important components of the living complex since its infancy. The role of the courtyard was reinforced by the nature of wooden framework of dwellings, of which the trusses tend to expand in lineage order. Buildings are generally arranged around courtyards as a whole so that becomes Chinese ideal home beyond house matching both functional and spiritual needs.

X. Chai · Y. Chen (✉)

Architecture Department, College of Landscape and Architecture, Zhejiang A&F University, Hangzhou, Zhejiang, China

e-mail: 20080036@zafu.edu.cn

Y. Chen

Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan

But with the process of urbanization and modernization, like any other developing nations and region, the actual ritual courtyard space, rather than the gardenized ones, has been under threat due to value shifts driven by economic changes. The situation becomes worse when the users of the house gradually lose their discourse power in construction which is dominated by administrative institution and professional designers. The design gap reveals the problem of the theoretical gap, and Space Syntax is a theory and technology that is relatively advanced from an architectural perspective to bridge the gap. Its core view is that space is the junction point of material form and social and economic activity. Through the shaping of space, spatial form supports or restricts social and economic activities, and social and economic activities realize the construction of material form through its organization in space. Specifically, people achieve their own social, economic, and cultural goals by constructing spatial patterns; this spatial construction activity itself is a part of social and economic activities. Only people can perceive and use this spatial structure by walking through the space [1]. Thus, from the standpoint of system theory and overall theory, space syntax attempts to split the complex relationship between different spatial analyses and different scales, and the model of the relationship between human activities, and to intuitively reveal spatial phenomenon quantitatively under those unable to describe the social rules of logic and the space, spatial structure, and variation of self-organization model is put forward [2]. This analysis method considers the connections between spaces at the micro- and mesoscopic levels, as well as the representation of spatial patterns at the macroscopic level.

2 Cases and Methods

2.1 Methods

Space Syntax was developed by Professor Bill Hillier, who formally proposed the theory in 1974. In his book *Space is the Machine*, he formally identified Spatial Configuration as his main study focus. “According to this theory, after returning to the space itself, “relationship” is the core of all problems. Every connection in a complex spatial system depends on every other connection that is related to it.” [3] (p. i). The word “configuration” is defined as “for the two spaces, if we take the spatial relations as any form of connection between them, such as adjacency or permeability, then the two space no matter which in any way connected to the third space, will change the original relationship between them, and the configuration is in them.” [3] (pp. 23–24).

This study used gamma analysis based on syntactic calculation software Depthmap X (1.0) and Jass. In Depthmap X, Convex, Isovist, and Axial are the three fundamental space division methods that can be analyzed using space syntactic theory. This study uses the corresponding analysis path based on convex to describe the configuration characteristics of courtyard space in the object building in a more

objective, comprehensive, and similar to a natural person's environmental cognition way. The degree of aggregation or dispersion of a space and other spaces in the system is represented by the integration value.

2.2 Cases

This paper takes Quzhou as the research area mainly because of two aspects. To begin with, Quzhou has been referred to as the “head of the five roads” since the Song Dynasty because it is a valley plain at the intersection of the provinces of Zhejiang, Fujian, Jiangxi, and Anhui [4]. In a prehistoric society with primitive transportation, Quzhou served as the intersection of several significant waterways and dry roads, developing into a significant hub for economic and cultural exchange. The area is therefore abundant in cultural resources. Second, since the release of the Quzhou Urban and Rural Construction Planning and Implementation Action Plan (2016–2035) in 2016, there has been a strong push for the integration of local rural land as well as a significant number of agricultural buildings that require modernization, making the research findings more applicable.

Rural housing construction and social governance in New China are intertwined. Based on the age of the research object and the representative stage of China's rural construction within the construction period, it can be divided into five categories: before 1949, the traditional rural construction period before the founding of New China; 1949–1978, collectivization period; 1978–2005, household contract system period; 2005–2013, New rural construction period; and the new urbanization period under the b (Fig. 1). The rural collectivization movement after the founding of the People's Republic of China in 1949, the household contract system reform in 1978, new socialist rural construction in 2005, and new urbanization in 2013 are its segmentation nodes [5].

This paper examined 30 courtyard cases: 7 during the traditional rural construction period, 13 during the collectivization period, and 10 during the new urbanization period (there were no courtyard cases during the household contract system and new rural construction periods). No research was done. In order to distinguish farmhouses, rural housing is numbered, such as “QZ-DP-008,” where “QZ” is Quzhou, “DP” is the village name, and “008” is the number of rural housings in the survey.

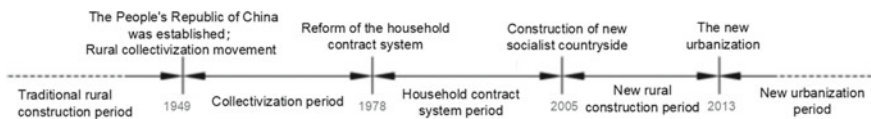


Fig. 1 The rural development stage

3 Results

3.1 Evolution of Syntactic Features of Courtyard Based on Gamma Analysis

Figure 2 displays the J-graphs of rural homes in the Quzhou region over three periods, and Table 1 displays the average Integration values during the period with courtyards. J-graphs for cases from the traditional rural construction period are shown in Fig. 2a. According to the J-graphs of this collection of rural homes, the structures have “tree-like” shapes. The typical depth of rural homes at this time was three. The courtyard, which is situated at the first fork point, is currently the shallowest (depth 1) space when compared to the functional labels of each space. Further research reveals that the rural homes at this time did not have many spaces (the average value is 6.8). Only one entrance and exit were designed to connect the indoor and outdoor areas of the buildings, which had extremely close internal and external boundaries. The trunk had short forks at depths 1 and 2, and all the cases were sequence types without rings. According to Table 1, its Integration value was significantly higher than the average Integration value of houses, with the majority of cases exceeding two times. Figure 2b displays the J-graphs for cases over the course of the collective period. This group of J-graphs, as shown in the figure, comes in two varieties: Because the tree-like structure basically inherits the traditional form, the morphological characteristics of J-graphs are extremely similar. In the other samples with rings, the building has two or more entrances and exits that are directly connected to the outside. The building’s relationship to its exterior and interior, as well as how it is used, is more adaptable, and the inner ring’s appearance enlivens the path of internal activities. Lobby, as a courtyard variant, continues to occupy the shallowest bifurcating point in the above two plane types. According to Table 1, the courtyard continues to have the highest Integration values among the tree-like rural houses, with average values that are more than twice as high as the overall average Integration values. Figure 2c depicts the J-graphs of rural housing cases during the period of new urbanization. Unlike the previous four periods, “non-ring tree-like sequence structure” has vanished in this period, and all sample J-graphs have multiple rings. The number of rural house spaces increased significantly during this period in comparison to earlier ones. A rural home’s typical depth was 3, which remained constant over time, showing that the entire plane system’s spatial organization was flatter. The existence of outer and inner rings, especially deep rings (rings connecting the deepest spaces), suggests that the connections between the inside and outside of the house, as well as the circulation of the inside of the house, are more varied and adaptable. During this time, the independent courtyard returned to the minimum living complex, but its status in the topological relationship has significantly declined. It is typically situated on a ring that is directly connected to the outside, but it does not cause bifurcations and is directly connected to the kitchen or lobby, becoming a part of the auxiliary space. According to Table 1, the Integration value of courtyards is close to the average Integration value of the entire rural house.

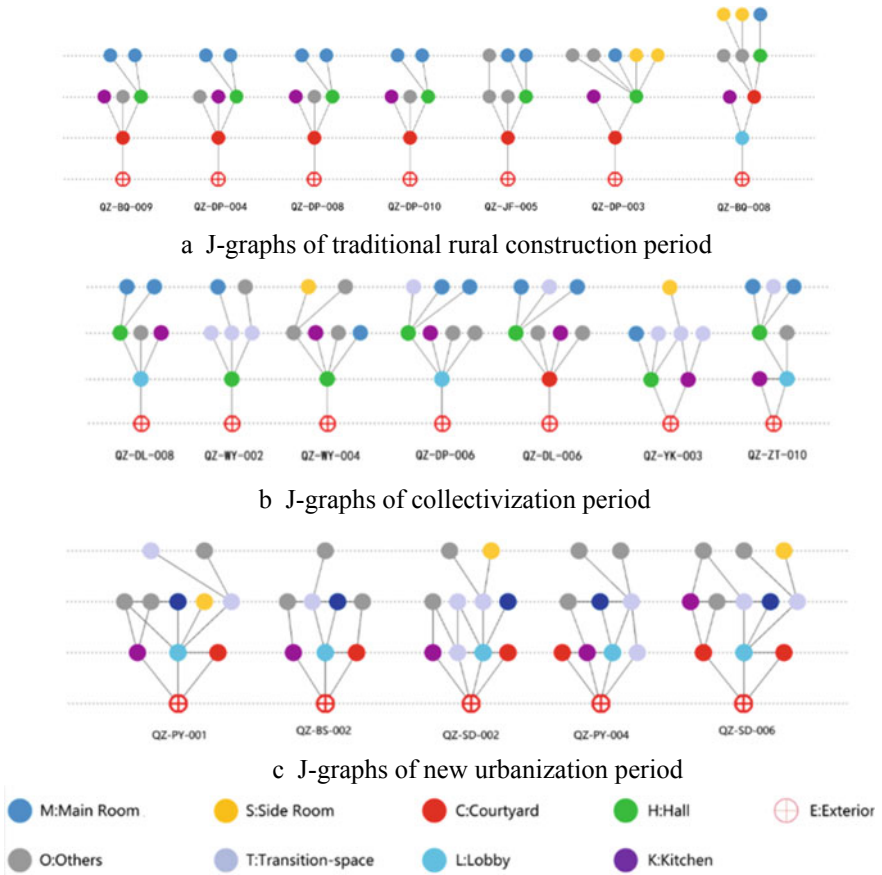


Fig. 2 J-graphs of three periods

Table 1 The average integration value of living space in Quzhou

Period	Average courtyard integration value	Average integration value
Traditional rural construction period (before 1949)	2.35	1.06
Collectivization period (1949–1978)	2.44	1.14
New urbanization period (after 2013)	0.99	1.08

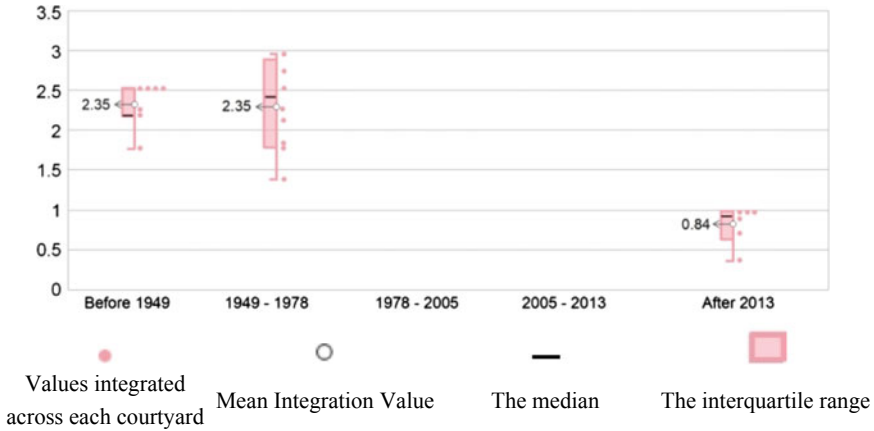


Fig. 3 Value box plot for yard space integration

By using gamma analysis, the general plane space structure of the rural house was found to have gradually changed from “non-ring tree-like sequence structure” to the trend of ringy structure. The plane function is more sophisticated, but at the same time a flattening trend is visible, the streamline and internal flow are more varied and adaptable on the inside as well as the outside, and the level of the residential interior space is weakened, exhibiting strong modern characteristics.

In this context, it is simpler to comprehend how the syntactic features of courtyard space have evolved. First, in the second period, the emergence of its variant coincides with the loss of the courtyard space’s spiritual function. The courtyard’s social function is deteriorating, as shown by the Integration value’s gradual decline (as shown in Fig. 3). The courtyard gradually becomes marginalized and separated from the main body of the house or large ring in J-graphs.

4 Conclusions

Space syntax theory revealed the following configuration of courtyard spaces in rural Quzhou homes: (1) Gamma analysis shows that the spatial structure in the rural house plane progresses from “non-ring tree-like sequence structure” to ringy structure. The plane function is refined but flattening, internal, and external flow is more diverse and flexible, residential interior space level is weakened, and the courtyard function is gradually removed. As its Integration value decreases, its social function deteriorates, and in J-graphs, it becomes marginalized and separated from the main or large ring.

The study found that rural house architecture developed through courtyard development. This paper uses gamma analysis in space syntax to reveal the invisible characteristics of rural house courtyard space and combine them with social and

humanistic development logic to inform relevant government departments. Architects need more resources and advice to guide rural construction scientifically. To achieve “external and internal cultivation” with national support for the revival of rural society and culture, rural home construction must go beyond copying and processing superficial features and deepen the social logic of the architectural space of rural homes.

References

1. Hillier B, Hanson J (1984) *The social logic of space*. Cambridge University Press, London, UK
2. Hillier B (1996) *Space is the machine: a configurational theory of architecture*. Cambridge University Press, London, UK
3. Hillier B (1999) *Space is the machine: a configurational theory of architecture*. Cambridge University Press, Cambridge
4. Quzhou Annals Compilation Committee (1994) *Quzhou annals*. Zhejiang People's Publishing House, Hangzhou, p 27
5. Lu Y (2017) *Research on the evolution sequence and design intervention mechanism of contemporary rural construction*. Tongji University, Shanghai

The Syntactic Changes of Traditional Rural House and Its Influence by Modernization



Jiacheng Xu and Yu Chen

Abstract After nearly 100 years of modernization in rural areas, the life and production mode of the new generation of rural people have changed. Using Space Syntax as a theoretical tool, this paper analyzes the syntactic characteristics of rural residential living space in Quzhou, China, in five periods to reveal the evolution of living space and its internal mechanism: (1) The main body of living space is gradually transformed from functional space to transportation space, which reflects the modern industrial civilization based on the concept of “efficiency first” and the modernization of home space. This feature is further strengthened with the in-depth involvement of external design forces in the construction of rural housing. (2) The evolution of the living space reflects the weakening of the spatial hierarchy of the rural house (gradual equivalence of functions) and the strengthening of the privacy domain of the individual or nuclear family.

Keywords Rural house · Living space · Space syntax

1 Introduction

Traditional rural house designs are crucial in preserving the regional spatial context, and rural construction must take into account how local culture will be passed down. There is no denying, however, that China’s rural areas have undergone an almost century-long modernization process and that this has altered the way of life and the mode of production for the new generation of rural residents. Understanding how farming property space has evolved to be more modern requires striking a balance between the needs of the rural and urban lifestyles and values. This is significant

J. Xu · Y. Chen (✉)

Architecture Department, College of Landscape and Architecture, Zhejiang A&F University, Hangzhou, Zhejiang, China
e-mail: 20080036@zafu.edu.cn

Y. Chen

Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan

because rural construction in recent years has gradually shifted its emphasis from “quantity” to “quality” of the background.

In terms of the farm house, the “living space” primarily refers to the hall, living room (which typically serves as a dining room), and the lobby space that is closely associated with it. The living space system serves as the “configuration core” of residential space and serves as the primary location for people to perform a variety of daily activities [1]. Using the living space system of rural houses as the research object, this paper aims to clarify (1) from a theoretical standpoint, the status evolution of living space in the rural house space system and (2) from a practical standpoint, the path of sustainable development of living space and the future design method.

2 Cases and Methods

2.1 Cases

In this paper, 17 representative villages were chosen from 14 towns (a total of 21 towns) subordinate to Qujiang District of Quzhou City in China, and 56 farm houses were chosen at random in the village for field investigation and mapping. The selection is based on *Quzhou’s 2016 Urban Rural Construction Planning and Implementation Action Plan (2016–2035)*, which aims to provide theoretical reference for future rural house design by combining with Quzhou’s current rural land integration work. Furthermore, Quzhou, as a valley plain at the cross-roads of Zhejiang, Fujian, Jiangxi, and Anhui provinces [2], was historically an important junction of water and dry roads and has been known as “the head of five roads” since the Song Dynasty. As a result, the local regional construction culture is distinct, providing a rich reference sample for the subsequent comparative analysis.

This paper divides the research objects into five categories based on the important events of the change of rural social governance in New China and the construction time: before 1949, the traditional rural construction period before the founding of New China; 1949–1978, collectivization period; 1978–2005, household contract system period; 2005–2013, new countryside construction period; and the new urbanization period against the backdrop of new urbanization from 2013 to the present [3].

2.2 Methods

Space Syntax is a mathematical method developed by architects to describe and analyze space. It uses Depthmap software to study spatial relations. The core feature of Space Syntax is that space itself is not important, but the relationship between spaces is important [1]. Depthmap X (1.0) was used in this paper to adopt genotypes

Table 1 Methods for genotype analysis

Spatial configuration features	Method	Analysis target	Analytical significance
Stability form	Integration analysis	Through sorting, refining stable type	Find and express the hidden steady-state spatial layout framework

under one spatial partitioning method based on convex space (Table 1). This paper examines the configuration change of the ground floor space of five-period rural houses in Quzhou (because the living space is all distributed on the ground floor, in order to exclude other variables, the other floors are not considered).

3 Analysis and Results

3.1 Evolution of Syntactic Features of Living Space in Stability Form Analysis

Table 2 shows stable rural house types, and Table 3 shows typical rural house plans. The stable-type living space system (hall/living room/lobby) has changed significantly over five periods: Since the collectivization period, the hall gradually replaced the courtyard to dominate the stable type of rural house, and it remained stable in the first three periods (before 2005) until 2013, when it suddenly disappeared. Since 1978, rural houses have had living room space. As a family activity space, the living room has always ranked low. The modern lobby space (collectivization period, the lobby is not in the modern sense of the lobby, but a patio with roof form temporary lobby) was born in 2005, and in the fourth period has been located in stable sequence, the importance of its syntactic meaning also gradually improve, until the fourth period to become the leading space in the space system.

In syntactic theory, integration value is an important index for determining a space’s reachability. Simply put, it can indicate whether a space is “busy” or “quiet” [4]. Combined with the integration value, the stable type’s results can be analyzed more deeply (Table 4). In the first four periods, the living room’s integration value decreased with increasing depth, while the hall’s increased with decreasing depth, indicating that the rural house’s living space system changed over time. The fifth-period hall disappearance ended this trend. What’s more, the modern lobby has also been the rural house’s most integrated space. The lobby’s integration value decreased during urbanization. The lobby’s control degree and integration value decreased due to the central ring’s appearance in rural house spatial structures and the increase in transition space. Further combined with integration average value and the period of the integration value range, the average rural house integration value during the

Table 2 Stability form of rural houses in Quzhou five periods

Period	Before 1949	1949–1978	1978–2005	2005–2013			After 2013
Stability form	Courtyard–hall	Lobby–hall Hall–	Hall–living room Hall–transition space	Hall–transition space–living room	Lobby–hall	Lobby–living room Hall–	Lobby–

Table 3 Floor plan of typical cases in different periods

Period	Floor plan	Period	Floor plan	Period	Floor plan
Before 1949		1949–1978		1978–2005	
2005–2013		After 2013			

period is basic stable, and the most integrated space in living space system values falls in the fifth period, which deduces the gap of the integrated value between the spaces less, namely in space appeared on the spatial properties of equivalence (Fig. 1).

3.2 Evolution of Transition Space Ratio

A key metric for evaluating residential privacy is the transition space ratio [4]. The quality of the transition space controls how frequently people interact with one another and establishes the purpose of the space by dividing and connecting activities [5]. Therefore, it is essential to specify and control how people and space interact. The quantity and distribution of transition space for all samples are shown in Table 4. In contrast, over the course of the five periods, both the quantity and the ratio of transition space have increased rapidly. Particularly during the new urbanization period, large-scale external design forces interfered with the construction of rural homes, and the transition space ratio of samples during this time showed another peak (Fig. 2). A tightening of the distinction between public and private spaces, along with the differentiation and improvement of spatial functions, can be inferred from the increase in the transitional space ratio. The entire house is connected by a network of connecting transitional spaces, which together form a spatial cluster resembling a “domain.” These spaces share a ring, and the independence of this ring allows us to distinguish between the “public sphere” and the “private sphere.” The increase in transition space ratio leads to the differentiation of the privacy attribute of living space.

4 Conclusions

Rural houses are referred to as “buildings without architects” due to their distinctive construction method, which enables the plane shape of rural houses to more accurately reflect the lifestyle of users. This is in contrast to the standardized residential design in cities. Analyzing the changes in the living space in Quzhou over five periods reveals two traits:

1. The general trend that the transition space has gradually taken over as the dominant space in modern rural housing is reflected in the evolution of the overall spatial structure of rural housing, including the change of overall integration value, the stable type, and the increase in the transition space ratio. The “efficiency first” philosophy, which denotes modernization at the level of the home, is the foundation of modern industrial civilization, which is what is happening in this instance [6]. However, it is important to note that due to the extensive use of outside design forces in the construction of rural homes as well as the

Table 4 Average integration value and depth of living space in Quzhou in five periods

Period	Living room		Hall		Lobby		Average integration value	Average number of transition spaces	Average transition space ratio (%)
	Average integration value	Average depth	Average integration value	Average depth	Average integration value	Average depth			
Before 1949	1.22	1.00	1.62	2.25			1.07	0	0.00
1949-1978			2.51	1.60	2.75	1.00	1.21	0.60	8.57
1978-2005	1.03	2.00	2.59	1.00			1.03	1.40	20.90
2005-2013	1.00	2.00	2.71	1.40	2.87	1.00	1.17	1.00	21.67
After 2013	0.87	2.00			1.99	1.00	1.05	2.00	26.59

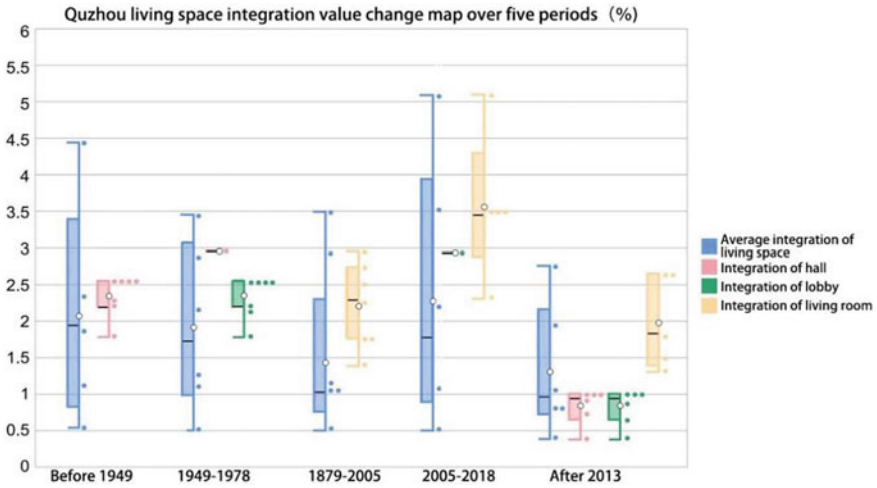


Fig. 1 Changes in overall integration value and significant spatial average integration value of typical Quzhou samples over five periods

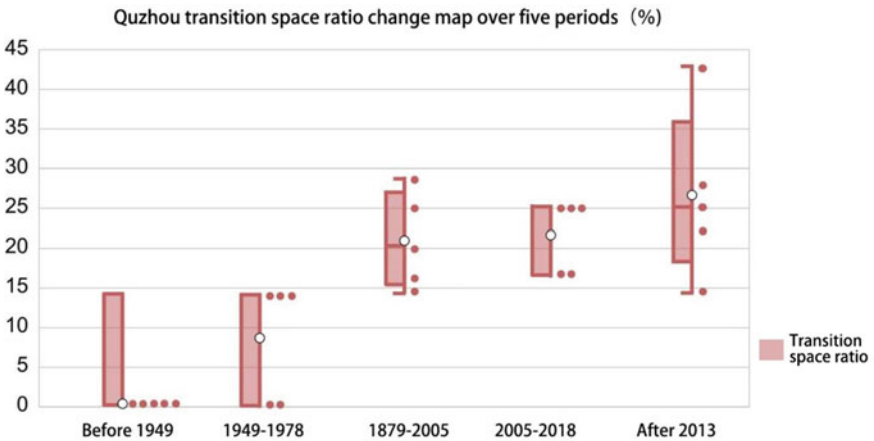


Fig. 2 Changes in transition space ratio of typical Quzhou samples over five periods

phenomenon of special mutation in the configuration of rural homes, this feature has been further strengthened (the sudden disappearance of the hall space).

2. The findings of the aforementioned spatial genotype analysis also point to a weakening of the spatial hierarchy of rural houses (gradual equivalence of functions) and a strengthening of the privacy domains of individuals or nuclear families [7]. This feature displays the user’s values and lifestyle in the spatial layer according to the current fashion.

References

1. Hillier B (1996) *Space is the machine: a configurational theory of architecture*. Cambridge University Press, London, UK
2. Quzhou Annals Compilation Committee (1994) *Quzhou annals*. Zhejiang People's Publishing House, Hangzhou, p 27
3. Lu Y (2017) *Research on the evolution sequence and design intervention mechanism of contemporary rural construction*. Tongji University, Shanghai
4. Hanson J (1998) *Decoding homes and houses*. Cambridge University Press, Cambridge, p 23
5. Lawrence RJ (1987) *Housing, dwellings and homes: design theory, research and practice*. Wiley, Hoboken
6. UGO SANTANA (2017) *Spacial configuration in single family houses. Study about the work of Marcos Acayaba*. In: 11th international space syntax symposium
7. Major MD, Sarris N (1999) *Cloak and dagger theory: manifestations of the mundane in the space of eight Peter Eisenman houses*. In: 2nd international space syntax symposium

A Quantitative Study on the Spatial Mechanism of Zhejiang Traditional Dwellings Based on Space Syntax



Keyou Xu and Yu Chen

Abstract Based on the space syntax theory, this study conducted a quantitative study on traditional residential spaces in Zhejiang. Through the use of the computer-aided program of space syntax, the RA value of each room in the building example was measured and analyzed. Thus, the spaces of traditional dwellings can be interpreted mathematically, which can greatly assist in the study of traditional architectural culture. Furthermore, quantitative analysis allows the entire research process to be tracked and verified and reduces the interference of researchers' subjective values. As a result, it will facilitate a deeper understanding of the social logic of architectural space and contribute to the sustainability of settlements.

Keywords Space syntax · Zhejiang traditional dwellings · Quantitative analysis

1 Introduction

This research is examining the space culture mechanism of traditional architecture in order to better serve the sustainability of human settlements as well as the democratic nature of settlements. By utilizing computer-aided operations based on space syntax, the subjective intentions of the government and architects can be avoided, allowing the internal laws of architectural space to be presented more objectively.

The vast majority of research into Chinese traditional architecture is based on historical and sociological models that lack the verification and analysis of mathematical quantification [1, 2]. By using space syntax, we can transform building information into mathematical models, providing a clearer research path into the internal mechanisms of buildings [3–5]. Therefore, it means that the discussion of

K. Xu · Y. Chen (✉)

Architecture Department, College of Landscape and Architecture, Zhejiang A&F University, Hangzhou, Zhejiang, China
e-mail: 20080036@zafu.edu.cn

Y. Chen

Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan

architectural space will be more scientific in nature, and the results will be based on the calculation results of objective computer programs, which will replace the subjective judgments of humans that have been used previously.

Due to the research perspective, there are various points of view regarding traditional residential buildings in Zhejiang, which are usually disturbed by subjective value judgments of architectural culture [6–8]. This study replaces the subjective value orientation of researchers with an objective quantitative analysis of space syntax. The rules are summarized based on the data of the quantitative analysis. In this way, the discussion on the mechanism of architectural space can be traced and verified throughout the research process on the basis of the data of the quantitative analysis.

2 Materials and Methods

As a source of materials, this study examined traditional dwellings in the central region of Zhejiang Province. Based on the spatial topological relationship, building information is transformed into quantifiable data using the computer-aided program of space syntax. Each space in the examples of buildings has been measured for its RA value. According to the space syntax theoretical model, the RA value is used to reflect the relationship between the selected node and the entire space system, and it also represents the level of crowd aggregation in a region from a social culture perspective [4, 5]. By assessing the RA values of several key spaces (main room, courtyard, and hall), it will be possible to gain an understanding of the type characteristics of those traditional dwellings in this area and examine their interactions in further detail.

Based on computer-assisted methods of operation (Depthmap), the building plan is analyzed for topological relationship (Fig. 1). The building information will be identified as the topology between N space nodes, and the path depth from each space to the remaining space will be calculated by the program of Depthmap. Based on the calculation, Depthmap will further count the RA value of each space according to the mathematical formula. The physical meaning of RA in architectural research is the accessibility of space nodes in the system. It is mathematically based on the Mean depth of the chosen point.

$$RA = \frac{\text{Mean depth} - 1}{\frac{N}{2} - 1}$$

In Chinese traditional residential buildings, the main room, courtyard, and hall are considered to be the most important spaces [9]. These spaces are the main target of this study, and RA values are used to analyze their internal laws. Additionally, the interaction between architecture and exterior is also an important component of the study of traditional architecture's spatial relationship [3, 10]. Exterior refers to a different spatial topological relationship and, from a practical standpoint, it indicates whether a passage exists between residential buildings and the external environment

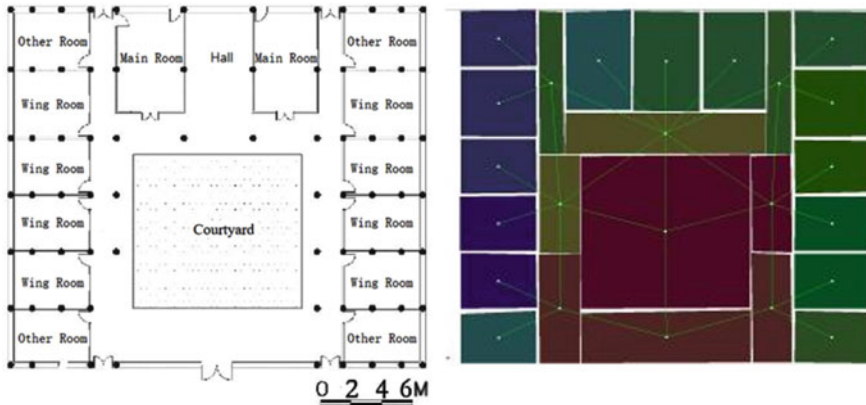


Fig. 1 Topological relationship converted from building plan in Depthmap software

[11]. It is also used as a variable when calculating the RA value of each space in this study.

3 Data Analysis

3.1 Type Characteristics Analysis

For the investigation of the type characteristics, the RA values of each space of each building example were determined. Whether or not the exterior space is taken into account, the courtyard has a lower value than any other part of the building. It is clear from the data analysis that courtyard is a significant space reflecting traditional dwellings in the region, and our field survey of the region confirms this result. Data in one instance differed slightly from expectations, but the reason for this can be explained by the contingency inherent in self-construction of traditional residential buildings (Fig. 2).

Cultural studies suggest that settlements in central Zhejiang pay more attention to cluster activities as part of their cultural activities [12]. As demonstrated by the field survey, courtyards are important places for social interaction in traditional Zhejiang dwellings. Data analysis results are consistent with those obtained from social research and field surveys. Further, it is found that exterior, as a variable, has no impact on this architectural characteristic. This means that the particularity of courtyard is less affected by the change of building state. It is a key space regardless of whether the building is connected to the external environment.

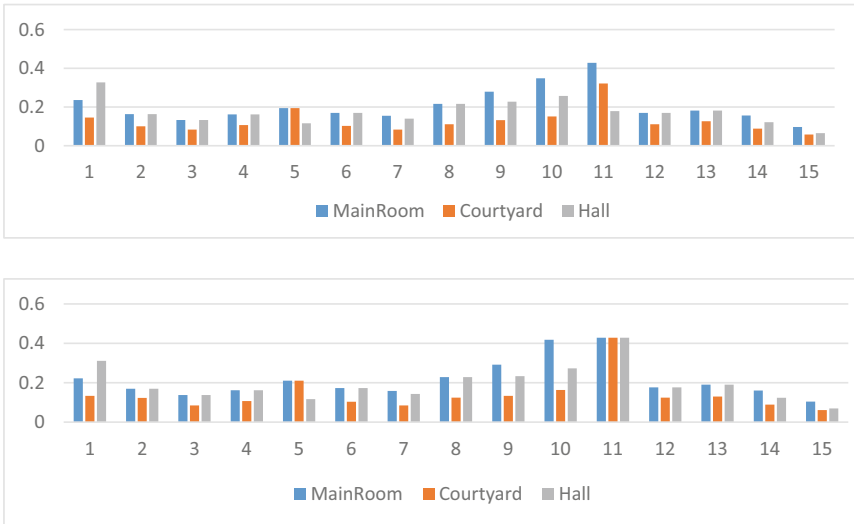


Fig. 2 RA value of various spaces (with exterior/without exterior)

3.2 Spatial Relationship Analysis

To further explore the traditional dwellings in Zhejiang, this study examines the relationships between key spaces (main room, courtyard, and hall). They are analyzed for correlation (main room-courtyard; main room-hall; hall-courtyard) and their goodness of fit is calculated based on their RA values. In addition, the exterior is still considered to be a variable (Fig. 3).

During the analysis of spatial relationship, the exterior, as a variable, has a significant impact on the result value. There is a significant increase in correlation between spaces when taking the exterior into account (better goodness of fit). A comparison of the goodness of fit of three groups of regression analysis shows that hall and courtyard, both public spaces, have a weak correlation. And among the three groups, the strongest correlation can be observed between the main room and the hall.

In combination with the practical significance, this shows that when the traditional dwellings in central Zhejiang are open to the outside world, the connection between the key spaces is strengthened. Zhejiang’s rural society places a greater value on neighborhood relations and cluster culture, which may be due to the internal social logic of this spatial characteristic.

According to the results of the data analysis, the spatial relationship between the main room and the hall is of particular interest. Although the two spaces differ from a conventional architectural perspective (the main room is relatively private, while the hall is relatively public), the data shows that they have some relevance. These studies demonstrate that the social attribute of the architectural space affects the residential attribute of the space. Traditional residences in central Zhejiang are not conventional

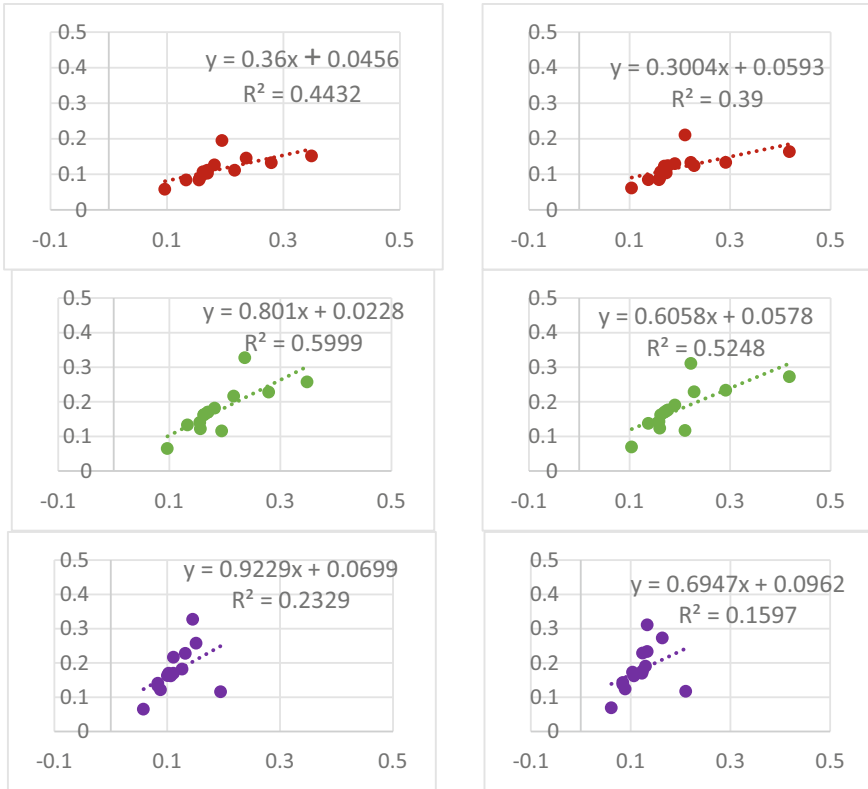


Fig. 3 Regression analysis of spatial relationships (with exterior/without exterior) (Red: Main room-courtyard; Green: Main room-hall; Purple: Hall-courtyard)

residences, but rather micro-communities. As a result, we are now able to analyze traditional dwellings from a new perspective.

4 Conclusion and Discussion

This paper examines traditional Zhejiang dwellings using a quantitative analysis of space syntax. Data analysis results are in agreement with historical data research and field observations. It provides a detailed explanation of the internal mechanisms of traditional dwellings, including their type characteristics and spatial relationships.

This study focuses on the composition of architectural space mechanism in a more precise manner than previous studies. Meanwhile, it is based on an ideological model of comparative analysis. Taking into account objective data, courtyard is considered to be a representative space. In terms of type characteristics, exterior influence is less important, but it has a greater impact on spatial relationship. And according to

the analysis of the spatial relationship, the high correlation between the main room and the hall demonstrates that Zhejiang traditional residential space places a greater emphasis on community culture.

Additionally, this study examined a new research idea in addition to the above conclusions. An effective way to verify conjectures about the spatial characteristics of related buildings is to establish a comparative analysis model based on the quantitative model of space syntax. The process is based on objective, computer-aided operations, which are traceable and correctable throughout. Compared to the traditional way of interpreting architecture based on humanistic perspectives, this will be an exciting innovation.

References

1. Chan CC, Xiong Y (2005) What forces that shape and retain the Beijing courtyard houses? In: Proceedings of the 3rd Hawaii international conference on arts and humanities, Honolulu, HI, USA, 13–16 Jan 2005, pp 794–805
2. Knapp RG (1990) *The Chinese house: craft, symbol, and the folk tradition*. Oxford University Press, Oxford, UK
3. Huang B-X, Chiou S-C, Li W-Y (2019) Study on courtyard residence and cultural sustainability: reading Chinese traditional Siheyuan through space syntax. *Sustainability* 11
4. Hillier B, Hanson J (1984) *The social logic of space*. Cambridge University Press, Cambridge, UK
5. Hanson J (1998) *Decoding homes and houses*. Cambridge University Press, Cambridge, UK
6. Qi L, Zaiyi L, Yongfa W, Dagmawi MD, Yiwei Z (2019) Cultural sustainability and vitality of Chinese vernacular architecture: a pedigree for the spatial art of traditional villages in Jiangnan region. *Sustainability* 11
7. Wei Z (2015) Home beyond the house: the meaning of home for people living in Yanxia Village, Zhejiang Province, China. University of Illinois at Urbana-Champaign, Champaign, IL, USA
8. Junqing D, Xiping Y (2009) *Zhejiang vernacular house*. China Architecture Publishing & Media Co., Ltd., Beijing, China
9. Zhuoqun Y (1992) Chinese traditional vernacular dwellings and culture. In: Yuanding L (ed) *Investigation into the subconsciousness in vernacular dwellings*. China Architecture Publishing & Media Co., Ltd., Beijing, China
10. Chen Y, Xu K, Liu P, Jiang R, Qiu J, Ding K, Fukuda H (2021) Space as sociocultural construct: reinterpreting the traditional residences in Jinqu Basin, China from the perspective of space syntax. *Sustainability* 13
11. Ding J, Ma S (2020) Comparative analysis of habitation behavioral patterns in spatial configuration of traditional houses in Anhui, Jiangsu, and Zhejiang provinces of China. *Front Architectural Res* 9(1):54–66
12. Tillman HC, Alkin MC (1982) *Utilitarian Confucianism: Ch'en Liang's challenge to Chu His*. Harvard University Asia Center, Cambridge, MA, USA

Construction and Verification of the Evaluation Scale for Architectural Design Scheme Based on Knowledge Engineering



Yingjie Liu, Lan Chen, Chuanlong Wang, Dongmiao Zhao, and Chao Liu

Abstract In order to realize the rationality and correctness of the evaluation of architectural design schemes, based on using knowledge engineering to intelligently process data and information, and to be able to conduct knowledge mining on expert scoring data, an architectural design scheme is proposed as the object. The evaluation scale of the architectural design scheme was constructed by psychometrics. The results are as follows: Exploratory factor analysis is used to prove that the evaluation dimensions of the scale are divided reasonably. The validity of the scale was verified by the Kaiser criterion and gravel test criterion. The Cronbach α coefficient was used to verify the high reliability of the scale. The feasibility and validity of the scale are verified by testing the designed scale. This paper proposes a new way with psychometrics method to evaluate the building scheme and introduces knowledge engineering in the process of scale development.

Keywords Architectural design scheme · Knowledge engineering · Scale · Reliability · Validity

1 Introduction

Before the construction of building projects, it is necessary to comprehensively consider various factors to design the scheme. With the constant change of architectural design schemes, how to evaluate their rationality and scientificity has become the key problem of current research. This paper introduces the method of psychometrics, using the scale to judge the rationality of architectural design schemes. At present, the environmental assessment of community and open space is relatively developed. The environmental scale designed by other researchers can measure the characteristics of the built environment and evaluate the reliability of built environment data [1]. In the field of architecture, scales for post-use assessment of buildings

Y. Liu · L. Chen · C. Wang · D. Zhao · C. Liu (✉)
Qingdao University of Technology, Qingdao 266520, China
e-mail: liuchao@qut.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_74

555

have been designed [2]. The scales to assess the built environment associated with walking and the built environment around the school that affects youth habits have been developed [3, 4]. However, scale research for the architectural design phase is still blank.

Because the item selection and expert scoring data of the scale need knowledge capture, storage and screening, the construction of the scale needs the support of knowledge engineering. Knowledge engineering is the development of artificial intelligence in knowledge and information processing. It uses computer systems to complete the collection, collation, construction, and application of knowledge in the form of 'rules' [5]. Knowledge engineering mainly includes knowledge acquisition, knowledge representation, knowledge verification and knowledge utilization. Knowledge acquisition is to supplement the knowledge acquired outside the system to the knowledge base studied, and process the external data [6]. Knowledge representation and knowledge verification methods can verify the accuracy of knowledge through computers [7]. The idea of knowledge engineering has been perfectly reflected in human-computer interaction design.

Therefore, this paper combines the relevant knowledge in the field of architecture and uses the principle of knowledge engineering to construct a scale for evaluating architectural design schemes, so as to facilitate and quickly complete the evaluation of specific architectural schemes by experts. And through knowledge extraction to collect and screen the data, using knowledge verification to test the reliability and validity of the scale. This method provides a convenient new way for the evaluation of architectural design schemes and provides an effective and concise measurement tool for the research of architectural design scheme evaluation.

2 Methodology

The ultimate purpose of this research is to verify whether the constructed scale is suitable and accurate for the evaluation of architectural design schemes. The scope of the research is limited to the context of architectural scheme design. More precisely, the focus of this research is that knowledge engineering methods can be used to develop the need to collect and organize relevant knowledge in the field of architectural design solutions, so as to use knowledge engineering to solve the purpose of knowledge acquisition and utilization in artificial intelligence, especially expert systems. Based on the research object and purpose of this study, the main research questions are clarified as follows:

2.1 Knowledge Acquisition and Utilization

Pretreatment and Adaptability Analysis. The first part is the pretreatment, through the data center and Pearson correlation coefficient matrix two methods to achieve, to facilitate the subsequent factor analysis of the relevant calculations; the mathematical formula for data centralization is as follows:

$$X_{ii} = X_i - \bar{X} \tag{1}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \tag{2}$$

where X_i represents the original data, X_{ii} represents the centralized data and \bar{X} represents the average of the original data. n represents the number of data.

The Pearson correlation coefficient matrix formula is used to calculate the correlation coefficient matrix of the sample data. The calculation formula is as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \tag{3}$$

where r represents the correlation coefficient, X_i and Y_i represent the horizontal and vertical coordinates of the sample points, respectively, and \bar{X} and \bar{Y} represent the average value of the sample.

The second part is the adaptability analysis, which is tested by KMO and Bartlett methods to detect whether the data is suitable for factor analysis. KMO tests the correlation and non-correlation between variables. The KMO value ranges from 0 to 1. It is generally believed that the KMO value is above 0.7. The representative factor analysis method is very suitable. When it is lower than 0.5, it indicates that the factor analysis method cannot be used. The KMO test is calculated as follows:

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2} \tag{4}$$

where KMO is an indicator of Pearson correlation coefficient and partial correlation coefficient between variables, r represents Pearson correlation coefficient and a represents partial correlation coefficient.

The Bartlett test is used to determine whether factor analysis is applicable based on significant level values. The larger the value is, the more suitable the data is for factor analysis. The Bartlett test formula is as follows:

$$-1 \times \left(n - 1 - \frac{2p + 5}{6} \right) \times \ln \det(R) \quad (5)$$

where n represents the number of items, $\det(R)$ is the determinant of the correlation matrix and p is the number of variables.

2.2 Knowledge Verification

Validity test. The scale of this study has the characteristics of weak correlation of evaluation items and rich practical meanings represented by items and dimensions, so factor analysis is more suitable for scale validity verification. KMO test and Bartlett test were used to verify the validity of the data, so exploratory factor analysis can be used to verify the validity of this scale. First, explore the number of evaluation dimensions through the Kaiser criterion method and the gravel inspection criterion; then, the rationality of the number of evaluation dimensions is clarified by means of variance accumulation and variance explanation rate. The above can prove the eligibility of the scale setting and complete the validity test.

Reliability test. The reliability of the scale was tested by internal consistency coefficient and verified by Cronbach α coefficient. The Cronbach α coefficient is used to test the item's consistency of the same content in the scale. The formula is as follows:

$$\alpha = \frac{K}{K - 1} \left(1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_X^2} \right) \quad (6)$$

where K is the total number of items in the scale; the intra-test variance of the score of question i ; the variance of the total score for all items.

3 Results and Discussions

3.1 Pretreatment and Adaptability Analysis

Bartlett test results. The results of KMO and Bartlett values (Table 1) show that based on the data of the scale pre-test, the Bartlett test statistic is 1327.25, and its corresponding significance (P value) is 0.00. Therefore, it can be considered that the correlation coefficient matrix and the unit matrix are significantly different, suitable for factor analysis. After testing, the KMO and Bartlett values of the sample data meet the requirements of factor analysis. The variables are correlated and not independent, and the factor analysis method can be applied.

Table 1 KMO and Bartlett test

KMO value	0.72	
Bartlett test	Statistic	1327.25
	Significance (<i>P</i> value)	1.06×10^{-105}

KMO test results. According to the data collected from the pre-test of the scale, with the help of the KMO test formula, the KMO value of the pre-test is 0.72, which indicates that the correlation between factors is strong, and the factor analysis method is more applicable.

3.2 Analysis of Validity Test Results

Comprehensive use of Kaiser criterion and gravel test criteria to determine the number of scale evaluation dimension. The results show that in the gravel diagram (Fig. 1), the number of factors is 2, 8, 11 three points from steep to flat, all in line with the gravel test criteria. The eigenvalues corresponding to the three points are 3.28, 0.91 and 0.41, respectively. According to the eigenvalue criterion, the points with the eigenvalue of 3.28 and 0.41 do not conform to the criterion, and the number of evaluation dimensions is 8. Therefore, this scale sets up 8 evaluation dimensions.

The variance interpretation rate can indicate that the selected evaluation dimension represents the degree of evaluation items. In this section, the variance interpretation

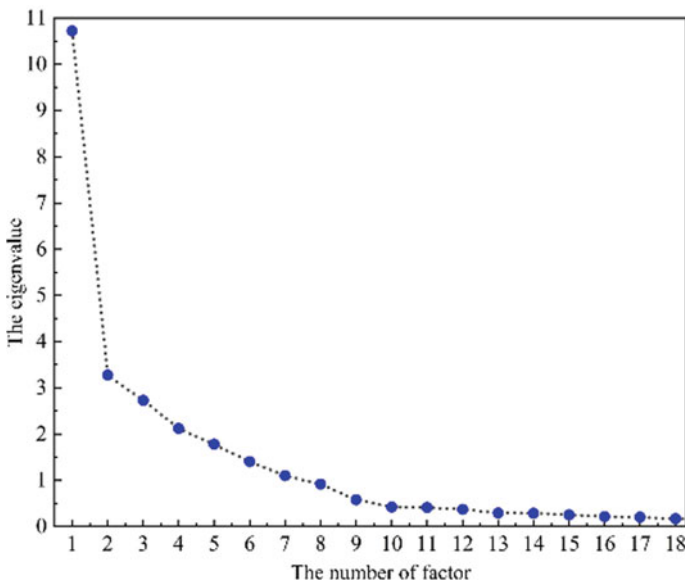


Fig. 1 Gravel diagram

rate is calculated by the maximum rotation method (varimax) and the factor load matrix. The results of variance interpretation rate are that the variance interpretation rates of the first 18 factors are 0.12, 0.12, 0.10, 0.10, 0.10, 0.09, 0.09, 0.08, 0.02, 0.02, 0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, and the results are greater than or equal to 0.01. After the number of factors in the initial eigenvalue exceeds 18, the corresponding variance interpretation rate is less than 0.005, so it is not considered. The cumulative variance interpretation rate of the first 18 factors increased by 12%, 24%, 34%, 44%, 54%, 63%, 72%, 80%, 82%, 84%, 86%, 87%, 88%, 89%, 90%, 91%, 92% and 93%. The variance interpretation rates after rotation were 0.12, 0.12, 0.11, 0.09, 0.09, 0.09, 0.09, 0.09, and the cumulative values increased by 12%, 24%, 35%, 44%, 53%, 62%, 71%, 80%. In the process of factor analysis, the main concern is the data part after rotation, and when the number of factors after rotation is 8, the cumulative variance interpretation rate is 80%, which shows that the eight dimension factors can represent 80% of the scale information.

3.3 Analysis of Reliability Test Results

The reliability verification results of the scale showed that the Cronbach α coefficient was 0.923, the confidence interval was array ([0.868, 0.961]), the coefficient value was greater than 0.9 and the width of the confidence interval was less than 0.1. The scale had good reliability, strong internal consistency and high reliability.

4 Conclusions

This paper expounds on the design method of evaluation scale based on knowledge engineering through the case of building evaluation scheme. Knowledge engineering is applied to the construction of psychometric scales, which integrates relevant knowledge in this field into the design and collects expert scoring data. Knowledge engineering can reduce errors. Maximize the accuracy of the design through this human-computer interaction design method. The scales can be used to collect expert knowledge widely, and can guarantee the objectivity, scientificity, and comprehensiveness of expert knowledge. The research results are helpful to the relevant research on the mesoscale test of architectural design schemes and provide data support and solid reference for further establishment of evaluation models.

References

1. Frank LD, Saelens BE, Chapman J, Sallis JF, Kerr J, Glanz K et al (2012) Objective assessment of obesogenic environments in youth. *Am J Prev Med* 42(5):e47–e55

2. Fox EH, Chapman JE, Moland AM, Alfonsin NE, Frank LD, Sallis JF et al (2021) International evaluation of the microscale audit of pedestrian streetscapes (MAPS) global instrument: comparative assessment between local and remote online observers. *Int J Behav Nutr Phys* 18(1)
3. Deng C, Dong X, Wang H, Lin W, Wen H, Frazier J et al (2020) A data-driven framework for walkability measurement with open data: a case study of triple cities, New York. *Isprs Int J Geo-Inf* 9(1):36
4. Solbraa AK, Anderssen SA, Holme IM, Kolle E, Hansen BH, Ashe MC (2018) The built environment correlates of objectively measured physical activity in Norwegian adults: a cross-sectional study. *J Sport Health Sci* 7(1):19–26
5. Chen Y, Chen Y, Wang C, Chu H, Tsai T (2005) Developing a multi-layer reference design retrieval technology for knowledge management in engineering design. *Expert Syst Appl* 29(4):839–866
6. Quintana-Amate S, Bermell-Garcia P, Tiwari A (2015) Transforming expertise into knowledge-based engineering tools: a survey of knowledge sourcing in the context of engineering design. *Knowl-Based Syst* 84:89–97
7. Chandrasegaran SK, Ramani K, Sriram RD, Horváth I, Bernard A, Harik RF et al (2013) The evolution, challenges, and future of knowledge representation in product design systems. *Comput Aided Des* 45(2):204–228

Development of Indoor Air Quality Management Using Automatic Control of Home Appliances



Efforts to Keep CO₂ Concentrations Below Standard Values

Yuko Kuma  and Fumiya Sakai

Abstract We developed a system to control CO₂ concentration in a room by substituting CO₂ for air pollution and controlling a circulator using a CO₂ sensor and a microcontroller. The system is compact and can be installed anywhere. The system is highly scalable by changing the combination of sensors and home appliances and is compact enough to be easily installed in a building. An occupancy experiment was conducted, and it was confirmed that an environment in which CO₂ concentration did not exceed the standard value could be maintained.

Keywords Internet of things · Indoor air quality · Automatic control of home appliances · CO₂

1 Introduction

In the recent years, automatic control of home appliances has become common place, providing a comfortable living environment. It has become easy to link sensors, microcontrollers and home appliances. We have investigated a system that detects air pollution by COVID-19 and automatically exhausts the polluted air. We propose a system to control CO₂ concentration in a room by substituting CO₂ for air pollution and controlling a circulator using a CO₂ sensor and a microcontroller.

Y. Kuma (✉)

Shonan Institute of Technology, Fujisawa, Kanagawa 2518511, Japan

e-mail: kuma@info.shonan-it.ac.jp

F. Sakai

Alpha Systems Inc., Kawasaki, Kanagawa 2110053, Japan

2 System Overview

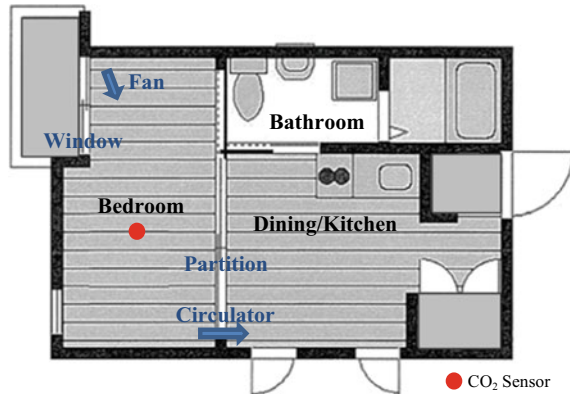
An overview of the system as follows; we get CO₂ concentration from CO₂ concentration sensor (Sensirion SCD30) to microcontroller (M5Stack). The measurement interval is 5 min. The measured data is sent and recorded in a data sheet (Google spreadsheet), it is also used as the basis for circulator operation. The circulator is controlled using Google Apps Script (GAS) and a smart remote controller (Nature Remo3). The upper and lower limits of CO₂ concentration to run the circulator can be set arbitrarily. The circulator starts running when the measured value exceeds upper limit and stops the value falls lower limit. Automatic control is a feature of this system.

3 Experiment

3.1 Outline of Experiment

This experiment was conducted using student residences. Figure 1 shows a plan of the residence. It has a bedroom, dining kitchen and bathroom. Total floor area is 21.7 m² and ceiling height is 2.4 m. While the CO₂ concentration in the bedroom (approx. 18 m²: 1970 mm * 3910 mm * 2370 mm) exceeds 1000 ppm [1, 2], the circulator is activated and air sent to the dining and kitchen side.

Fig. 1 Plan of the residence



3.2 *CO₂ Concentration Measurement Before System Introduction*

We measured changes in CO₂ concentration over time generated from the start of bedtime, assuming situations that could occur in daily life in summer season. One male in his 20s lives in this room. The room is ventilated 0.5 times/h all day. Table 1 gives the conditions of the experimental environment. Measurements are taken 8 conditions, which are a combination of 3 items: window, fan and partition (Fig. 1), those details as follows:

Case 1: The window closed, the fan turned off and the partition closed.

Case 2: The window closed, the fan turned off and the partition opened. Compared to Case 1, the air volume in the dining kitchen has increased.

Case 3: The window closed, the fan turned on and the partition closed. Compared to Case 1, the air in the bedroom will be agitated.

Case 4: The window closed, the fan turned on and the partition opened. Compared to Case 1, the air volume in the dining kitchen has increased and the air in the bedroom will be agitated.

Case 5: The window opened, the fan turned off and the partition closed. Compared to Case 1, the bedroom window is open to let outside air in.

Case 6: The window opened, the fan turned off and the partition opened. Compared to Case 1, the bedroom window is open to let outside air in, and the air volume in the dining kitchen has increased.

Case 7: The window opened, the fan turned on and the partition closed. Compared to Case 1, the bedroom window is open to let outside air in and the air will be agitated.

Case 8: The window opened, the fan turned on and the partition opened. Compared to Case 1, the bedroom window is open to let outside air in and the air will be agitated, and the air volume in the dining kitchen has increased.

Figure 2 shows the change overtime in CO₂ concentration in each case. These experiments were performed by the same subjects on different days. In all cases,

Table 1 Conditions of the experimental environment

	Window	Fan	Partition
Case 1	Close	OFF	Close
Case 2	Close	OFF	Open
Case 3	Close	ON	Close
Case 4	Close	ON	Open
Case 5	Open	OFF	Close
Case 6	Open	OFF	Open
Case 7	Open	ON	Close
Case 8	Open	ON	Open

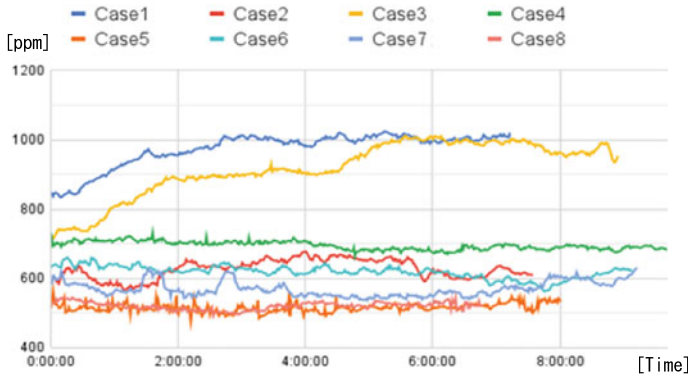


Fig. 2 Change over time in CO₂ concentration

changes over time were observed from the beginning of sleep. Of the eight cases, Cases 1 and 3 were found to have elevated CO₂ concentrations, exceeding 1000 ppm. The similarity between Cases 1 and 3 is that both windows and partitions are closed. In the other 6 cases, there was almost no change from the CO₂ concentration at the start of bedtime.

3.3 CO₂ Concentration Measurement After System Introduction

In this measurement, the system was installed in two cases (Cases 1 and 3) with high CO₂ concentrations, and the effectiveness of the system in reducing CO₂ concentrations was confirmed. Figures 3 and 4 show the change overtime in CO₂ concentration in each case. Cases 1 and 3 were measured again with and without the system. In both cases, we confirmed when the CO₂ concentration in the bedroom exceeds 1000 ppm, the circulator is activated and the air sent to the dining kitchen. The circulator then stops when the CO₂ concentration falls below 800 ppm.

4 Conclusion

We proposed an easy-to-install system to improve air quality in living rooms. The system consists of a CO₂ sensor, a microcontroller and a circulator. Before the system was installed, measurements showed that the CO₂ concentration exceeded 1000 ppm when the windows and partitions were closed. However, after the system was installed, it was confirmed that the indoor CO₂ concentration could be kept lower than 1000 ppm. The system can easily change its operating conditions depending on

Case1_System IN

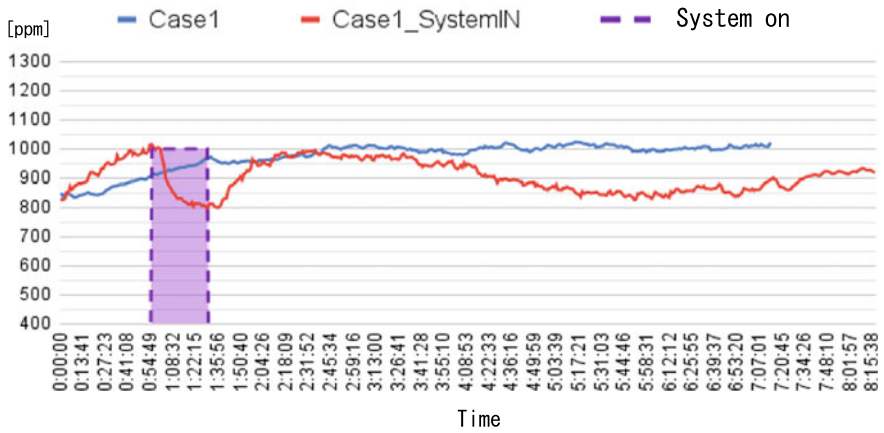


Fig. 3 Change over time in CO₂ concentration (case 1)

Case3_System IN

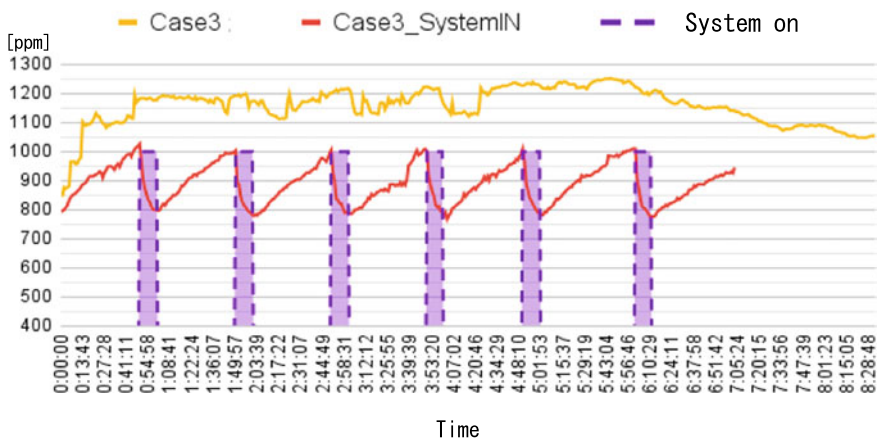


Fig. 4 Change over time in CO₂ concentration (case 3)

the size of the room and the type of contaminant, making it suitable for a variety of situations. The system is highly scalable.

References

1. Ministry of Health, Labour and Welfare Website. <https://www.mhlw.go.jp/content/11130500/000771215.pdf>. Last accessed 2021/10/10
2. Ministry of Health, Labour and Welfare Website. <https://www.mhlw.go.jp/bunya/kenkou/seikatsu-eisei10/>. Last accessed 2021/10/10

A Model for Detection of the Susceptible in Hospital Contact Network Based on Social Network Analysis



Yixuan Yang , Doo-Soon Park , Seok-Hoon Kim , Sony Peng ,
Sophort Siet , Sadriddinov Ilkhomjon Rovshan Ugli ,
and Phonexay Vilakone 

Abstract COVID-19 can be modeled as a SEIR disease transmission model, and early detection and prevention of susceptible is a great way to stop the spread of the virus. The purpose of this paper is to conduct community detection on the hospital's contact network (including patients, nurses, doctors, and managers), obtain frequently contacted communities (susceptible communities), and analyze the independent contact networks of different roles to obtain high-influence communities and rank the node influence, if nodes with high influence are included in the susceptible community, we will recommend focusing on observing these nodes to prevent the spread of virus. In this paper, we introduce a model that combines social networks analysis with disease transmission model to analyze hospital contact networks, apply the classic overlapping community detection method CPM to community detection of contact networks, and utilize PageRank algorithm to calculate the node's influence. The experimental results show that the model can effectively solve our problem.

Keywords Community detection · Contact network · Key nodes · k-cliques community

Y. Yang · S. Peng · S. Siet

Department of Software Convergence, Soonchunhyang University, Asan, South Korea

D.-S. Park (✉) · S.-H. Kim · S. I. R. Ugli

Department of Computer Software Engineering, Soonchunhyang University, Asan, South Korea
e-mail: parkds@sch.ac.kr

S.-H. Kim

e-mail: seokhoon@sch.ac.kr

P. Vilakone

Department of Computer Engineering and Information Technology, National University of Laos, Vientiane, Laos

1 Introduction

In the recent years, with the outbreak and epidemic of coronavirus disease 2019 (COVID-19), early detection of infectious diseases with COVID-19 becomes a research hotspot in computational epidemiology and complex network science. The outbreak of COVID-19 has the characteristics of unpredictability, rapid spread, wide infection range, and difficulty to control. It caused a huge loss of life and property in human society. Usually, models of infectious diseases include considering the overall behavior of the population, or interpersonal relationships in the contact network. In our work, we intend to apply the social network analysis on contact networks and combine the contact network with the epidemic to discuss how to build a mode to detect the susceptible, and it is a meaningful and useful way to reduce the damage of COVID-19.

COVID-19 is a pandemic respiratory illness spreading from person-to-person caused by a novel coronavirus and poses a serious public health risk, it is available to be modeled as an SEIR model (which is a type of the disease transmission model) [1]. SEIR model contains susceptible (S), exposed (E), infectious-asymptomatic (IA), and recovered (R). Based on the existing study, the susceptible state transfers to the latent state with the infection probability per unit time β , and the exposed state transfers to the infection state with the infection probability per unit time γ [2]. Therefore, designing effective methods for the detection of susceptible is an effective means of preventing and controlling COVID-19. Figure 1 is an explanation of SEIR in contact network.

In this work, we focus on designing a mode of detecting some communities in the hospital contact network and mining some key nodes in these communities, and we utilize the social network analysis method in this work.

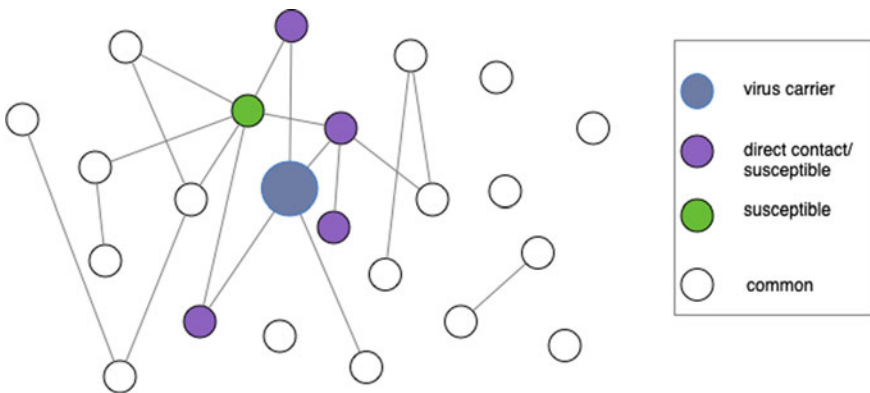


Fig. 1 Simulation of SEIR model in a contact network

2 Related Work

Social network analysis is an analytical method for the detailed study of human communication, interaction, and other complex behavioral patterns [3]. It is used to study individuals/organizations and their connection structures. Social network data is usually presented in the form of a social graph (shown in Fig. 1), where nodes represent individuals and edges represent relationships between individuals, where relationships may be personal friend relationships, contact records with other individuals, and so on. At present, there is a lot of research work on social networks and diseases. Ref. [4] applied social network analysis to the informative interactions of doctors, nurses, health professionals, and clerks in the social network of the Australian Metropolitan Teaching Hospital. This research suggests that social network analysis can be used to examine complex medication advice-seeking interactions among hospital ward staff, providing useful quantitative baseline data for comparing the impact of interventions on interactions. Ref. [5] applied a modified SEIR compartmental mathematical model for prediction of COVID-19 epidemic dynamics incorporating pathogen in the environment and interventions. The results of the study show that quarantine of contacts and isolation of cases is useful for halting the spread of COVID-19.

3 Model for Detection of the Susceptible in Hospital Contact Network

3.1 Problem Statement

Usually, there are several roles are included in a hospital contact network, such as: doctors, nurses, managers, and patients. The patient will carry the virus of some infectious diseases (COVID-19), so some nurses, doctors, and other staff are possible to become virus carriers (exposed, susceptible) in the process of multiple contacts (or prolonged contact) with the patient, and will carry the virus to more staff who don't contact with patients directly (other nurses, doctors, or managers). Therefore, the purpose of this paper is to design a model that can detect staff who have high contact frequency with patients, long contact time, and more contact with other staff in their respective office areas. For these kinds of staff, conduct stricter epidemic prevention tests on them, to reduce the possibility of virus transmission.

3.2 Detection of the Susceptible in Hospital Contact Network

To understand our method well, at the beginning of this subsection, we will briefly introduce some basics of social network analysis methods that we will apply in our model.

Cliques Percolation Method (CPM) [6]. The CPM is used to discover overlapping communities, and a clique is a collection of nodes where any two nodes are connected, i.e., a complete subgraph. The nodes in the community are closely connected, the edge density is high, and it is easy to form a clique. Therefore, edges within communities are more likely to form large complete subgraphs, while edges between communities are almost impossible to form large complete subgraphs, so that communities can be discovered by finding cliques in the network. A k -clique represents a complete subgraph with k nodes in the network. If a k -clique overlaps with another k -cliques by $k - 1$ nodes, we will regard the two k -cliques as connect to each other. The set of all connected k -cliques is a k -clique community.

PageRank [7]. PageRank algorithm, also known as web page ranking algorithm, is a technology that is calculated by search engines according to the mutual hyperlinks between web pages (nodes) to reflect the relevance and importance of web pages (nodes). The main idea is: (1) if a web page is linked to by many other web pages, it means that this web page is more important, that is, its PageRank value will be relatively high. (2) if a page with a high PageRank value links to other pages, the PageRank value of the linked page will increase accordingly.

In a social network, we replace the concept of a “web page” with a “node” and calculate the importance of that node in the network. The formula is as follows:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)} \quad (1)$$

Among them, u, v represent the nodes in the graph, B_u represents the node connected to the node u , and L_v represents the number of neighbor nodes of the node v .

Our Method

In our method, in order to detect the community and the high-influence nodes inside, we use CPM method to mining overlapping community, then we apply the PageRank method on four different roles contact network (nurse, patient, doctor, and manager), respectively, get the top- N node list, then to find the nodes which not only belong to top- N node list but also community.

Figure 2 explains our goal in detail. At first, we will detect the communities in the contact network; then we will rank the nodes in sub-contact networks with different roles (nurse, doctor, patient, and manager); finally, we will search the nodes (from top- N node lists) in the communities which contains patients.

The procedure of our method is described as the following pseudocode:

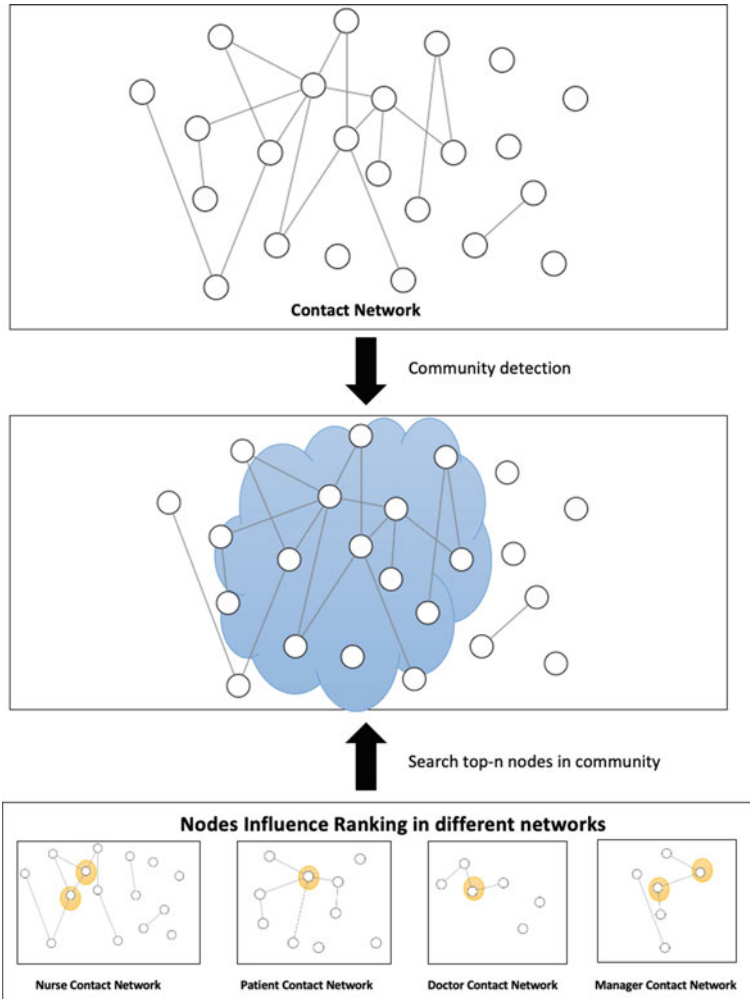


Fig. 2 Overview of our model

Algorithm

- 1 Input total contact graph $G = (V, E)$, nurse contact graph $G_n = (V, E)$, patient contact graph $G_p = (V, E)$, doctor contact graph $G_d = (V, E)$, manger contact graph $G_m = (V, E)$, k, n, RS
- 2 Initial Community Set C , Top-n node lists of contact network $AllNodeList$, nurse network $NurseList$, patient network $PatientList$, doctor network $DoctorList$ and manger network $ManagerList$

```

3 C ← CPM (G, k)
4 NurseList, PatientList, DoctorList, ManagerList, AllNodeList ←
  PageRank(Gn, n), PageRank(Gp, n), PageRank(Gd, n), PageRank(Gm, n),
  PageRank(G, n)
5 for each c in C:
6   for each node in c:
7     if node ∈ NurseList || Patient || DoctorList || ManagerList:
8       RS ← node
9 Return RS

```

4 Experiment and Results Analysis

4.1 Dataset

This dataset is a contacts network between patients, patients, and healthcare workers (nurse, doctor, and manager) and among healthcare workers in a hospital ward within 4 days in Lyon, France [5]. The study included 46 healthcare workers and 29 patients. In our work, we only focus on the first day data and take every 1 min as a count unit. In our experiment data, there are 18 patients and 36 healthcare workers.

4.2 Results and Analysis

The results of k -cliques community with different k are given as Table 1. It is observed the highest value of EQ will be obtained when $k = 6$. So in our work, we will select the 6-cliques community for our study.

The node set of 6-cliques community is given as Table 2, and the top-8 node lists in different networks are given in Table 3.

According to the results in Tables 2 and 3, we can get a group with the highest probability of spreading the disease, which contains 5 patients, 1 nurse, and 1 doctor, and the nurse and doctor are top 1 in their respective role networks. Hence, No. 33

Table 1 Results of k -cliques community

	Number of community	Number of nodes	Value of EQ (Modularity)
$K = 4$	2	42, 5	0.0235
$K = 5$	1	27	0.048
$K = 6$	1	21	0.0593

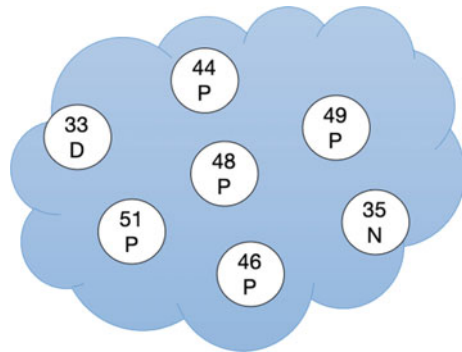
Table 2 Node list of 6-cliques community

Nodes of 6-cliques community	
$K = 6$	{6, 33, 44, 51, 2, 10, 4, 14, 12, 19, 31, 18, 11, 49, 47, 1, 53, 13, 0, 27, 5, 16, 46, 25, 48, 35, 21}

Table 3 Top-8 node lists

	Top-8 node list
All node	{51, 44, 49, 35, 52, 33, 48, 46}
Doctor	{33, 28, 21, 17, 15, 14, 13, 11}
Nurse	{35, 31, 34, 27, 25, 32, 22, 23}
Patient	{51, 49, 44, 52, 38, 47, 36}

Fig. 3 Group of the highest probability of spreading COVID-19



and No. 35 should be paid more attention and strictly controlled. The detail shows in Fig. 3.

5 Conclusion

Overall, since COVID-19 can be modeled as a SEIR model, an effective way to stop the spread of the virus is to detect and prevent susceptible individuals as early as possible. Based on this premise, this paper uses CPM community detection algorithm to detect the contact network of hospitals to obtain susceptible communities and analyzes the independent contact network of different roles by PageRank to obtain high-influence communities and rank the influence of nodes. In the experimental part, an experiment is conducted on a real hospital contact data and the most suitable K -cliques community k parameter is found. The results show that the model can effectively solve our problem.

Acknowledgements This research was supported by the National Research Foundation of Korea (No. NRF2022R1A2C1005921) and BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048).

References

1. Yang Y, Hao F, Park DS et al (2021) Modelling prevention and control strategies for COVID-19 propagation with patient contact networks. *Hum Cent Comput Inf Sci* 11:45
2. Fan RG, Wang YB, Luo M et al (2020) SEIR-based novel pneumonia transmission model and inflection point prediction analysis. *J Univ Electron Sci Technol China*. <https://doi.org/10.12178/1001-0548.2020029>
3. Yang Y, Peng S, Park DS et al (2022) A novel community detection method of social networks for the well-being of urban public spaces. *Land* 11(5):716
4. Creswick N, Westbrook JI (2010) Social network analysis of medication advice-seeking interactions among staff in an Australian hospital. *Int J Med Inform* 79(6):e116–e125
5. Vanhems P, Barrat A, Cattuto C et al (2013) Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One* 8(9):e73970
6. Palla G, Barabási AL, Vicsek T (2007) Quantifying social group evolution. *Nature* 446(7136):664–667
7. Wang R, Zhang W, Deng H et al (2013) Discover community leader in social network with PageRank. In: *International conference in swarm intelligence*. Springer, Berlin, pp 154–162

A Modern Recommendation System Survey in the Big Data Era



Sony Peng , Doo-Soon Park , Dae-Young Kim , Yixuan Yang , Sophort Siet , Sadriddinov Ilkhomjon Rovshan Ugli , and HyeJung Lee 

Abstract Recommendation Systems are a prominent field that is widely used and popular for assisting people in making appropriate automated selections. It is a mechanism that assists the users in identifying relevant information from a range of available data. In this paper, we create a survey of state-of-the-art recommendation models and focuses primarily on providing a concise summary of the many strategies for recommendation, ranging from traditional to modern approaches, to open the door for further investigation into recommendation systems. In addition, the specific domain of the recommendation models was also illustrated in this study.

Keywords Recommendation systems · State-of-the-art · Modern approaches

1 Introduction

We live in the era of flood information, with data significantly increasing from various sites. With the massive data information, recommendation systems are the well-known technique to filter the knowledge data (e.g., news, movies, songs, point-of-interests, books, journals, products, etc.) information and then return suitable lists to the user. From the traditional data mining approach to the deep learning model, the model of the recommendation is a promising technique at the root of state-of-art technology [1]. In the realm of eCommerce, users have several options in finding items of interest, where recommendation systems assist people in locating and analyzing the preferred items. In the Healthcare area, recommendations have been used to suggest

S. Peng · D.-S. Park (✉) · D.-Y. Kim · Y. Yang · S. Siet · S. I. R. Ugli
Department of Software Convergence, Soonchunhyang University, Asan, South Korea
e-mail: parkds@sch.ac.kr

D.-Y. Kim
e-mail: dyoung.kim@sch.ac.kr

H. Lee
Institute for Artificial Intelligence and Software, Soonchunhyang University, Asan, South Korea

suitable items (e.g., food, drug, physical activity, healthcare professional, hospital place). Moreover, location-based social networks (LBSNs), including Foursquare, Facebook, Instagram, LinkedIn, and Yelp are the centric sites where users can share their physical location, user tag, comment/feedback, like/dislike, rating, real-time check-ins, and user friend connections. In addition, users are closely connected and certainly know about their real-time locations.

The existing studies have created recommendations in various domains by applying simple recommendation techniques to complex models. At first, the recommendation models are categorized into two methods. These two models are collaborative filtering (CF), and content-based filtering (CBF). According to the demand for state-of-art applications, more models are introduced including knowledge-based, demographic-based, and hybrid method, etc.

Due to the popularity of the recommendation model, researchers are exploring the recommendation model to enhance its performance and the model's complexity.

To the best knowledge of our study, Sect. 2 introduces the detail of modern recommendation approaches. Section 3 provides simple explanations about recommendation-specific domains. Finally, Sect. 4 is described conclusion of the work.

2 Modern Recommendation Systems in the Big Data Era

2.1 How Do We Filter the Related Journals?

Within the scope of this investigation, we spot the journals from various well-known sites. Google Scholar [2] is used as a primary search engine, and in addition to this, we have also embraced Web of Science [3], which we refer to as an essential tool to discover related published papers.

Moreover, several keywords are used to filter the journal papers. These keywords include recommendation system, collaborative filtering, content-based filtering, hybrid model, traditional recommendation system, modern recommendation system, product recommendation, movie recommendation, product recommendation, news recommendation, job recommendation, location recommendation, point-of-interest recommendation, healthcare recommendation system, and recommendation challenges.

2.2 Overview of Recommendation Systems

With the massive amount of data, the recommendation system aims to predict the close items to the user. Recommendation systems collect information on their users' preferences for items (e.g., books, journals, songs, movies, products, location, travel,

etc.), which then return the personalized items to the active/target users. Data is a necessary part of the model of the recommendation. Refer to Fig. 1, there are two types of the data, namely explicit data, and implicit data. Explicit data refer to the data that users are actively provide to the system through their interaction such as like/dislike, rating and especially it's directed collection. This type of data is powerful and beneficial to the recommendation model since the data is given the clear purpose of the user's mind in terms of analysis. The disadvantage of this data is difficult to obtain. Another type is Implicit data which is collected indirectly indicates preference and taste. The data is mainly stored in the session system like browsing history, landing page, purchase history, and even mouse movement.

Figures 2 and 3 are shown the recommendation model from collaborative filtering, content-based filtering, knowledge-based, demographic-based, and finally, the hybrid model. More precisely, the recommendation system's general process is displayed in Fig. 2.

Fig. 1 An overview of the basic the data

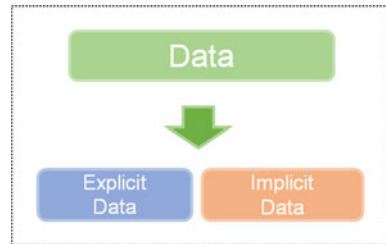


Fig. 2 Basic recommendation system model

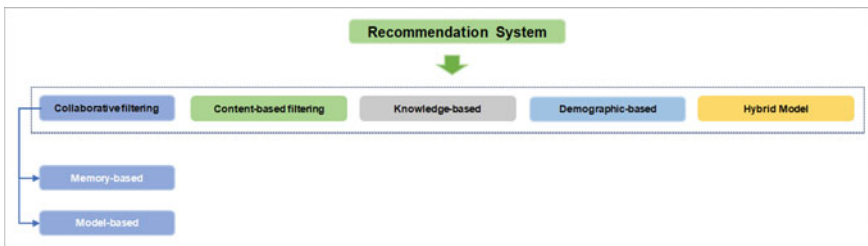


Fig. 3 Recommendation system types

2.3 Modern Recommendation Systems

Various recommendation models have been deployed in the study area's wide range. However, there are two main categories in the recommendation systems, including collaborative filtering and content-based filtering. In our current work, we are not focusing on only two models but aim to provide more studies with other models with the interaction of the big data.

Collaborative Filtering. Collaborative filtering is the most successful technique in the recommendation system and is widely used in the recommendation area. This technique uses the behavior of the group of users to suggest to other users [4]. Simply, the recommendation is based on nearest neighbor or similar users. There are two mainly types in the collaborative filtering including user-based and item-based.

User-based: In the mechanism, the system will suggest the items based on user's favorite and find the most similar item to the one user like [5]. Several similar metrics can be utilized to calculate the similarity between users/items, such as cosine similarity, adjusted cosine similarity, Pearson correlation coefficient (PCC).

Item-based: In collaborative filtering, item-based predicts user rating for the item similarity. For this model, their performance is better than the user-based since it suffers from several issues, such as sparsity and scalability [6]. Moreover, both user- and item-based suffer from cold-start problems (Item cold-start and User cold-start).

Content-based Filtering. The fundamental principle underlying content-based recommenders is to recommend products based on their content items' similarity [7]. By examining the descriptions of those items, the algorithms discover and separate the primary common attributes of a particular user's personal favorites. These choices are then saved in this user's profile. The system then recommends more similar things to the user's profile.

Furthermore, content-based recommendation systems can capture the user's individual preferences and propose products that would be of little interest to other people. However, because item feature representations are built manually to some extent, this technique necessitates a great deal of domain knowledge. Furthermore, content-based recommendation systems may only propose based on the user's existing interests, and in addition, it limits their capacity to broaden users' existing interests.

Knowledge-based model. The primary purpose of the knowledge-based recommendation model is to suggest to the users according to basic knowledge of the users, items, and the relationship between items or users [8]. More importantly, this model has no cold-start problem because the model doesn't depend on the user/item rating or purchase history. In addition, knowledge-based is particularly useful for complex domains where items are not frequently purchased [9]. However, this recommendation mechanism can run into trouble if people don't have the proper domain knowledge [10].

Demographic-based model. Demographic-based is the recommendation system type where the user's demographic data is utilized as an alternative input for making suggestions [11]. The data can be user age, gender, location, etc.

Hybrid model. Hybrid model is the method that merge multiple recommendation models to over-come the issue in the existing recommendation models.

3 Specific Domain of Recommendation Systems

Recommendation systems are utilized in various application perspectives. It is a state-of-art application where people use techniques because it assists the user to save time to filter the valuable item/information to reduce their finding times. Recommendation systems have been applied in many eCommerce companies. For instance, amazon deployed the product recommendation in their web sites to suggest the most suitable item to their target customers. Existing works [12–15] designed a recommendation model to suggest the products to the target customers.

Other applications in movie domains proposed by Ref. [16–19] created a model to recommend movie entertainment based on various machine learning mechanisms. Not only in the range of entertainment or eCommerce, but recommendation systems are also widely used in healthcare. Ozsoy et al. [20] proposed a recommendation model to identify new targets for known drugs. This model aims to over-come the problem related to traditional drug discovery by using the existing drugs to treat newly discover diseases. Ref. [21] deploys the model of recommendation to predict a patient's health condition by analyzing the patient's lifestyle, physical health record, and social activities.

4 Conclusion

The simple survey focuses on modern recommendation systems in the big data era. In the study, we provide a simple explanation and survey of current recommendation systems such as collaborative, content-based, knowledge-based, demographic-based, and hybrid approaches. Furthermore, we illustrate the several issues which existed in the traditional recommendation models. Due to the demand of the big data era, recommendation models are built and researched more critically to adapt to real-world application and to find valuable knowledge filtering to figure out the pattern in the massive amount of data (big data).

Acknowledgements This research was supported by the National Research Foundation of Korea (No. NRF-2022R1A2C1005921) and BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048).

References

1. Sahoo AK, Chakraverty S (2022) Machine intelligence in dynamical systems: a state-of-art review. *Wiley Interdisc Rev Data Min Knowl Discov* 12(4):e1461
2. Google Scholar Homepage. <https://scholar.google.com>. Last accessed 2022/10/31
3. Web of Science Homepage. <https://www.webofscience.com>. Last accessed 2022/10/31
4. Koren Y, Rendle S, Bell R (2022) Advances in collaborative filtering. In: *Recommender systems handbook*
5. Koochi H, Kiani K (2016) User-based collaborative filtering using fuzzy C-means. *Measurement* 91:134–139
6. Zhang Z, Zhang Y, Ren Y (2020) Employing neighborhood reduction for alleviating sparsity and cold start problems in user-based collaborative filtering. *Inf Retrieval J* 23(4):449–472
7. Mathew P, Kuriakose B, Hegde V (2016) Book recommendation system through content based and collaborative filtering method. In: *2016 international conference on data mining and advanced computing (SAPIENCE)*. IEEE, pp 47–52
8. Aggarwal CC (2016) Knowledge-based recommender systems. In: *Recommender systems*. Springer, Cham, pp 167–197
9. Tarus JK, Niu Z, Mustafa G (2018) Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artif Intell Rev* 50(1):21–48
10. Dong M, Zeng X, Koehl L, Zhang J (2020) An interactive knowledge-based recommender system for fashion product design in the big data environment. *Inf Sci* 540:469–488
11. Safoury L, Salah A (2013) Exploiting user demographic attributes for solving cold-start problem in recommender system. *Lect Notes Softw Eng* 1(3):303–307
12. Cheng T (2019) Product recommendation system design. In: *Proceedings of the 2019 2nd international conference on information management and management sciences*, pp 71–74
13. Ge X, Zhang Y, Qian Y, Yuan H (2017) Effects of product characteristics on the bundling strategy implemented by recommendation systems. In: *2017 international conference on service systems and service management*. IEEE, pp 1–6
14. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Comput* 7(1):76–80
15. Chiu MC, Huang JH, Gupta S, Akman G (2021) Developing a personalized recommendation system in a smart product service system based on unsupervised learning model. *Comput Ind* 128:103421
16. Vilakone P, Park DS, Xinchang K, Hao F (2018) An efficient movie recommendation algorithm based on improved k-clique. *HCIS* 8(1):1–15
17. Cui BB (2017) Design and implementation of movie recommendation system based on Knn collaborative filtering algorithm. In: *ITM web of conferences*, vol 12. EDP Sciences, p 04008
18. Ahuja R, Solanki A, Nayyar A (2019) Movie recommender system using K-means clustering and K-nearest neighbor. In: *2019 9th international conference on cloud computing, data science and engineering (confluence)*. IEEE, pp 263–268
19. Zhang J, Wang Y, Yuan Z, Jin Q (2019) Personalized real-time movie recommendation system: practical prototype and evaluation. *Tsinghua Sci Technol* 25(2):180–191
20. Ozsoy MG, Özyer T, Polat F, Alhajj R (2018) Realizing drug repositioning by adapting a recommendation system to handle the process. *BMC Bioinf* 19(1):1–14
21. Sahoo AK, Mallik S, Pradhan C, Mishra BSP, Barik RK, Das H (2019) Intelligence-based health recommendation system using big data analytics. In: *Big data analytics for intelligent healthcare management*, pp 227–246

Privacy-Aware Intelligent Healthcare Services with Federated Learning Architecture and Reinforcement Learning Agent



Prohim Tam, Inseok Song, Seungwoo Kang, and Seokhoon Kim

Abstract The convergence of digital health systems, Internet of Healthcare Things (IoHT), and Deep Learning (DL) offers intelligent healthcare services to deploy in medical software development with extended decision-making modules. However, the communication reliability and information privacy of data-driven model integration remain a challenging topic to further discuss by researchers and standard organizations. From a real-time communication perspective, the standard of Quality-of-Service requirements in smart healthcare is within real-time services, which permits extremely low packet delay budget and error loss rate. To jointly generate a delay-aware, reliability-aware, and privacy-aware approaches in intelligent healthcare services (e.g. DL-based medical software), there are three key essential aspects, namely distributed edge learning, resource virtualization, and model slicing orchestration. This paper presents optimized Federated Learning (FL) architecture based on autonomous agent policies (resource placement and slicing orchestration) using reinforcement learning. The system architecture outlines large-scale IoHT deployment in wireless cellular networks, network functions virtualization-enabled edge computing placement, and converged FL components. The communication and computation models tackle the utilization efficiencies of bandwidth, energy, and computing capacities. This proposed optimization approach strengthens privacy and reliability by orchestrating through different weights of slicing prioritization in multi-intelligent healthcare services such as non-real-time/real-time modelling, telemedicine, and remote operation (e.g. telesurgery).

Keywords Federated learning · Intelligent healthcare · Internet of healthcare things · Network functions virtualization · Reinforcement learning

P. Tam · I. Song · S. Kang · S. Kim (✉)
Department of Software Convergence, Soonchunhyang University, Asan, Korea
e-mail: seokhoon@sch.ac.kr

S. Kim
Department of Computer Software Engineering, Soonchunhyang University, Asan, Korea

1 Introduction

Deep Learning (DL)-based systems in healthcare have been deployed for several purposes, namely medical image detection, segmentation, classification, enhancement, and analysis [1]. The advancement in (1) resource-constrained segmentation with deep multi-scale convolutional neural networks, (2) disease diagnosis/classification (e.g. emphysema), and (3) feature learning with modified autoencoders has become active topics in state-of-the-art DL-based healthcare development. However, from communication perspectives, the real-time DL-based modules integrate within healthcare systems for time-sensitive operation by autonomously constructing the communication links between Internet of Healthcare Things (IoHT) devices and (centralized) DL engines. Furthermore, by using raw data sharing methods, the communication resources are deficient due to numerous transmission rounds, and the personalization information is leaked into the centralized model server. Therefore, model training, inference, and adaptation in real-time intelligent healthcare services require applying collaborative learning framework which considers privacy-aware conditions. The primary requirements that necessitate adhering are the end-to-end Quality-of-Service (QoS) and privacy of local data following the General Data Protection Regulation (GDPR).

Based on standardization releases, QoS requirements for IoHT applications concern delay, jitter, bandwidth, and error rate. The logical slicing concepts have become an interesting domain to orchestrate the resource efficiently and prioritize applications with mission-critical QoS restrictions. Figure 1 presents the decoupling multi-service control of logical concepts with sliced services compared to physical heterogeneous infrastructure. In DL-based healthcare services, the control systems require guaranteeing different priority tiers including the training time towards the final learning model. Table 1 presents the QoS requirements of two sample scenarios from smart healthcare applications, including QoS Class Identifier (QCI), resource type, priority, Packet Delay Budget (PDB), and packet error loss rate [2, 3]. Remote operation (e.g. telesurgery) can converge with DL in terms of motion scaling, performance analysis, or even AR predictive displays when overall delays exceed the upper-bound limitation. Telemedicine has been integrated with DL for advancing telemonitoring, diagnosis, and disease management. Since DL-based approaches require extra computation times, the control mechanism needs to ensure its applicability in the real-world deployment.

This paper is organized as follows. Section 2 presents the related works of (1) applying Federated Learning (FL) framework for collaborative DL-based modelling in intelligent healthcare services and (2) Reinforcement Learning (RL) for optimizing FL control policies. Section 3 describes the system architecture of the proposed approach (AutoFL-HIS), including the working flows, algorithm designs, and healthcare confidentiality concepts. Section 4 presents the simulation. Section 5 concludes the paper.

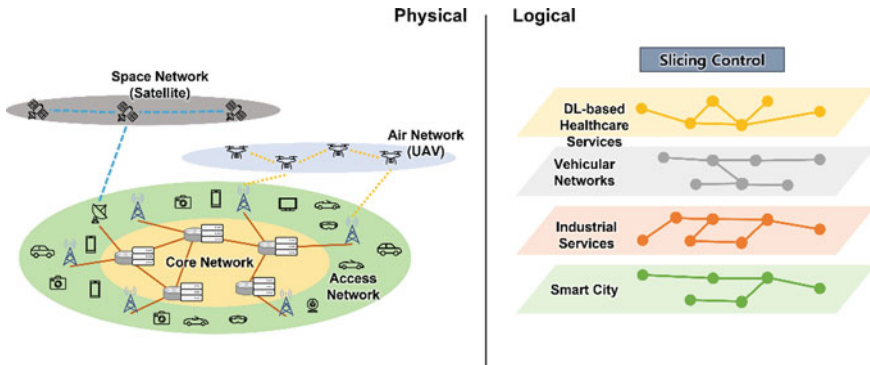


Fig. 1 Multi-service control of logical concepts compared to physical heterogeneous networks

Table 1 QoS requirements on IoHT example scenarios

Example scenarios	QCI	Resource	Priority	PDB (ms)	Packet error loss rate
Remote operation	69	Non-GBR	5.5	60	10^{-6}
Telemedicine	2	GBR	6.8	150	10^{-3}

2 Related Works

The traditional raw data sharing can be modified into model communications by following the working flow of distributed edge learning to accelerate the convergence speed and reduce the fronthaul congestion. FL framework enables collaborative (edge) learning by offering the interaction of model interchangeability between a global parameter server, edge aggregators, and local participants [4, 5]. The working flow of FL requires several control policies of participant selection, offloading decisions, update scheduling, resource orchestration, and aggregation rules. To optimize the control policies, RL is considered as a solving approach if the environment set-up is well-modelled following the RL problems and well-adapted to Markov decision process (MDP) components. The observability of state spaces, configurability of action spaces, and computability of reward evaluation in FL healthcare environment have to be fully set up for the RL agent engine [6].

Data exposure capabilities from multiple healthcare organizations in digital services can be converged with privacy awareness if the local model is computed for transferring within a local micro data centre. Several layers are required in the platform, including data holders with confidentiality layers, communication infrastructure, and FL manager [7, 8]. Data holder with confidentiality layer aims to respond by (1) processing its local data, (2) training the model according to the current streaming data batch and distributed global model hyperparameters, (3) optimizing the loss values compared to its experienced losses, and (4) uploading the loss-optimized

models for aggregation. For communication infrastructure, the controller and orchestrator require to configure the control policies. An extension of NFV-enabled edge aggregators for healthcare FL frameworks will be described in Sects. 3 and 4. FL manager is responsible for (1) executing the averaging process in every client-service model round communications, (2) aggregating with priority awareness of different services, and (3) performing the federated data analysis.

Healthcare remote systems in large-scale IoHT deployment can be within cellular networks [9], which leverage the enabling usage scenarios of enhanced mobile broadband, massive machine-type communications, and ultra-reliable and low-latency communications. Therefore, the system models for FL communication and computation are enhanced with higher uplink data rates and virtualized edge-assisted framework.

3 Optimized FL for Privacy-Aware Intelligent Healthcare

3.1 System Architecture

In this section, the optimized FL architecture based on RL for privacy-aware intelligent healthcare services, termed AutoFL-HIS, is presented. The system architecture of integrated FL components is discussed with NFV-enabled virtualization in cellular healthcare networks. Figure 2 presents the interactivity between (1) IoHT monitoring devices with on-device computing capabilities, (2) IoHT-oriented FL for privacy-aware modelling that corresponds to the model communications in local edge aggregators and central parameter for final learning, and (3) multi-service of IoHT applications.

Central parameter server refers to FL manager, which is denoted as P , equipped with a macro base station M . NFV-enabled edge aggregators referred to the central computation servers, which virtually isolated the resource capacities in NFV infrastructure, and denoted as $E = \{1, 2, \dots, e\}$. Each virtual edge aggregator e is equipped with a small base station, denoted as $S = \{1, 2, \dots, s\}$. In this context, if the control mechanism selects any particular base station to offload, the computation destination will be linked to the primary NFV entities with its defined resource block. Data holders as participants denoted by $U = \{1, 2, \dots, u\}$, where $v_{N(e)}^t$ is the selected subset of the participant to offload to aggregator e at t -iteration. Model parameter size from participant u in t -iteration with data batch d is denoted as $m_{u(d)}^t$, which subjects to offload towards e if $u \in v_{N(e)}^t$. The required resources for total aggregation following the schedule timeslot- t_{agg}^e in edge aggregator e indicates the decision-making of agent control. Therefore, the communication model is defined with the uplink data rate of model size $m_{u(d)}^t$ offloading to further specify the total latency transmission. The computation model of virtual edge node e considers the evaluation of allocated resources to fully compute all the edge-participants models $m_{u(d)}^t$ from the selected subset $v_{N(e)}^t$. In terms of energy consumption, each edge

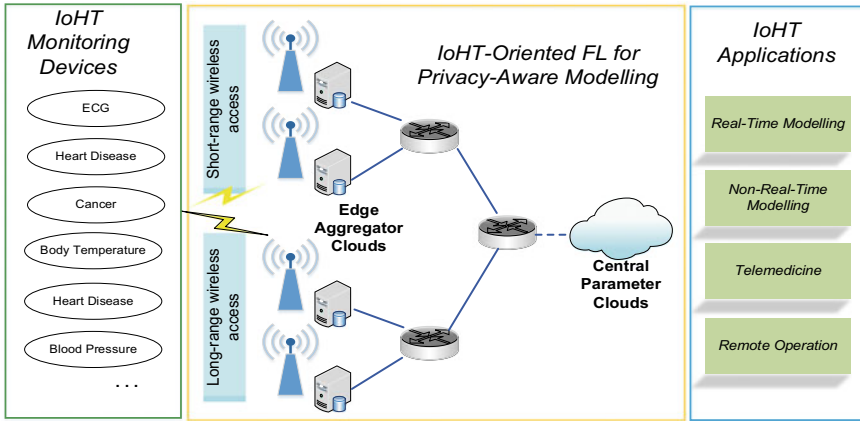


Fig. 2 Interactions between IoHT monitoring devices (FL participants) to multi-service IoHT applications in cellular edge-assisted networks

node considers the activeness of aggregation and frequency of model incoming per timeslot, which is relative to the model size aggregation. The optimization process jointly tackles minimizing energy utilization and placing the virtual resources in different update scheduling and congestion levels. In this study, the optimization procedures fully follow the concepts of RL and MDP components. The application destinations in this study only consider 4 primary slicing services of collaborative DL-based healthcare systems, such as real-time modelling (e.g. expected low-latency disease diagnosis), non-real-time modelling (e.g. disease classification for long-term noticing purposes), telemedicine, and remote operation.

3.2 Components for Optimization Procedures

The primary components to model the environment and trigger the observation parameters in this study focus on (1) the FL states in local and edge attributes, (2) actions in NFV orchestration, and (3) joint resource/energy/QoS reward efficiencies. The state components consider the essential characteristics of the model communications including model parameter sizes $m_{u(d)}^t$, virtual block remaining resource in node e at iteration- t , denoted as r_e^t , and other allocated communication variables (e.g. allocated bandwidth from selected subset IoHT to the offloading edge aggregators). The actions in NFV orchestration primarily tackle (1) the determination of the IoHT subset for the participant in any particular schedule timeslot- t_{agg}^e , which is jointly represented by $v_{N(e)}^t$, and (2) predetermination of bandwidth and computing resource capacities, which are jointly represented by $r_{alloc(e)}^t$.

In terms of reward functions, three objectives are integrated including the optimization of resources, energy, and QoS. The optimization in this concept refers to

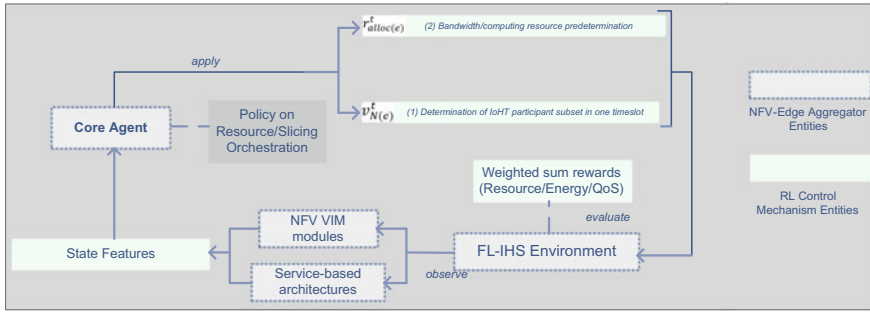


Fig. 3 Interaction interfaces and processing flows of AutoFL-HIS approach

the expected Q -value formulation from RL agent, which indicates the best action to take for maximizing the long-term rewards. The formulation of each model follows value approximation methods and bellman equations. Figure 3 presents the overview of system interactions between NFV-edge aggregator and agent control mechanism entities in FL-HIS environments.

4 Simulation Environment and Result Discussions

The experiment is conducted using Mini-NFV framework following the purposes of virtual network function manager deployment and network services on Mininet/RYU SDN controller testbed. The virtual resources are configured in 5 different capacities in the instantiated descriptors to represent the performance of edge aggregators. IoHT is configured in Mininet as hosts with constrained capacities compared to normal device modules. The model offloading is represented by normal task size uploads. The controller/orchestrator configures the offloading paths, resource placement, and forwarding paths following the RL agent outputs. The agent triggers the communication outputs from Iperf3 testing. The optimization process of RL agent is simulated using OpenAI-gym with prior knowledge from Iperf3 delays. Resource and energy efficiencies are calculated in RL agent separately. The merging process of triple-joint weighted sum models is formulated following the overall concepts described in Sect. 3.1. Figure 4 illustrates the total rewards (weighted sum) between the proposed scheme and two reference schemes (Greedy algorithms on resource and energy), respectively. The reward scoring valuation is set by accumulative negative and position scores if the QoS requirements are exceeded or not. The QoS requirements followed the indicators presented in Table 1. The optimal score of each immediate reward is 0. The negative scores occur mostly in the exploration phase. Overall, the agent control mechanism returned a reliable long-term expected reward which is efficient for autonomous control in FL-HIS systems.

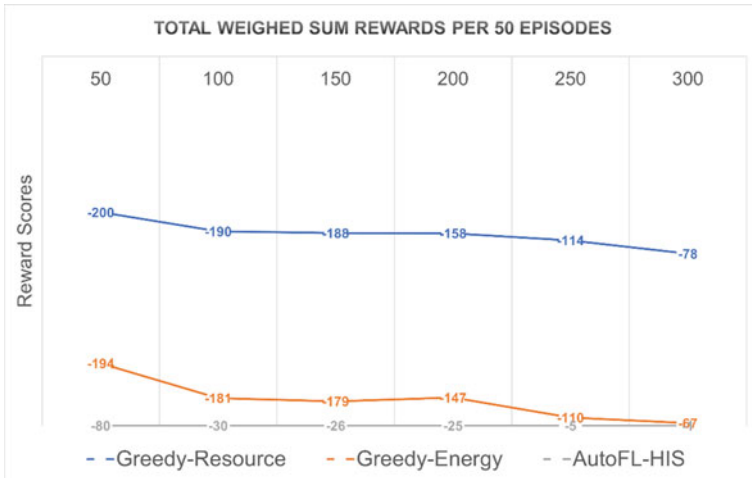


Fig. 4 Total weighted sum reward comparison between the proposed scheme, AutoFL-HIS, and reference schemes (greedy resource and energy)

5 Conclusion

In this paper, the set-up components, system architectures, working flows, and algorithm designs are contributed to point out the optimized FL architecture based on RL for privacy-aware intelligent healthcare services, termed as AutoFL-HIS. Multi-service intelligent healthcare applications were simulated in the set-up FL-based environment to illustrate the reward efficiencies of each resource/energy slicing management and QoS expectations. In future study, experimental simulation on AutoFL-HIS architecture will be extended by capturing key performance indicators of accuracies, losses, algorithm delays, resource efficiency, and reliability metrics using a real-world data set of the 2019 annual data (Version 2.0.1) of the Korea Medical Panel Survey jointly organized by the Korea Institute for Health and Social Affairs and the National Health Insurance Service.

Acknowledgements This work was supported by BK21 FOUR (Fostering Outstanding Universities for Research) under Grant 5199990914048.

References

1. Chen Y-W, Jain LC (2020) Deep learning in healthcare. In: Paradigms and applications. Springer, Heidelberg
2. Wijethilaka S, Liyanage M (2021) Survey on network slicing for internet of things realization in 5G networks. *IEEE Commun Surv Tutor* 23(2):957–994

3. Digital cellular telecommunications system (Phase 2+) (GSM), Universal Mobile Telecommunications System (UMTS), LTE, Policy and charging control architecture. V15.4.0, 3GPP TS 23.203 (2018)
4. Brendan McMahan H, Moore E, Ramage D, Hampson S, Agüera y Arcas B (2017) Communication-efficient learning of deep networks from decentralized data. In: *Artificial intelligence and statistics*, pp 1273–1282
5. Ye Y, Li S, Liu F, Tang Y, Hu W (2020) EdgeFed: optimized federated learning based on edge computing. *IEEE Access* 8:209191–209198
6. Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang YC, Kim DI (2019) Applications of deep reinforcement learning in communications and networking: a survey. *IEEE Commun Surv Tutor* 21(4):3133–3174
7. Antunes RS, André da Costa C, Küderle A, Yari IA (2022) Federated learning for healthcare: systematic review and architecture proposal. *ACM Trans Intell Syst Technol* 13(4):1–23
8. Nguyen DC, Pham QV, Pathirana PN, Ding M, Seneviratne A, Lin Z, Dobre O, Hwang WJ (2022) Federated learning for smart healthcare: a survey. *ACM Comput Surv* 55(3):1–37
9. Petroni A, Salvo P, Cuomo F (2021) On cellular network supporting healthcare remote monitoring in IoT scenarios. *Front Commun Netw* 2

ML/SDN-Based MEC Resource Management for QoS Assurances



Seyha Ros, Chanthol Eang, Prohim Tam, and Seokhoon Kim

Abstract Mobile edge computing (MEC) extracts the resource closer to end users; however, resource orchestration and management are still mentioned as the important issues. We propose the software-defined networking (SDN) technique to provide the control management based on the centralized global controller. The benefits of the SDN involvement for flexibility and efficient MEC resources management will be gained. In this paper, the MEC resource will be coordinated by SDN, in order to provide the efficient operation for MEC resources in the network. Moreover, we present the benefits of SDN within the heterogeneous Internet of Things (HetIoTs) in order to provide the effective quality of services (QoS) for massive traffic. With the overview of resourceful HetIoTs data-driven in 5G and the adoption of MEC-SDN benefits, the proposed approach presents the study with K-Mean-based assistance. Since the user traffic will be waiting for serving or disconnecting when there are insufficient resources in the packet data gateway (PGW). In this mentioned issue, priority queuing has been proposed to tackle network congestion problems in IP networking. In this paper, we also propose and employ priority queueing to backhaul network to manage the users' traffic in PGW. Based on K-Mean clustering algorithm, this approach possibly classifies the communication types and allows priority queue to schedule the traffic effectively. The main goal is to reduce network congestion by improving QoS and 5G network performance.

Keywords Quality of service · Software-defined networking · Mobile edge computing · K-mean · Resource management

S. Ros · C. Eang · P. Tam · S. Kim (✉)
Department of Software Convergence, Soonchunhyang University, Asan, Korea
e-mail: seokhoon@sch.ac.kr

S. Kim
Department of Computer Software Engineering, Soonchunhyang University, Asan, Korea

1 Introduction

Heterogenous Internet of Things (HetIoTs) has been employed in the last decade. Currently, there are many HetIoTs applications for different purposes, which have been introduced in 5G mobile communication to interconnect electronic devices reaching the Internet. This connectivity of HetIoT devices required a huge amount of radio network resources. To perform the requirement, mobile edge computing (MEC) has been proposed to increase the computing resources at the radio access network (RAN) while moving the necessary computing elements to RAN field [1, 2]. This method claims to enable local communication, and gains a numerous benefits such as radio services, saving the HetIoTs devices power consumption and offer longer battery life, decreasing end-to-end communication latency, and reducing outgoing traffic to remote gateway for handling the heavy utilization of optical resources.

Beside the above-mentioned statement, the operating expenditure (OPEX) and capital expenditure (CAPEX) will be occurred in fronthaul network area. The OPEX and CAPEX will be affected to RAN network infrastructure, while the management and orchestration problem will become the main considering issues. To prevent the overspending of the OPEX, CAPEX, and power consumption problems, network function virtualization (NFV) becomes the key candidate to fix the troubles [3]. NFV enables to generate a multiple virtual machines (VMs) to serve the variety of network functions. However, there is inefficiency management of the virtual network computing resources. In order to handle this trouble, software-defined networking (SDN) concept will be applied, while the entire resources will be controlled by the centralized controller. SDN controller provides the scalable resources management, flexibility, and reduces the complication at the control plane.

With HetIoTs deployment, each application field has various purposes of usage for communication or interaction in their specific needs which leads developers/researchers to build any types of applications in networking technology nowadays. These collections of services possibly lead networking traffic to chaos if any optimal methodologies or tools are not applied critically. Meanwhile, if any algorithms or controllers are implemented correctly, it will absolutely be beneficial for subscribers' communication needs. To accomplish this goal, the proposed approach in this paper will introduce K-Mean clustering algorithm for classifying the communication types, and priority queue method which is modeled as m/m/c for scheduler. The priority is based on real-time restriction communication level. These two methodologies work closely to provide an improvement in 5G communication, schedule an acceptable prioritization, reach users' satisfaction, and solve the chaos in MEC computation resources.

2 Related Works

2.1 Emergence of MEC, NFV, and SDN

As mentioned in the previous section, MEC was proposed to migrate to the lower capacity of MCC elements close to edge networking in order to enable local communication. Many mobile operators employ MEC to improve the QoS for service level agreement (SLA) customers. The legacy MEC was totally made by physical resources. Therefore, the investment cost is higher, and the infrastructure will be extended. To overcome these mentioned issues, several complementary studies present the involving of NFV into MEC. In NFV-assisted scenario, the physical MEC entities will be migrated to virtual MEC (vMEC) [4, 5]. The virtual function can be offered from dedicated MEC hardware. The hypervisor is the main candidate in order to provide the handling between hardware and guest operating systems such as Windows and Linux operating system. The variety of virtual function application will be installed to the VMs. In other scenarios, hypervisor take a role as the middle of host and guest operating systems.

The fully programmable networking environment was introduced and located at the SDN centralized control plane [6, 7]. The OpenFlow is the open protocol enabling the interaction between the decoupling data plane and control plane. The controller monitors the data plane for the global view capability. SDN controller located in the central entity between application plane and user plane. The southbound interface was introduced to enable monitoring, information collection, and control information from the requests of data plane. The northbound interface takes advantages as the communication interface between the servings from various applications such as QoS and load balancing. This interface provides the application programming interface (API) from the user, while the users able to control the network by service algorithms based on their requirement key indicators [8].

2.2 Machine Learning-Based and Priority Queue Control Mechanisms

Machine learning (ML)-based control system can be helpful to integrate with SDN platforms [9–11]. K-Mean clustering is one of the most well-known and convenient unsupervised ML algorithms in SDN-based network traffic classification. It is highly beneficial to network management, IP engineering control, organized resource allocation, etc. And priority queue (PQ) is a common sequence which has structural feature arrangement based on high or low priority. Both of these paradigms have applied in related studies and produced remarkable results. PQ was proposed in a cyber-network security detection system which combined with head of line and shortest processing time strategies to schedule the mission arrangement intelligently. Meanwhile, it was also implemented for advanced scheduling in big-throughput network.

In this domain study, it starts with constructing a pipeline which manage the operations. And it enables enormous priority queuing for acceptable demonstrating at a suitable cost. Additionally, with priority queuing, the categorized traffic flow of diversified applications or communication types are solved. It possibly designs a system based on many collections of self-similar traffic flow. In this area, PQ is enhanced with Empty Buffer Approximation method, so it become a Single-Server-Single-Queue collectively which is easy to control the challenging performance.

3 ML/SDN-Based Resource Management for QoS Assurances

3.1 SDN-Based Management of vMEC Resources

In this paper, we focus on the significant management of the vMEC resources. As presented in Fig. 1, the MECs are located close to each radio remote head (RRH). To overcome the legacy MEC management and orchestration, the centralized control by SDN controller will be involved in this approach. Each of MEC at the RRH provides either physical or virtualized computing components based on the resource condition. Flow monitor, flexibility management entity, and controller are located at the SDN controller. While the flows are monitoring and collecting the information of all MEC resources, MEC provides the services to HetIoT devices based on the rule from the SDN controller. The controller synchronizes with MEC servers. The entire HetIoT requesting information are monitored and stored at the controller. The orchestrator controls the generation of virtual function with VM of MEC servers. While the physical resources reached the defined threshold, the proposed orchestrator/controller will send a rule to the MEC server to generate the VM blocks and offload the physical resources for execution.

3.2 K-Mean-Based Multi-service QoS Prioritization

The user traffic has been classified into four different communication types as shown in Fig. 2, since the classification is constructed by K-Mean ML algorithm which cluster the user traffic by K value (number of clusters). In this approach, we assign the $K = 4$, because there are four types of communications will be classified such as conversation, streaming, interactive, and background communication. The classification data set is based on the bit error ratio (BER) of each communication type, while it has different BER. In this paper, we assumed the conversation BER equal to 10^{-3} or 10^{-4} , the streaming BER equal to 10^{-5} or 10^{-6} , the interactive BER equal to 10^{-7} , and the background communication BER equal to 10^{-7} or 10^{-6} .

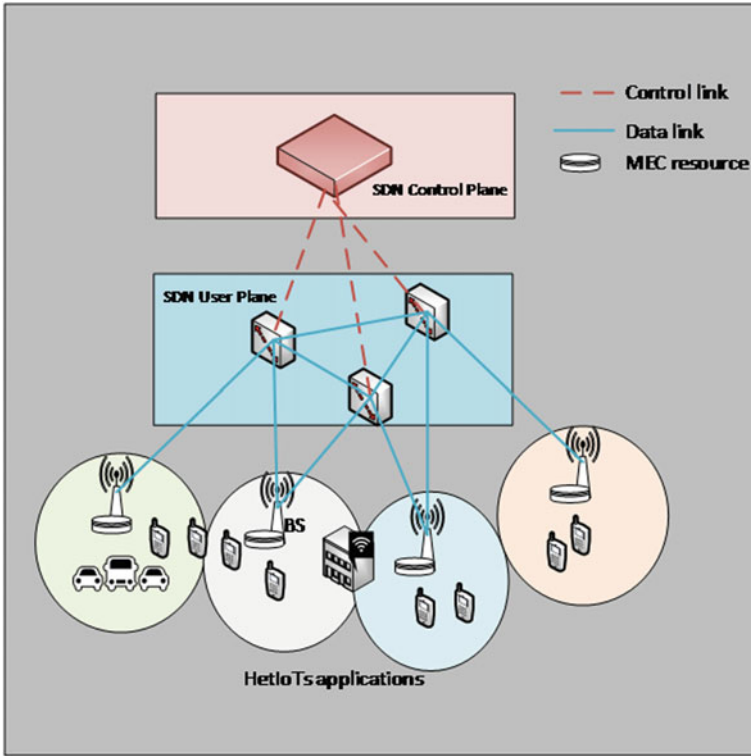


Fig. 1 The architecture of MEC controlled by SDN controller

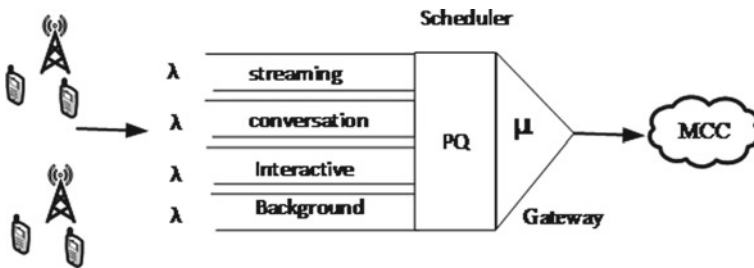


Fig. 2 Scheduler prioritization of four communication types based on PQ

After classification, each communication type will be queued into independence queueing system, so there are four difference queues for serving the difference of communication types. This queue system could be modeled as $m/m/c$. Then, this approach installs the scheduler for each queue that handles the traffic based on the priority queueing system. While the streaming and conversation communications are the real-time applications that have a higher priority to increase forwarding rate and

reduce queuing length in order to decrease the packet drop. Meanwhile, interactive and background communication are not real-time restriction communication types, therefore the lower priority will be assigned to both of these types. By decidedly prioritizing real-time communication, users' satisfaction, network performance, and QoS are superior. The resource in NFV-enabled vMEC can be orchestrated based on K-Mean prioritization, which is insightful for the control systems.

4 Conclusion

We present the effective virtual/physical MEC resources management based on SDN concept. This proposed approach claims to reduce the complicated management and orchestration of MEC/vMEC resources, and provides the scaling/flexibility management. MEC/vMEC servers has been assumed as the user plane in this approach, and we expected to improve the QoS as well as the network performance for both RAN area and evolved packet core gateway. Especially, MEC/vMEC provides the effective resources based on the needs of each HetIoT application. Furthermore, the utilization of K-Mean clustering algorithm and PQ m/m/c model are addressed to improve the prioritization for higher required real-time communication. Assume that there are four communication types including conversation, streaming video/voice, interactive, and background communication, which are classified based on BER criteria and K-Mean algorithm. After classification, the priority queue system starts installing scheduler which efficiently identifying which independent queue requires to be prioritized with higher vMEC resource allocation. By implementing this approach, we ensure to be advantageous for improving the 5G network performance, reducing HetIoTs traffic complexities, enhancing multi-service QoS, and adjusting the vMEC resource control efficiently. With SDN/NFV-enabled control, virtualization offers flexibility in MEC resource management.

In the future study, deep reinforcement learning will be applied to enable autonomy control systems, and experimental analysis will be added in terms of (1) agent control and (2) end-to-end SDN/NFV/MEC-related communication flows. The performance metrics will be captured for extending this current version, in terms of QoS parameters.

Acknowledgements This work was supported by BK21 FOUR (Fostering Outstanding Universities for Research) under Grant 5199990914048.

References

1. Goh Y, Choi M, Jung J, Chung JM (2022) Partial offloading MEC optimization scheme using deep reinforcement leaning for XR real-time M7S devices. In: IEEE international on consumer electronics (ICCE), Las Vegas

2. Jiang F, Dong L, Wang K, Yang K, Pan C (2022) Distributed resource scheduling for large-scale MEC systems: a multiagent ensemble deep reinforcement learning with imitation acceleration. *IEEE Internet Things J* 9(9):6597–6610
3. Kiran N, Liu X, Wang S, Yin C (2020) VNF placement and resource allocation in SDN/NFV-enabled MEC networks. In: *IEEE wireless communications and networking conference workshop (WCNCW)*, Seoul
4. Yang B, Cao X, Li X, Zhang Q, Qian L (2019) Mobile-edge-computing-based hierarchical machine learning tasks distribution for IIoT. *IEEE Internet Things J* 7(3):2169–2180
5. Shah SD, Gregory MA, Li S, dos Reis Fontes R, Hou L (2022) SDN-based service mobility management in MEC-enabled 5G and beyond vehicular networks. *IEEE Internet Things J* 9(15):13425–13442
6. Baek J, Kaddoum G (2021) Online partial offloading and task scheduling in SDN-fog networks with deep recurrent reinforcement learning. *IEEE Internet Things J* 9(13):11578–11589
7. Filali A, Mlika Z, Cherkaoui S, Kobbane A (2022) Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services. *IEEE Trans Netw Sci Eng* 9(4):2174–2187
8. Trinh B, Muntean GM (2022) A deep reinforcement learning-based resource management scheme for SDN-MEC-supported XR applications. In: *IEEE 19th annual consumer communications and networking conference (CCNC)*, Las Vegas
9. Malik A, de Fréin R, Al-Zeyadi M, Andreu-Perez J (2020) Intelligent SDN traffic classification using deep learning: deep-SDN. In: *Proceedings of 2nd international conference on computer communication and the internet (ICCCI)*, Nagoya
10. Jiang F, Wang K, Dong L, Pan C, Xu W, Yang K (2020) Deep-learning-based joint resource scheduling algorithms for hybrid MEC networks. *IEEE Internet Things J* 7(7):6252–6265
11. Kim K, Park YM, Hong CS (2020) Machine learning based edge-assisted UAV computation offloading for data analyzing. In: *International conference on information networking (ICOIN)*, Barcelona

Movie Recommendation System Using Community Detection Based on the Girvan–Newman Algorithm



Sadriddinov Ilkhomjon Rovshan Ugli , Doo-Soon Park ,
Daeyoung Kim , Yixuan Yang , Sony Peng , and Sophort Siet 

Abstract Recently, a lot of research on the movie recommendation system has been conducted. The most important element in the movie recommendation system is to recommend the most suitable movie for an individual. To increase this accuracy, this paper proposes a movie recommendation system using community detection based on boundaries. In particular, the demographic characteristics of users are used to establish implicit bonds between users. After building an implicit social network, we applied community detection based on the Girvan–Newman algorithm to discover potential communities. Mean Average Error (MAE) was applied to evaluate the accuracy of the proposed model. The method presented in this paper improves the performance of the recommendation system by considering the user’s profile-based similarity.

Keywords Recommendation system · Community detection · Implicit social network

1 Introduction

In the nascent era of big data, it has been anticipated those recommendation systems (RS) and other kinds of knowledge-mining techniques will play a pivotal role in data-driven research. In the course of the past three decades, longitudinal studies have been conducted on recommendation systems using the glut of information. Various kinds of recommendation systems were designed and applied to a wide range

S. I. R. Ugli · D.-S. Park (✉) · D. Kim (✉)

Department of Computer Software Engineering, Soonchunhyang University, Asan, South Korea
e-mail: parkds@sch.ac.kr

D. Kim

e-mail: dyoung.kim@sch.ac.kr

Y. Yang · S. Peng · S. Siet

Department of Software Convergence, Soonchunhyang University, Asan, South Korea

of fields. Recently, many areas like business, e-commerce, entertainment, education, and others are taking advantage of RS. Popular examples include Amazon, YouTube, Netflix, LinkedIn, Facebook, Goodreads, etc. One can refer to Sect. 2.1 for further information about the recommendation system.

Community detection is one of the branches of Social Network Analysis (SNA) which tries to harvest the relevant information from given social networks [1]. The graph represents such connections, while the people are the nodes and the connections among them are edges. Community detection does not merely concern with the connections between users, but also the importance (power) of nodes and edges are essential. In other words, finding how a particular node is popular among its counterparts may contribute to identifying the splitting whole network into smaller communities.

Finding underlying patterns inside a network for recommendation systems helps to recommend more personalized items to the target users. However, it is not always the case that the data contains implicit relationships among users. Unlike explicit social networks, implicit social networks do not contain direct ties. Users from movie streaming platforms can be taken as an example. An individual user mostly preserves the history of movies viewed and ratings. Therefore, techniques from social network analysis are not applicable, unless one performs several data preprocessing stages in advance. Conversely, in all social network services, explicit social network properties pre-existed. Considering these aspects of social networks, the accuracy and effectiveness of recommendation systems can be increased. Community detection and its nature were discussed in more depth, in Sect. 2.2.

This paper is organized as follows. Section 2 surveys the recommendation systems and community detection. We discussed the research experiment and our method in more detail, in Sect. 3. Section 4 represents the experiment and evaluation of RS. Finally, Sect. 5 performs preliminary conclusions and future work.

2 Related Work

2.1 Recommendation Systems

The recommendation system does not only have a knock-on effect on platforms where it was applied but also leads to immediate outcomes. Many cases can be given from real-life examples: How Netflix capitalized on developing a better recommendation system.

There are two dominating methods in RS: collaborative filtering (CF) and content-based recommendation [2]. The intuition behind the CF method is based on the situation in which people incline to ask for suggestions from their friends [3]. From the movie recommendation perspective, if multiple users rated a movie identically, then the model considers them as “friends”. From that step on, the system tracks their actions and then recommends movies based on revealed patterns. On the other

hand, the content-based algorithm mainly focuses on the profile of a user and/or characteristics of an item. If the recommendation engine finds two users same in accordance with the profile-based similarity, then it provides an identical set of items for both [4]. Both techniques can be implemented simultaneously by decreasing the disadvantages. This method known as the hybrid approach is being used by many companies, nowadays [5].

However, the intervention of social networks in data mining altered the research trend of recommendation systems, over the past decades. Utilizing the direct connections of users' recommendation systems increased their accuracy.

2.2 Community Detection

The idea of community detection has appeared in network science as an approach to finding the groups inside intricate systems [6]. Communities are derived from social networks. In Fig. 1, a snippet of a social graph is illustrated.

Graph G is the pair of two sets: (V, E) . Here, $V = \{v_1, v_2, \dots, v_n\}$ is a set of nodes from graph G , while $E = \{v_i, v_j\}$ is a set of pairs of different nodes. Such kinds of pairs are called an edge. There are different types of graphs: complete, regular, isomorphic, homomorphism, dense and sparse graphs, weighted and unweighted graphs, trees, etc. There are many important properties of a graph: the degree of nodes, and the path between nodes. Moreover, the common way to express the graph is by using an adjacency matrix. If the graph is undirected, then an adjacency matrix

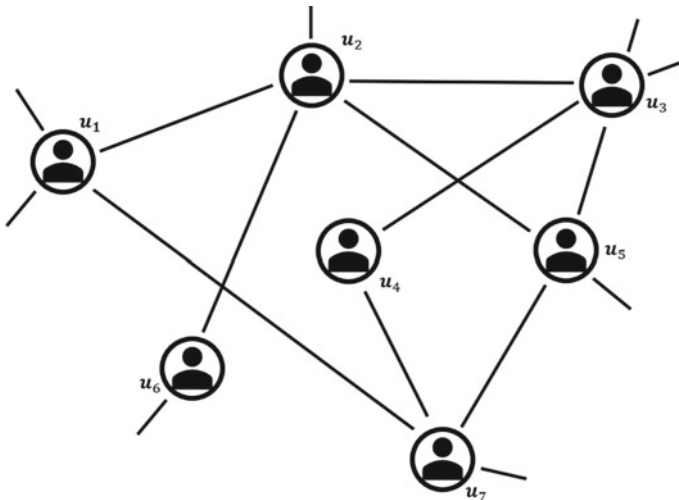


Fig. 1 Graph representation of a social graph

will be a symmetric matrix, otherwise (in the case of a directed graph) asymmetric square matrix.

Communities of a graph can be detected using several methods: Louvain, Surprise, Leiden, and Girvan–Newman Community detection algorithms. In this paper, the Girvan–Newman method was implemented to get communities. This method is based on values of edge betweenness [7]. This algorithm calculates the edge betweenness values first for all edges in a graph. Then, remove the edge with the highest edge betweenness value. This step is iteratively repeated until the favored community structure will be determined. To measure a particular division of the network once communities are detected, modularity is used [8].

3 Design for Movie Recommendation System Using Community Detection Based on the Girvan–Newman Algorithm

The environment for the experiment can be checked from the following Table 1.

Figure 2 elucidates the whole architecture of our method.

In the first stage, we performed some data preprocessing steps on our dataset. Only three attributes of users were used: user age, gender, and occupation. Since all three features were different, discretization was necessary. We created 5 discrete intervals from age values [9]. It was performed discretization for occupation attribute by minimizing the number of categories [10]. Some rows were eliminated from the dataset for sake of accurate recommendation. Then, we split our dataset into two sub-datasets: train (80% of the dataset) and test (20% of the dataset).

Based on these 3 values, we used the following equation to build a binary adjacency matrix:

$$(c_{ij}) = \begin{cases} 1, & D_i = D_j, i \neq j, D_{i,j} \in A, (1 \leq i, j \leq n) \\ 0, & D_i \neq D_j, i \neq j, D_{i,j} \in A, (1 \leq i, j \leq n) \end{cases} \quad (1)$$

Here, A is a user matrix contains n rows and three demographic attributes, and D is a user vector. While c_{ij} is the entity of the new adjacency matrix C which we will

Table 1 Experimental environment

Division	Detailed contents
CPU	Intel(R) Core(TM) i5-7500 CPU @ 3.40 GHz
RAM	8 GB
HDD	128 GB
OS	Windows 10 Pro 64 bit
DEV.TOOL	R version 4.0.3 (2020-10-10) RStudio Version 2022.07.1

Once we built an adjacency matrix, we applied a Girvan–Newman method to detect communities from our undirected graph. And from 24 communities found, we removed communities that contain less than 6 members.

4 Experiment and Result Analysis

All users from the test sub-dataset were classified into the communities calculating Euclidean distance. Once, the classification stage has finished we moved to the MAE evaluation part.

10 random users were selected, and then for each user, 10 random movies were chosen. We predicted the ratings of all randomly selected movies. Finally, we calculated the Mean Absolute Error between actual and predicted ratings. We got 0.94 for MAE which is relatively better.

5 Conclusion

In this paper, we applied a community detection method for a movie recommendation system. More particularly, the Girvan–Newman algorithm was used to determine communities from our constructed network. And we removed communities containing few members from the detected 24 communities. Finally, we received 0.94 for MAE.

In the future, we want to use different types of community detection algorithms and try to apply other datasets.

Acknowledgements This research was supported by the National Research Foundation of Korea (No. NRF-2022R1A2C1005921) and BK21 FOUR (Fostering Outstanding Universities for Research) (No.5199990914048).

References

1. Lie H, Maes P (2005) InterestMap: harvesting social network profiles for recommendations. In: Proceedings of IUI beyond personalization 2005: a workshop on the next stage of recommender systems research, Jan 9, 2005, San Diego, CA, USA, pp 54–59
2. Vilakone P, Park DS, Xinchang K et al (2018) An efficient movie recommendation algorithm based on improved k-clique. *Hum Cent Comput Inf Sci* 8:38
3. Ben-Shimon D, Tsikinovsky A, Rokach L, Meisles A, Shani G, Naamani L (2007) Recommender system from personal social networks. In: Wegrzyn-Wolska KM, Szczepaniak PS (eds) *Advances in intelligent web mastering. Advances in soft computing*, vol 43. Springer, Berlin
4. Park D-S (2013) Improved movie recommendation system based-on personal propensity and collaborative filtering. *KIPS Trans Comput Commun Syst* 2(11):475–482

5. Vilakone P, Xinchang K, Park D-S (2020) Movie recommendation system based on users' personal information and movies rated using the method of k-clique and normalized discounted cumulative gain. *J Inf Process Syst* 16(2):494–507
6. Allman A, Tang W, Daoutidis P (2018) Towards a generic algorithm for identifying high-quality decompositions of optimization problems. In: Eden MR, Ierapetritou MG, Towler GP (eds) *Computer aided chemical engineering*, vol 44. Elsevier, Amsterdam
7. Newman M, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69(2):026113
8. Newman MEJ (2001) Modularity and community structure in networks. *Phys Rev E* 64:016132
9. Our World in Data. https://ourworldindata.org/grapher/population-by-broad-age-group?country=~OWID_WRL. Last accessed 2022/10/31
10. International Labour Organization. <https://ilostat.ilo.org/resources/concepts-and-definitions/classification-occupation/>. Last accessed 2022/10/31

Efficient Container Management Scheme Based on Deep Learning Model



Byeonghui Jeong , Jueun Jeon , Seungyeon Baek ,
and Young-Sik Jeong 

Abstract The container orchestration platform provides services and applications to users by automatically managing containers in cloud-native clusters. However, the conventional container management technique operates based on a reactive mechanism, so it is difficult to guarantee availability for a rapidly changing workload, and a problem arises in which resources are wasted. Therefore, this study proposes an efficient container management scheme (ECMS) that guarantees high availability and scalability of cloud services and applications composed of various workloads and minimizes idle resources. ECMS predicts future workload by training accumulated past resource usage metrics on a deep learning model. ECMS creates a container by scaling the size of the resource to prevent the generation of idle resources, then selects a node that guarantees high availability and deploys the container. Lastly, it provides scalability of service by performing horizontal resource container autoscaling based on proactive mechanism.

Keywords Cloud computing · Container automation system · Resource usage forecasting · Deep learning

B. Jeong · J. Jeon · S. Baek · Y.-S. Jeong (✉)
Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea
e-mail: ysjeong@dongguk.edu

B. Jeong
e-mail: qudgm16323@dgu.ac.kr

J. Jeon
e-mail: jry02107@dongguk.edu

S. Baek
e-mail: tmddusdls12@dongguk.edu

1 Introduction

The container orchestration platform manages multiple containers configured in cloud-native cluster by utilizing automated container management technique. At this time, container scheduling and container autoscaling, which operate based on a reactive mechanism, cannot immediately respond to a temporarily rapidly changing workload, so it is difficult to guarantee availability and generating idle resources [1, 2].

Therefore, this study proposes an efficient container management scheme (ECMS) to improve the service overload and idle resource generation problems of the container scheduling and container autoscaling techniques used in the container orchestration platform. ECMS predicts future workload by training accumulated past resource usage metrics on a bi-directional gated recurrent unit (Bi-GRU) model. In addition, ECMS creates a container by calculating an elastic resource size through future workload. The created container is deployed on a node that guarantees high availability. Lastly, ECMS performs horizontal container resource autoscaling based on a proactive mechanism to guarantee service scalability for various workloads.

This study proposed ECMS, which provides a cloud service with high availability and scalability by minimizing the problem of generating idle resources for workload composed of various data distributions.

2 Related Works

Various studies have been conducted to efficiently manage cloud-native cluster using the container automation system [3, 4]. In this section, several universal methods for improving service scalability by scaling container resources are described.

Dang-Quang et al. [5] proposed autoscaling using the Bi-LSTM model to improve the problem of reactive mechanism-based autoscaling. This provides proactive mechanism-based autoscaling through the predicted resource usage to processes a burst workload more efficiently than conventional scaling.

Rossi et al. [6] proposed an Elastic Docker Swarm (EDS) consisting of threshold-based heuristics to prevent idle resource generation of cloud-native applications from various workloads. EDS reconfigured the threshold policy through multi-reinforced learning and performed hybrid autoscaling to provide efficient resource management.

The ECMS proposed in this study scales the size of the container by using the resource usage predicted through the deep learning model, and then deploys it on the optimal node. In addition, which is guarantees high availability and scalability for cloud-native services by performing scaling immediately from various workloads.

3 ECMS Scheme

In this study propose ECMS, a container automation management technique, to guarantee the high availability and scalability of cloud-native service. First, ECMS predicts future workloads by utilizing Bi-GRU model to guarantee quality of service (QoS). At this time, the size of the initial container resource set by the resource estimation of administrator is scaling to an elastic size. Furthermore, for service availability, the container is deployed on a node with a resource size larger than the future peak point. Lastly, proactive mechanism-based horizontal container resource autoscaling is performed to prevent the idle resource generation for loads of various sizes from future workload. Therefore, ECMS guarantees the availability of cloud-native services from various workloads and reduces idle resource generation.

4 Conclusion

This study proposed ECMS, to provide high availability and scalability of cloud-native service configured of multiple containers. The ECMS predicted future workload with high accuracy by training decomposed past workload records to Bi-GRU model. ECMS had scaling the container resource size to an elastic size to guarantees availability from rapidly changing workload and deploys the container on the optimal node. In addition, proactive mechanism-based horizontal container resource autoscaling was performed to provide scalability for loads of various sizes. As a result of evaluating the performance of ECMS, the number of overload occurrences and idle resource generation were reduced compared to the conventional automation management techniques of container orchestration platform.

The ECMS is difficult to guarantee full continuity of service due to the prediction error of Bi-GRU model. Therefore, we plan to conduct a study to add a threshold to ECMS to take into account the error of resource usage prediction.

Acknowledgements This work was supported by Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00047, Development of microservices development/operation platform technology that supports application service operation intelligence).

References

1. Toka L, Dobreff G, Fodor B, Sonkoly B (2021) Machine learning-based scaling management for kubernetes edge clusters. *IEEE Trans Netw Service Manage* 18:958–972
2. Carmen C (2022) Kubernetes scheduling: taxonomy, ongoing issues and challenges. *ACM Comput Surv*, 1–36
3. Zhiheng Z, Minxian X, Maria AR, Chengzhong X, Rajkumar B (2022) Machine learning-based orchestration of containers: a taxonomy and future directions. *ACM Comput Surv* 54(10):1–35

4. Zhicheng C, Rajkumar B (2022) Inverse queuing model-based feedback control for elastic container provisioning of web systems in kubernetes. *IEEE Trans Comput* 71(2):337–348
5. Dang-Quang NM, Yoo M (2021) Deep learning-based autoscaling using bidirectional long short-term memory for kubernetes. *Appl Sci-Basel* 11:1–25
6. Rossi F, Nardelli M, Cardellini V (2019) Horizontal and vertical scaling of container-based applications using reinforcement learning. In: *Proceeding of the ieee international conference on cloud computing*, 329–338

Actual Resource Usage-Based Container Scheduler for High Resource Utilization



Sihyun Park , Jueun Jeon , Byeonghui Jeong , Kyuwon Park ,
Seungyeon Baek , and Young-Sik Jeong 

Abstract Kubernetes select node and deploy pod based on request to ensure the size of resources for containers with various requirements. In this case, containers are inefficiently managed due to idle resources which are generated by workload configured in various sizes. Therefore, in this study, we propose an Actual Resource Usage-based Scheduler (ARUS), which utilizes the resource usage of each component to perform scheduling to improve the problem of resource waste. ARUS forecasts future resource usage from collected resource usage by utilizing DLinear model. In this case, the optimal node is selected through the scoring for efficient resource utilization (SERU) algorithm. Therefore, ARUS improves resource utilization over conventional kube-scheduler.

Keywords Cloud computing · Container orchestration · Scheduling · Time series forecasting · Deep learning

S. Park · J. Jeon · B. Jeong · K. Park · S. Baek · Y.-S. Jeong (✉)
Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea
e-mail: ysjeong@dongguk.edu

S. Park
e-mail: psh0430@dongguk.edu

J. Jeon
e-mail: jry02107@dongguk.edu

B. Jeong
e-mail: qudgml6323@dgu.ac.kr

K. Park
e-mail: q1park3817@dgu.ac.kr

S. Baek
e-mail: tmddusdls12@dongguk.edu

1 Introduction

Kubernetes provides automated container managing techniques, such as scheduling, auto-scaling, and load balancing to efficiently manage the cloud configured with large amounts of containers [1]. Among them, the kube-scheduler, which performs the container scheduling, selects nodes by performing filtering and scoring steps based on the request to execute the pod [2]. First, in the filtering step, nodes with resources consisting of the size of the request or more are selected so that the pod can be stably executed. The scoring step evaluates the state of the node based on the resource size and pod request selected through the previous step, and then deploys the pod to the optimal node [3]. In this case, if the resource in the pod allocated by the request size is not shared with other components and the resource usage decreases, idle resources occur. Therefore, due to idle resources created in the cluster, additional pods cannot be generated, resulting in low cluster resource utilization [4].

In this study, we propose an Actual Resource Usage-based Scheduler (ARUS) that utilizes future resource usage to minimize idle resources in the cluster [5]. First, ARUS forecasts future CPU and memory usage by learning CPU and memory usage records of the collected components in the cluster into the DLiner model. At this time, the optimal node is selected based on the actual resource usage through the scoring for efficient resource utilization (SERU) algorithm that utilizes the forecasted resource usage. Therefore, ARUS minimizes idle resources to ensure high resource utilization.

2 Related Works

Studies of scheduling techniques using mathematical, heuristic, machine learning, and deep learning are being conducted to improve the problem of resource waste in cloud computing. This section describes studies that perform container scheduling using the deep learning model.

Huang et al. [6] proposed an RLSK that performs scheduling based on a reinforcement learning model. RLSK uses a scheduling algorithm based on Deep Q-Network (DQN) reinforcement learning model to minimize differences in other types of resource utilization in the cluster, thereby improving overall resource utilization.

Liu et al. [7] proposed Deep Learning-driven Scheduler for Deep Learning clusters (DL²) that considers the characteristics of the workload calculated from the deep learning model. DL² learned the existing scheduling policy through offline supervised learning. After that, resources were dynamically allocated by adjusting the scheduling policy in detail through online reinforcement learning. Therefore, DL² minimizes overhead and improves the accuracy of policy learning.

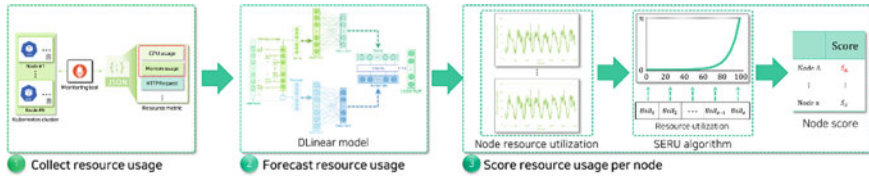


Fig. 1 Scheme of actual resource usage-based scheduler (ARUS)

3 Scheme of Actual Resource Usage-Based Scheduler (ARUS)

In this study, ARUS is proposed to efficiently utilize the resources of the container-based cluster. The overall scheme of ARUS is shown in Fig. 1.

ARUS, which consists of three steps, performs scheduling based on resource usage to minimize idle resources in a cluster. First, CPU and memory usage are extracted from the metrics of components collected through the monitoring tool. The extracted resource usage is learned in the DLinear model to predict future resource usage. Finally, resource utilization is calculated for each node configured in the cluster by applying a SERU algorithm that utilizes the predicted resource usage, and then the optimal node is selected. Therefore, ARUS efficiently manages the container by preventing idle resource generation and node overload.

4 Conclusions

In order to efficiently utilize resources in a container-based cluster, ARUS is proposed to reduce the occurrence of idle resources generated during container scheduling. ARUS predicted future resource usage with high accuracy through a DLinear model that considers temporal characteristics. In addition, in order to select nodes based on resource usage through the SERU algorithm, the resource utilization of each node was calculated, and the optimal node was selected. As a result of evaluating the performance of ARUS, cluster resource utilization was improved over Kubernetes' kube-scheduler.

As the SERU algorithm is based on the exponential function, the effect on scheduling increases rapidly as the resource utilization increases. Therefore, we are planning to conduct a study that adds scoring algorithms to ARUS considering various functions.

Acknowledgements This work was supported by the Culture, Sports, and Tourism R&D Program through the Korea Creative Content Agency (KOCCA) grant funded by the Korea government (MCST) (No. R2022020120).

References

1. Rodriguez MA, Buyya R (2019) Container-based cluster orchestration systems: a taxonomy and future directions. *Software: practice and experience* 49(5), 698–719
2. Carrión C (2022) Kubernetes scheduling: taxonomy, ongoing issues and challenges. *ACM Comput Surv*, 1–36
3. Menouer T (2021) KCSS: Kubernetes container scheduling strategy. *J Supercomput* 77(5):4267–4293
4. Yang Y, Chen L (2019) Design of Kubernetes scheduling strategy based on LSTM and grey model. In: 14th international conference on intelligent systems and knowledge engineering. IEEE, pp 701–707
5. Zeng A, Chen M, Zhang L, Xu Q (2022) Are transformers effective for time series forecasting? arXiv preprint [arXiv:2205.13504](https://arxiv.org/abs/2205.13504), 1–15
6. Huang J, Xiao C, Wu W (2020) Rlsk: a job scheduler for federated kubernetes clusters based on reinforcement learning. In: International conference on cloud engineering. IEEE, pp. 116–123
7. Peng Y, Bao Y, Chen Y, Wu C, Meng C, Lin W (2021) DL2: a deep learning-driven scheduler for deep learning clusters. *IEEE Trans Parallel Distrib Syst* 32(8):1947–1960

Solving Pre-erection Area Block Placement Problem Using Deep Reinforcement Learning



Yoon Hyun Kim  and Gijoo Yang

Abstract In shipbuilding, Pre-erection (PE) area block placement planning is one of the complex and time-consuming processes that require the ship launch schedule, locations of the deployed blocks, and the travel distance of the Goliath cranes and transporters. It takes several weeks even for human experts to produce a reasonable plan for block placement. In this paper, we develop an efficient and time-saving placement method by applying reinforcement learning (RL) in order to maximize the usage of PE field in shipbuilding environment. We created a batch simulation using digital twin for the environment of RL. To model shipyards, we used Unity 3D, a 3D game engine based on actual ship block data and shipyard data. To check the availability of block placement, first, from Unity 3D representation, we decide in which part of the PE area the given block can fit. Second, since the shape and size of the given block can widely vary, we use the convolution concept for the discretization of the placement information of PE area. We used this information as the state of our RL model. To capture the characteristics of the given state and promote the efficiency of our RL model, we use a CNN-based auto-encoder for our policy network. We designed the reward value for our RL model so that the travel distance of Goliath crane can be minimized. In our experiment, we showed that our RL agent performs better than human experts by about 15%.

Keywords Deep reinforcement learning · Digital twin · PE area in shipyard

Y. H. Kim

Korea Shipbuilding and Offshore Engineering, Seongnam-Si 13591, South Korea
e-mail: yhkim@ksoe.co.kr

G. Yang (✉)

Dongguk University, Seoul 04620, South Korea
e-mail: gjyang@dgu.edu

1 Introduction

In the recent years, significant progress has been made in solving challenging problems using artificial intelligence technologies. Among them, deep reinforcement learning (Deep RL) techniques are emerging where agents directly interact with given environments and learn through experiences.

RL can also be applied to block placement problem. However, since the space of PE area is limited, major problem can arise. After placing specific number of blocks within the limited spaces, blocks are erected at the dock. If the block placement is not efficiently performed, the PE space can be filled before all the blocks are placed. In addition, the size, shape, and quantity of blocks to be handled in a specific construction batch vary every time. Furthermore, the available space of the PE area continues to change according to the plan for carrying in and out blocks from/to the PE area. For this reason, it is difficult to obtain consistent information about the PE area. Also, since the placement process is continuously performed, the target batch is affected by the previous and subsequent batch processes. Therefore, we must handle two or more batches that must be processed at the same time. For this reason, even for human experts, several weeks are needed to produce a reasonable plan for block placement. To address this challenge, we propose a Deep RL-based block placement approach, where we train an agent so that it tries to maximize the usage of PE area. To implement the environment for our RL model, we use Unity 3d to model the process of the shipbuilding based on the actual shipyard data and ship block data. In each episode, all the blocks of the ships are sequentially placed by the RL agent according to the erection sequential order. The objective is to place blocks on the PE area, such that the travel distance of the Goliath crane and the usage of the PE area are optimized. Training is guided by a reward function using the travel distance of the Goliath crane movement for each of the RL agent's block placement action.

In the previous methods [1], PE area is divided into grids and all the shapes of the blocks are assumed to be same. However, since the real-world shipyard PE area and blocks are not, our approach checks the availability of placement of the given block in the PE area by using the Unity3D. And, since the size of the real-world shipyard and PE is huge (e.g., 55 m \times 738.5 m), in order to handle the huge size of the agent's action space, we use the convolution concept for discretization of the placement information of PE area. To decrease the size of the action space we adapt the masking method [2]. If this method is not used, the action space will become too large, and as the result, the agent will not learn well. This approach dramatically reduces the size of the action space.

To apply Deep RL to the simulation, we use Unity ML-Agents python low level API. In order to compare human expert's PDF practice with our work results, we transformed human expert's practice into Unity 3D representation.

The contribution of our work is as the following:

- (1) To our knowledge, our work is the first attempt to solve the actual shipyard PE problem by using deep RL

- (2) Our method is the first placement approach endowed with the ability to consider the irregular sizes and shapes of the blocks. With this ability, we are able to apply our method to the PE process of the shipyard based on actual data
- (3) In order to handle the huge size of the agent's action space, we use the convolution concept for the discretization of the placement information of PE area
- (4) Our experiments show that, our method is both faster and better at generating optimized placements for target blocks than human experts

2 Related Works

Deep RL [3, 4] is the combination of reinforcement learning and deep learning. RL is the area of machine learning that deals with sequential decision-making problems. It can be formalized as an agent that has to make decisions in an environment to maximize cumulative rewards. In RL a software agent collects state information through observations and takes actions within an environment, and in return it receives rewards. Its objective is to learn to act in a way that it will maximize its expected cumulative rewards. Deep RL has been successful in complicated sequential decision-making tasks with lower prior knowledge thanks to its ability to learn different levels of abstractions from data.

Minh [5] approximated the action value function by applying a deep neural network to the existing Q-learning technique and introduced the concept of a replay buffer that reuses previous experiences. Through these methods, learning was stabilized and made a great performance improvement.

However, in the case of continuous action space, it is difficult to use the above methods. The method that emerged to solve this problem is the policy-based approach. In this work, we use policy-based method. The policy-based method is defined as $\pi_\theta(s, a)$ using deep learning method. This policy network is trained using the gradient ascent method, and the loss function (= the objective function) is defined as the following.

$$\nabla_\theta J(\theta) = E_{\pi_\theta} [\nabla_\theta \log \pi_\theta(s, a) * Q_{\pi_\theta}(s, a)] \quad (1)$$

$$\nabla_\theta J(\theta) = E_{\pi_\theta} [\nabla_\theta \log \pi_\theta(s, a) * G_t] \quad (2)$$

In Eq. (1), the method of changing $Q_{\pi_\theta}(s, a)$ to G (return) is called REINFORCE.

Lim and Rho [1], Pan and Jason [6] tried to solve the placement problem based on genetic algorithms [7]. Genetic algorithms are computational models based on the evolutionary process of the natural world. Starting from the initial population, the group evolves, and evolution proceeds with genetic operations of selection, mutation, and crossover. Through these genetic operations, the optimal solution is found. To minimize the cost of the relocation, the distance travelled between each block is used as an objective function. The blocks in the PE area are expressed as chromosomes

through encoding. The cost of relocation is minimized using selection, mutation, and crossover operations as optimal placement algorithms. However, the size of the PE area is simplified into 5×5 , 10×10 , 25×25 grids. In addition, it is assumed that the shape of blocks is all same. This method is difficult to apply to the real-world shipyard environment.

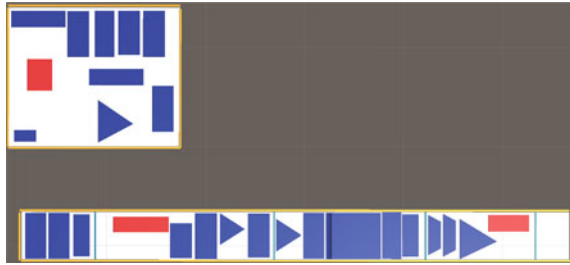
Google [2] developed a method for placing chips by using RL. The agent starts with an empty chip canvas state s_0 and sequentially places one macro through an action a_0 . Action is the position on the chip canvas to be placed. The policy network that determines action used Actor-Critic. In this method, network is updated through the state-of-the-art algorithm, Proximal Policy Optimization (PPO) [8] method. The policy network acts as a decoder. It receives the state information and outputs the probability distribution. Based on this probability distribution matrix, the agent selects one cell. A masking method is used to reduce the size of the action space. As a reward function, the wire-length and routing congestion is used. It is set to receive higher reward if the wire length is short and the routing congestion is low. This method shortened the working time to less than 6 h, while it takes months for human experts to work. However, in the process of discretizing chip canvas into grid cells, each cell size is set to be same. In this setting, it is impossible to place blocks of various sizes and shapes.

3 Methods

3.1 Data

In this paper, Unity3D, a 3D game engine, is used to implement the simulation of the PE area. We implement real-time simulation that can place blocks in PE area using Unity 3D and actual shipyard data. Blocks are created using AutoCAD based on the actual ship block data. The distance of 1 m of the shipyard is set to 1 of the X, Y, Z coordinates in Unity 3D. These allow us to implement the simulation that reflect the real-world shipyard. The order of the block placement is created in the form of a csv file using our own module by receiving the real-world shipyard sequential order that has been used before. Simulation operates in the order of the data items in the csv file. Blocks consist of two to six sub-blocks. Figure 1 is a picture taken with a composition by overlooking the blocks placed in advance before the start of the simulation. The blue blocks are the blocks of the ship. The red block is an area that cannot be placed. The sky-blue blocks are railroads which the Goliath crane passes through. And any block cannot be placed on them. For convenience, PE area shown above will be PE 1 and below will be PE 2.

Fig. 1 PE area and blocks implemented by unity 3D



3.2 Reinforcement Learning

Environment

An agent interacts with the environment through actions. The environment causes a change in a state through the agent's actions. The environment gives the agent the next state and a reward. The agent receives the next state and a reward accordingly. The agent of RL learns to maximize a cumulative reward received in the episodes. As a simulation is performed, an episode sample is created. By using the episodes created in this way, the agent learns from episodes.

We implemented our environment by using Unity3D and AutoCAD. To shorten our simulation time, only the necessary steps for the placement of PE area are visualized in our simulation. And for the same reason, actual Goliath crane is not visualized. Hence, the block placement in PE area is shown as a block movement. Other unnecessary steps are omitted visually but made to work internally using C# scripts. PE area is implemented in Unity3D based on the actual size of the PE area, and the spots that cannot be used in PE area are also implemented. It is noteworthy to mention that this is similar to the real-world shipyard PE area.

Reward

Reward is evaluated and given after a single episode. When the agent fails to place even one block, it is considered as a failure and the reward is set to -10 points. If the agent succeeds, a positive reward is given. Our reward function is determined by the travel distance of the Goliath crane (D). By subtracting D from the sufficiently large value of 1000, our reward value is obtained, i.e. the smaller the D gets, the higher the reward becomes. Our reward function is shown in Eq. (3) as the following:

$$R = \begin{cases} |1000 - \sum D| & \text{if succeeds} \\ -10 & \text{if fails} \end{cases} \quad (3)$$

Policy Network

The RL algorithm we propose in this paper is REINFORCE [9]. The reason why we chose REINFORCE is: (1) REINFORCE algorithm updates the policy network after

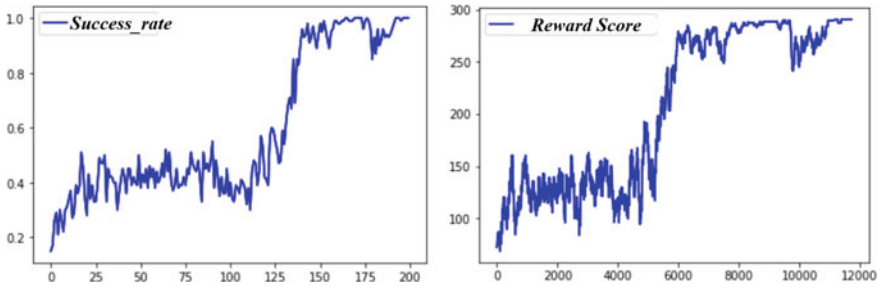


Fig. 2 Left, the success rate. Right, moving average of the reward score

every episode and thus, it needs episodes that always end (2) the episodes of PE area block placement problem always ends because the number of blocks is limited (e.g., 56 blocks in our case). As the input for the policy network, fixed size state matrix is used. Then, the state matrix is used as an input for a CNN-based Auto-encoder [10]. We implement our RL with PyTorch. We use two GPUs to update the network. We use Adam optimizer. We have learned 20,000 episodes and the total learning time was approximately 45 h.

4 Experiment

Figure 2 shows the success rate and the score of the block placement for the reinforcement learning agent. Figure 2 left represents the success rate with respect to the number of episodes. Figure 2 right represents the reward score with respect to the number of episodes. As shown, the success rate increases as the episode increases. And after 20,000 episodes, it converges to 1.0.

To evaluate the agent, we use the rewards that the agent received at every episode as a score. Since all failed episodes get -10 points, we represent only the successful episode data on the graph. It represents the moving average in units of 100 episodes. As shown in Fig. 2 Right, after about 12,000 episodes, it converges to approximately 300. The reward values achieved by human experts and our RL are 260, 298 respectively.

References

1. Lim B, Roh M (2010) Study on optimistic placement in shipyard PE area based on genetic algorithm. *J CDE Acad*, 553–561
2. Mirhoseini A, Goldie A, Yazgan M et al (2021) A graph placement methodology for fast chip design. *Nature* 594:207–212
3. Szepesvári C (2010) Algorithms for reinforcement learning. *Synth Lect Artif Intell Mach Learn* 4(1):1–103

4. Kaelbling LP, Michael LL, Andrew WM (1996) Reinforcement learning: a survey. *J Artif Intell Res* 4:237–285
5. Mnih V et al (2013) Playing ATARI with deep reinforcement learning. arXiv preprint [arXiv:1312.5602](https://arxiv.org/abs/1312.5602)
6. Pan JC-H, Po-Hsun S, Ming-Hung W (2015) Order batching in a pick-and-pass warehousing system with group genetic algorithm. *Omega* 57:238–248
7. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison Wesley, Reading, MA
8. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) CoRR
9. Sutton RS, McAllester D, Singh S, Yishay Policy gradient methods for reinforcement learning with function approximation. Mansour AT&T Labs—Research, 180 Park Avenue, Florham Park. NJ 07932
10. Hinton G, Salakhutdinov R (2006) Reducing the dimensionality of data with neural networks. *Science* **313** (5786):504–507

Design of Reliable Electronic Document Based on HTML and Blockchain



Andrew Hwang and Ji Su Park

Abstract An electronic document based on HTML via digital communication is the mainstream in a mobile environment, and the document containing personal legal information gets to be delivered in this way. However, HTML does not have the content integrity verification specification; therefore, it is hard to prove the content integrity once an electronic document has been delivered. We have designed the Reliable Electronic Document System using a metatag definition in HTML to store a content digest value and a metadata definition in a blockchain to store the PKI certificate to verify the content integrity. Furthermore, we confirmed that the Reliable Electronic Document System generates a reliable document and verifies content integrity. As a result, we expect this system to be used in customer communication by providing a reliable electronic document.

Keywords Reliable electronic document · HTML · Blockchain · Customer communication

1 Introduction

Digital communication, such as email, has been the mainstream rather than physical communication based on paper for customer communication between an enterprise and a customer after being paperless era. Digital communication still has a vulnerability in content and delivery integrity, whilst traditional communication has been guaranteed by government law. Documents which have a legal effect on an enterprise and a government also has been delivered through the digital communication channel recently. However, these electronic documents based on HTML do not

A. Hwang
Principal Solution Architect, Quadient, Singapore, Singapore
e-mail: a.hwang@quadient.com

J. S. Park (✉)
Department of Computer Science and Engineering, Jeonju University, Jeonju-si, South Korea
e-mail: jisupark@jj.ac.kr

provide content integrity verification, whereas legal documents should have content integrity and proof of the delivery to a receiver to avoid a possible legal dispute. In this research, we have designed the Reliable Electronic Document System based on HTML and Blockchain to generate a reliable document which has content integrity. And we did experimental verification and confirmed that the Reliable Electronic Document System provides content integrity.

2 Related Research

2.1 Electronic Document

The electronic document is the structured data consumed by a human [1], and there are many types of electronic documents in various domains. The layout of an electronic document intended to be read by a human has a similar layout to a traditional paper document. PDF, which is the de facto standard specification for an electronic document, is the ISO 32000 standard specification and was developed by Adobe [2]. It is an independent format against OS and devices and does not require loading an external resource. HTML is the standard web language, and it is used for content display on a web or mobile environment [3]. As the mobile era is coming, electronic document based on HTML is getting a more critical document format, providing a better user experience in a mobile environment. However, an HTML document consists of multiple files, such as an external image. Therefore, the content can be changed if an external resource is altered. Moreover, there is no content integrity verification specification [3].

2.2 Blockchain

Blockchain technology was invented for cyber cryptography in 2008 [4]. It stores the data in a block that connects to the previous block, and each block has the previous block's hash digest value. It is impossible to modify the content in a block unless all blocks are modified from the first block, which is called a genesis block. Blockchain technology provides non-fungibility by using that mechanism. And all transaction data in the blockchain, called the ledger, is de-centralized, and anyone can attend the blockchain ledger. Blockchain provides decentralization, anonymous access and non-fungibility characteristics [5]. Many enterprises adopt blockchain technology to store the data needed to secure content integrity and transaction history with non-fungibility.

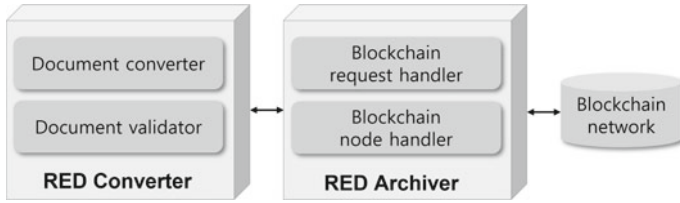


Fig. 1 Architecture of the reliable electronic document system

3 Reliable Electronic Document System

3.1 Design of Reliable Electronic Document System

Reliable Electronic Document System is a system to convert an electronic document based on HTML into a reliable electronic document to provide content integrity. This system receives an HTML document and responds to the reliable electronic document to a requestor, as shown in Fig. 1. Also, the system creates a new transaction in the Reliable Electronic Document Blockchain and stores the metadata of the generated reliable electronic document.

It consists of the Reliable Electronic Document (RED) Converter, which processes converting to a reliable electronic document, and the RED Archiver component, which stores metadata in the Reliable Electronic Document Blockchain.

3.2 Design of Reliable Electronic Document Converter

The Reliable Electronic Document is the limited specification of HTML to have content integrity. The document converter in the RED converter, as shown in Fig. 1, converts the original document to a Reliable Electronic Document as well as requests to create a new transaction to the blockchain request handler in the RED Archiver to store the metadata of the document in the Reliable Electronic Document Blockchain. First, all related resources in the Reliable Electronic Document must be embedded using Data URI Scheme as shown below in Fig. 2 [6].

Second, there is no content integrity verification even if the Reliable Electronic Document consists of a single file because the document can be altered in an unauthorized manner. Therefore, we defined the extended Reliable Electronic Document

```
data: [<media type>] [;base64], <data>
```

Fig. 2 Data URI scheme

```
<meta name="rd-range" content="byte position"/>  
<meta name="rd-digest" content="labeled message digest"/>  
<meta name="rd-signed-digest" content="signed message digest"/>  
<meta name="rd-address" content="blockchain transaction hash"/>
```

Fig. 3 Extended meta tag definition

meta tag to insert a content digest and an encrypted content digest by using a PKI certificate, as shown below in Fig. 3.

The rd-range meta tag indicates the byte position of the Reliable Electronic Document in a file. The rd-digest meta tag indicates the message digest value, which is calculated by using a hash algorithm. The original content can be verified using the rd-digest meta tag value to determine whether the original content is altered. The rd-signed-digest meta tag has the encrypted message digest value, encrypted using a private key in a PKI certificate. The rd-digest meta tag value can be verified using the rd-signed-digest meta tag value whether the rd-digest is altered. The rd-signed-digest meta tag value can be decrypted only using a matched public key in a PKI certificate. The public key and the PKI certificate information are stored in the Reliable Electronic Document Blockchain. The rd-address meta tag has the blockchain transaction address to find the public key and the certificate information to verify the Reliable Electronic Document. These meta tags are used by standard HTML < meta > tag specification. Therefore it provides compatibility and reusability compared to the Chained Document in the previous research [7]. The Reliable Electronic Document can be validated by the Document Validator in the RED converter if there is a request to validate the content integrity of the document.

3.3 Design of Reliable Electronic Document Archiver

The Reliable Electronic Document metadata is stored in the Reliable Electronic Document Blockchain, as shown in Fig. 4. The metadata is stored in the data area of the blockchain transaction in JSON format. The public key and the PKI certificate are stored in the CERTIFICATE key to verify the content integrity later. The value of RED-CONTENT-DIGEST is the same as the rd-digest in the Reliable Electronic Document as it could be used for double verification.

4 Experimental Validation

We developed the prototype system to validate the Reliable Electronic Document System. We implemented the system using python script and used the PKI certificate of Let's Encrypt to sign the document. We prepared the sample electronic document

```
{  
'DOCUMENT-NAME': Document Name,  
'RED-CONTENT-DIGEST': Labeled Message Digest,  
'CERTIFICATE': Digital Signature Certificate,  
'TIMESTAMP': Reliable Electronic Document Creation Time,  
'CREATOR': Document Sender,  
'RECIPIENT': Document Recipient,  
'UUID':uuid  
}
```

Fig. 4 The definition of the metadata format in the reliable electronic document

and used the prototype system to generate the Reliable Electronic Document, as shown in Fig. 5. And the PKI certificate, including the public key, is stored in the Reliable Electronic Document Blockchain, as shown in Fig. 6. We confirmed that the Reliable Electronic Document has the same level of security as the PKI certificate, and it expects to take more than 4006×10^{12} years to alter the content in an unauthorized manner as it used the RSA-2048 bit key. Furthermore, the metadata, including a public key of the PKI certificate, is stored in the blockchain network. The metadata has a non-fungible value as it has inherited the non-fungibility of the blockchain characteristic.

5 Conclusion

In this paper, we have designed the Reliable Electronic Document System based on HTML and Blockchain technology. Electronic document based on HTML is getting a more critical document format than PDF in a mobile environment, and document which needs to validate the content integrity getting to be delivered via the mobile channel. And the Reliable Electronic Document System inherits the characteristic of digital signing using a PKI certificate and the non-fungibility of blockchain. Therefore, we expect the Reliable Electronic Document System can be a middle-ware between a legacy system in an enterprise and a customer to provide a reliable electronic document. In future research, we will research this system in the cloud environment to confirm that this theory can work properly in a real environment.

5. Kim KJ, Lee MS (2022) Blockchain technology and the creation of trust: focusing on transparency, immutability and availability. *J Korea Soc Comput Inf* 27(3):79–90
6. Masinter L (1998) The “data” URL scheme.” RFC 2397
7. Hwang HC, Kim WJ (2022) Design of chained document html generation technique based on blockchain for trusted document communication. *Electronics* 11(7):1006

Multi-armed Bandit-Based Rate Control with Logarithmic Rates in CSMA/CA Wireless Networks



Soohyun Cho 

Abstract Herein, we propose to use the log values of the data transmission rates of modulation and coding scheme (MCS) levels for multi-armed bandit (MAB) algorithms that adjust the MCS levels of packets in carrier-sensing multiple access/collision avoidance (CSMA/CA)-based wireless networks. We argue that the utility of the data transmission rates of the MCS levels may not be proportional to their nominal values and suggest using their log values instead of using the data transmission rates directly for computing the expected throughputs in the MAB algorithms. For the demonstrations, we introduce two MAB algorithms that adopt log rates. The algorithms also use a sliding window to cope with non-stationary wireless channel environments. Using the simulation results obtained from the ns-3 network simulator, we show that the proposed MAB algorithms outperform the MAB algorithms that do not use log rates in rapidly changing wireless channel environments.

Keywords Multi-armed bandit · CSMA/CA · Rate adaptation

1 Introduction

Algorithms to solve multi-armed bandit (MAB) problems [1] (referred to as MAB algorithms) have received tremendous interest from both academia and industry because they can be applied to solve many optimization problems in vast areas, including commercial applications such as online advertisements and finance. Among the MAB algorithms, the Kullback–Leibler upper confidence bound (KL-UCB) [2] and Thompson Sampling [3] are some of the most popular algorithms owing to their performance in dealing with the trade-off between the exploitation and exploration of knowledge.

Recently, researchers have introduced MAB algorithms to improve the performance of communication networks, such as carrier-sensing multiple access/collision

S. Cho (✉)

Hongik University, 94 Wausan-Ro, Mapo-Gu, Seoul 04066, Korea

e-mail: cho.soohyun@hongik.ac.kr

avoidance (CSMA/CA)-based wireless communication networks (e.g., IEEE 802.11 standards). In [4] and [5], the authors proposed MAB algorithms to control the data transmission rates of wireless nodes by adjusting the modulation and coding scheme (MCS) levels that incorporate the MCS indices and number of communication links if multiple-input multi-output (MIMO) is available.

As an example, for a system with K possible MCS levels, i.e., MCS level $i \in \{0, \dots, K-1\}$, the modified Thompson Sampling algorithm (MTS) in [5] uses the (physical) data transmission rate r_i of each MCS level i and its success probability estimate $p_i(t) \in [0, 1]$ that is obtained from the history of transmission experiences up to time t , to select an i that maximizes the expected throughput $r_i \bullet p_i(t)$ as follows:

$$i = \operatorname{argmax}_i(r_i p_i(t)) \quad (1)$$

The data transmission rate r_i can be, e.g., one of $\{6, 9, 12, 18, 24, 36, 48, 54\}$ Mbps in IEEE 802.11g [6] if we consider only the extended rate physical-orthogonal frequency-division multiplexing (ERP-OFDM) scheme. In addition, to cope with the inherently non-stationary channel conditions in wireless communication environments, researchers have adopted some change-detection mechanisms, such as the sliding window or discounting of rewards [7].

Although packets can be transmitted faster using higher MCS levels, the strong preference to use higher MCS levels by directly using the data transmission rates as in Eq. (1) can lead to low utilization of network resources because higher MCS levels require higher signal-to-interference and noise ratio (SINR) at the receivers of the packets. Furthermore, a data sending node in a CSMA/CA-based wireless network should wait for the arrival of acknowledgement (ACK) messages from its receiver or timeout events before transmitting subsequent packets. In addition, the data sending node needs to wait for the end of the random backoff time for every data packet to avoid possible collisions. Therefore, we argue that the utility of the data transmission rates is not proportional to their nominal values and propose using the log (base e) values of the rates for computing the expected throughputs as follows:

$$i = \operatorname{argmax}_i(\log(r_i) p_i(t)) \quad (2)$$

Here, we assume that $r_i > 1$. To demonstrate the effectiveness of the proposed method, we introduce two transmission rate control algorithms based on KL-UCB and Thompson Sampling with a sliding window in time. We used the event-driven network simulator ns-3 [8] to evaluate their performance in rapidly changing channel environments. The experimental results show that the use of log values of the rates significantly improves the system performance for both types of MAB algorithms.

2 Multi-armed Bandit Algorithms Using Logarithmic Rates

2.1 Thompson Sampling Using Log Rate (TS-LogR)

In this subsection, we introduce an MAB algorithm for transmission rate control in CSMA/CA wireless networks based on Thompson Sampling. We call it Thompson Sample with log rate (TS-logR). The difference between TS-logR and MTS is that TS-logR uses the log values of the rates when computing the expected throughput. In addition, TS-logR explicitly describes the use of the sliding window in time.

Algorithm 1 shows the pseudocode of TS-logR. In Line 1, α_i and β_i are lists used to keep track of the success and failure events of each MCS level i , respectively. They are initialized as empty lists. In line 4, the success probability estimate $p_i(t)$ of each MCS level i is sampled from the beta distribution using the histories of success and failure events during a sliding window time τ . The two arguments used by the beta distribution, $\text{sum}(\alpha_i \text{ during } [\max(0, t-\tau), t])$ and $\text{sum}(\beta_i \text{ during } [\max(0, t-\tau), t])$, indicate that we only use the numbers of success and failure events, respectively, that occurred between $t-\tau$ and t , i.e., $[t-\tau, t)$, if time t is greater than τ .

Algorithm 1: TS-logR algorithm

```

1: for each  $i \in \{0, \dots, K-1\}$ , set  $\alpha_i = []$ ,  $\beta_i = []$ .
2: for each transmission at time  $t$ , do:
3:   for each  $i$ , do:
4:     Sample  $p_i(t)$  from Beta( $\text{sum}(\alpha_i \text{ during } [\max(0, t-\tau), t]) + 1$ ,  $\text{sum}(\beta_i \text{ during } [\max(0, t-\tau), t]) + 1$ ).
5:   end for
6:   Use  $i = \text{argmax}_i(\log(r_i) \bullet p_i(t))$  for transmission and wait for the outcome.
7:   if success, append 1 to  $\alpha_i$ , else append 1 to  $\beta_i$ .
8:   Record the time at which the outcome is observed.
9: end for

```

Line 6 shows how the expected throughput of each MCS level is computed using the logarithms of r_i and $p_i(t)$. Line 7 shows how we track the success and failure events. If a data sender received an ACK message before a timeout period, the event was considered a success (i.e., the reward was one) and 1 was appended to α_i . If the ACK was not received in the given timeout period, the event was considered a failure (i.e., the reward was zero). It was recorded by appending 1 to β_i . Finally, the time of each event was recorded on a separate list (not shown here) to be used for the sliding window. If rate r_i was used without the logarithm in line 6, we called the algorithm TS-R for performance comparison purposes.

2.2 KL-UCB Using Log Rate (KL-UCB-LogR)

In this subsection, we introduce another MAB algorithm for transmission rate control in CSMA/CA wireless networks based on KL-UCB, referred to as KL-UCB-LogR. Algorithm 2 presents the pseudocode for the algorithm. In line 1, for each MCS level i , an empty list R_i is prepared to track the history of rewards that occur using i . After K initial steps, for each MCS level i , the sum of the rewards (S_i) and number of transmissions using i (N_i) during the sliding window time τ were computed. They were then used to estimate the success probability $p_i(t)$ in line 10. $d(\bullet, \bullet)$ in line 11 is a function that computes the Bernoulli KL divergence. Line 12 indicates that the expected throughput of each i is computed using the log value of its rate. If the rate was used directly without the logarithm, we called the algorithm KL-UCB-R for performance comparison.

Algorithm 2: KL-UCB-logR algorithm

```

1: for each  $i \in \{0, \dots, K-1\}$ , set  $R_i = []$ . step = 0.
2: for each transmission at time  $t$ , do:
3: if step  $\leq K-1$ :
4: Use  $i = \text{step}$  for transmission and wait for the outcome.
5: else:
6:  $N =$  number of all transmissions during  $[\max(0, t-\tau), t)$ .
7: for each  $i$ , do:
8:  $S_i =$  sum ( $R_i$  during  $[\max(0, t-\tau), t)$ .
9:  $N_i =$  number of the transmissions using  $i$  during  $[\max(0, t-\tau), t)$ .
10:  $p_i(t) = \max\{q \in [0, 1]: d(S_i/N_i, q) \leq (\log(N) + c \bullet \log(\log(N)))/N_i\}$ 
11: end for
12: Use  $i = \text{argmax}_i(\log(r_i) \bullet p_i(t))$  for transmission and wait for the outcome.
13: if success, append 1 to  $R_i$ , else append 0 to  $R_i$ .
14: Record the time at which the outcome is observed. step = step + 1.
15: end for

```

3 Experiments

3.1 Simulation Scenarios

For simulations, we used the IEEE 802.11g implementation in ns-3 (version 3.35) as an example of CSMA/CA-based wireless networks. We used only the ERP-OFDM scheme in the 802.11g (single link) for simplicity. Thus, there are eight possible MCS levels, i.e., $K = 8$. The data transmission rate can be any one of $\{6, 9, 12, 18, 24, 36, 48, 54\}$ Mbps. In all the simulations, two wireless nodes operate in the distributed coordination function (DCF) mode [9]. One node (sender) sends a constant bit-rate UDP traffic of 54 Mbps to the other node (receiver). During the simulation time of 10 s, the sender moved away from or toward the receiver at a speed of 6 m/s. When

Table 1 Simulation parameters

Parameter	Value	Parameter	Value
Carrier frequency	2.412 GHz	Noise floor	~ -94 dBm
Channel bandwidth	20 MHz	Clear channel threshold	-82 dBm
Nodes transmit power	100 mW	Data packet size	1500 bytes
Rx sensitivity	-101 dBm	Device queue size	500 packets
Slot time	20 μ s	Guard interval	800 ns

the sender moved away from the fixed receiver, the distance between the two nodes changed from 5 to 65 m. In the opposite direction, the distance started at 65 m and ended at 5 m.

For the path loss of the signal, we used the urban microcell non-line-of-sight model in [10]. Thus, for a given distance $d > 1$ m, the path loss (in dB) was computed as $40.198 + 38 \bullet \log_{10}(d)$. For the error probabilities of the arriving packets, *NistNetErrorRate* model in the ns-3 was used. The request-to-send/clear-to-send (RTS/CTS) scheme was not used. The sliding window time was set to 100 ms. The configuration parameter c in Algorithm 2 was set to one. The noise figure was set to 7 dB. Table 1 provides more details regarding the simulation environment.

3.2 Simulation Results

In Fig. 1, we show the changes in the application throughputs achieved by the receiver from the four MAB algorithms. In Fig. 1a, TS-logR shows better performance than TS-R, particularly when the sender has moved far away from the receiver. This is because TS-logR selects lower MCS levels more accurately when adequate. KL-UCB-logR also shows better performance than KL-UCB-R. The latter shows low performance from approximately the middle of the simulation. In Fig. 1b, the results show that TS-logR and KL-UCB-logR achieve higher throughputs than TS-R and KL-UCB-R, respectively, particularly when the sender is far from the receiver. The results in Fig. 1a, b also show that both TS-logR and KL-UCB-logR achieve high throughputs when the distance is small.

In Fig. 2, we compare the total application throughputs achieved by the four MAB algorithms during the 10 s simulation time. We repeated the simulations ten times with different seeds. The results in Fig. 2 show their averages. In the figure, we also show the simulation results using Minstrel [11] as a reference. The results clearly show that by adopting the log values of the rates, the MAB algorithms perform better in rapidly changing channel environments. Additionally, we can observe that TS-logR performs better than KL-UCB-logR and Minstrel in both simulation scenarios.

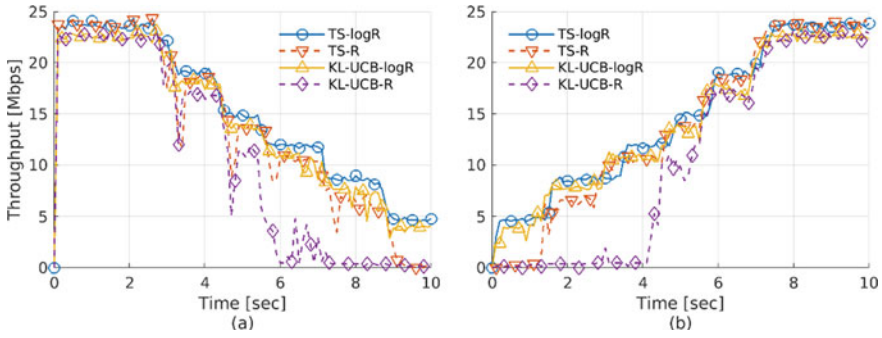


Fig. 1 Changes of the application throughputs (0.1 s average) achieved by the receiver, **a** the sender is moving away from the receiver, **b** the sender is moving toward the receiver

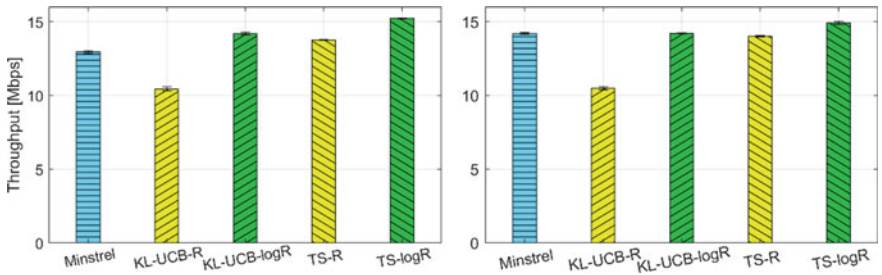


Fig. 2 Averages of total application throughputs from 10 repetitions, (left) the sender is moving away from the receiver, (right) the sender is moving toward the receiver

4 Conclusion

In this study, we aimed to control the transmission rates of data packets more accurately in CSMA/CA-based wireless networks using MAB algorithms. We proposed the use of log values of the transmission rates of MCS levels instead of using them directly in the computation of the expected throughput estimations. To evaluate the effectiveness of the proposed method, we introduced two MAB algorithms by adopting the log rates and a sliding window in time. The simulation results demonstrated that they outperform the MAB algorithms that do not adopt log rates when wireless channel environments are rapidly changing. We expect that log rates can also be applied to improve the performance of MAB algorithms for various types of wireless communications.

Acknowledgements This work was supported by 2020 Hongik University Research Fund.

References

1. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction, 2nd edn. Bradford Book, Cambridge
2. Garivier A, Cappé O (2011) The KL-UCB algorithm for bounded stochastic bandits and beyond. In: Proceedings of conference on learning theory, pp 359–376. JMLR
3. Agrawal S, Goyal N (2013) Further optimal regret bounds for Thompson sampling. In: Proceedings of international conference on artificial intelligence and statistics, pp 99–107
4. Combes R, Ok J, Proutiere A, Yun D, Yi Y (2019) Optimal rate sampling in 802.11 systems: theory, design, and implementation. *IEEE Trans Mobile Comput* 18(5):1145–1158
5. Gupta H, Eryilmaz A, Srikant R (2018) Low-complexity, low-regret link rate selection in rapidly-varying wireless channels. In: Proceedings of IEEE conference on computer communications. IEEE, pp 540–548
6. IEEE (2003) IEEE 802.11g-2003: further higher data rate extension in the 2.4 GHz band. IEEE
7. Garivier A, Moulines E (2011) On upper-confidence bound policies for non-stationary bandit problems. In: Proceedings of ALT 2011, pp 174–188, Springer, Berlin, Heidelberg
8. ns-3 Homepage. <https://www.nsnam.org>. Last accessed 29 Aug 2022
9. IEEE Computer Society (2016) IEEE Standard 802.11–2016 Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE computer society, Washington DC, USA
10. IEEE Computer Society (2007) IEEE Standard 802.20-PD-08r1: IEEE 802.20 channel models document. IEEE computer society, Washington DC, USA
11. Rate Adaptation for 802.11 Wireless Networks: Minstrel. <http://blog.cerowrt.org/papers/minstrel-sigcomm-final.pdf>. Last accessed 29 Aug 2022

The Design and Implementation of Medicine Recognition System Using the TensorFlow Lite



Ji un Park, Hee Jun Moon, and Dong Hyun Kim

Abstract People are hard to know the type, effects and side effects of the medicine which they purchase since the explanations of the medicine exploit the medical terms and have the difficulty to find. In this paper, we propose the medicine recognition system using a camera of a smartphone and the TensorFlow lite. If a user only shoots a photo of the medicine, the system recognizes the medicine by using the MobileNet V2 and displays candidates of the medicine on the screen. The benefits of the system are easy to search for the medicine conveniently.

Keywords Medicine recognition · MobileNet · Faster R-CNN

1 Introduction

The domestic pharmaceutical market is growing steadily. In 2019, the domestic pharmaceutical market is KRW 24.31 trillion, up 5.2% from 2018, which is 23.175 trillion won. The officinal pharmaceutical market sold at convenience stores has also tripled in size in seven years since introduced in 2012. As the market grows, more and more people are purchasing officinal medicines and prescription-based medicines which should be purchased with a prescription from a hospital.

In the case of the normal people, there is a limit to distinguish the type of pill if they lose their medicine bags or their pharmacist's explanation. Since the domestic pharmaceutical market is growing steadily every year, more and more patients are purchasing prescription-based medicines by using prescriptions from hospitals. However, in order to get more information about medicines, information can be obtained only by the description of the attached document included in the medicines. However, the documents exploit the medical terms and complex explanations, it is hard to understand the effects and the side effects correctly.

J. un Park · H. J. Moon · D. H. Kim (✉)

Department of Software, Dongseo University, Jurye-Ro 47, Busan 47011, Republic of Korea

e-mail: pusrover@dongseo.ac.kr

Second, it is not easy to get information of medicines. If the patient has not the expertise of medicines, he/she have to go to the website and search for more detailed information about the pills. However, it is cumbersome process to search for names or enter detailed texts such as letters, dividing lines and patterns to search. If there is an input mistake, the wrong medicine information may be drained. As a result, the patient becomes dependent on his experiences, and the lack of correct information on medicines leads to misuse and abuse, which raises fatal side effects to health.

In [1], the pill classification system using the TensorFlow is proposed. Gyeongin [2] compares of the performance three kinds of deep learning model: YOLO, Faster R-CNN and RetinaNet for the classification of pills. In [3], the license plate of a vehicle recognition system using SSD-MobileNet and ResNet is presented. Seon-Woo [4] evaluates the performance of the in-depth neural networks using double pruning method.

In this paper, we propose the real-time medicine recognition and identification system using the Tensorflow lite on the Android. To recognize a pill, the system exploits the MobileNet V2 which is the light-weighted deep learning algorithm performed on a mobile device. To do this, the system shoots the pill and transforms the photo to the tflite format in order to recognize the drug.

In Sect. 2, related works are presented and the design of proposed recognition system is described in Sect. 3. The implementation result is presented in Sects. 4 and 5 describes the conclusion.

2 Related Works

Soo-Hyoung et al. [1] proposed the pill classification system using the TensorFlow based on image processing. They extracted the part of a pill from a photo using the OpenCV first. To increase the correctness, the noise of the image is eliminated by the PyrMeanShiftFiltering filter. The only part of the pill was extracted by using the edge of the pill from the selected part. From the extracted image of the pill, the texts of the pill were drained. Based the edge of the pill and the texts, the CNN deep learning model recognized what the pill is. In [2], three kinds of deep learning model, YOLO, Faster R-CNN and RetinaNet, were compared to classify pills. The dataset for the experiment was composed of 100 classes, which had 100 images per a class. The evaluation result showed the sensitivity of the YOLO was 91.05%, that of the Faster R-CNN was 99.6% and the RetinaNet was 98.31%. The Faster R-CNN showed the best performance.

Woonki [3] presented the license plate of a vehicle recognition system using SSD-MobileNet and ResNet. The plate was recognized by following 3 steps, license plate detection, character segmentation and character recognition. To detect the license plate, the SSD-MobileNet was exploited and the ResNet for character area segmentation. In [4], the performance of the in-depth neural networks was evaluated using double pruning method. To improve the speed without compromising the learning

accuracy, they reduced the parameters that are not important to the existing learning. The evaluation result showed the MobileNet V3 has the best performance.

3 The Design of Recognition System

3.1 Drug Recognition Model

To identify a medicine image, we exploit the MobileNetSSD. The SSD stands for Single Shot Detector which is capable of training and detecting a single photo without transforming it. It is an algorithm that balances speed and accuracy when detecting objects. Multi-Scale Feature maps are used to detect objects of different sizes with only one picture. By making the feature map in multiple sizes, the detection of small objects in the large map and the detection of large objects in the small map are possible. This method diminishes the position estimation and resampling of the input image, while yielding high-accuracy results.

We also the MobileNet V2 to train the recognition of the extracted medicine image. The MobileNetV2 shows better performance than the existing MobileNet and ShuffleNet while dramatically reducing Params, MAdds, and operation speed. In addition, it shows excellent performance in detection and segmentation with a small amount of parameters. The method of mapping characteristics such as size and shape of drugs to recognize a pill has a disadvantage in that the failure rate of matching increases when the characteristics of tablets are similar. For a higher recognition rate, we exploit a method of recognizing data labeling based on its own shape and texts based on the overall image of the drug. The color of the pill is not used because accurate results cannot be obtained due to several variations depending on the shooting location, weather, color of light and, etc.

3.2 System Concept and Architecture

Figure 1 shows a conceptual diagram of a real-time medicine recognition and identification system using a camera. The user takes a picture of the medicine which he or she wants to know through the Android camera. The tflite type model recognizes and identifies the pill image taken in real time. The recognized result is shown the candidate list of the pills and the recognition rate on the screen being photographed together.

Figure 2 shows the architecture and modules of the system proposed in this paper. The camera module performs the function of photographing medicines using the android rear camera in real time. The captured values are passed to the TensorFlow Lite module for the recognition. The TensorFlow Lite module identifies the value of the image received from the camera and recognizes it using the built-in

Fig. 1 System concept

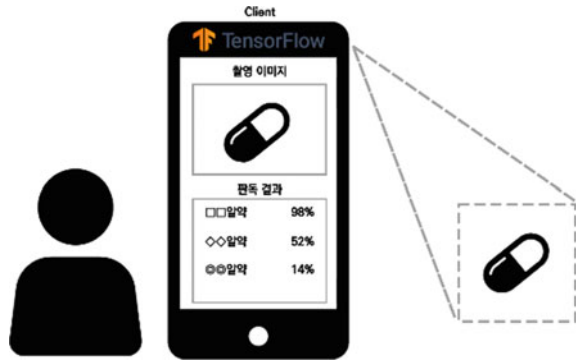
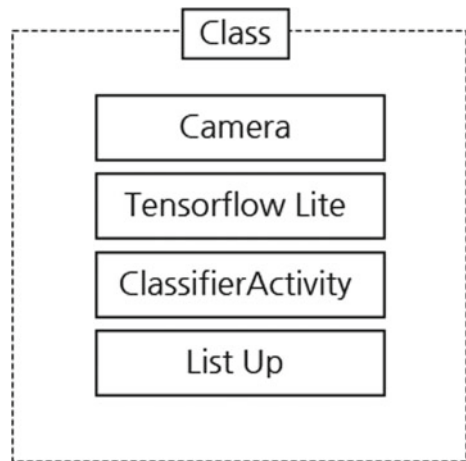


Fig. 2 System architecture



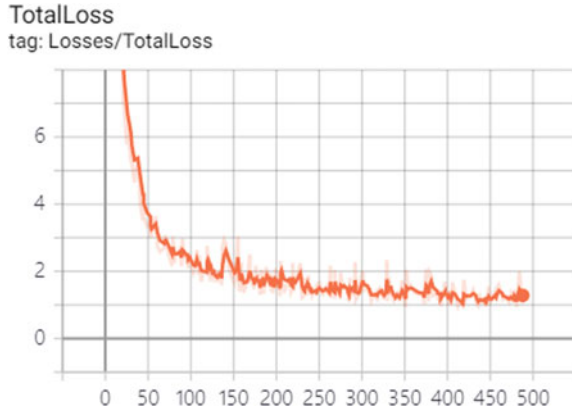
lightweight learning model to indicate which the medicine is. The List Up module displays medicines recognized through images as three user-defined list according to the recognition rate. The ClassifierActivity module integrates three modules as the Android main module.

4 Implementation

For the implementation, the deep learning server was configured with NVIDIA GeForce GTX 1060 graphics processing unit, Intel Core i5-7300HQ CPU and 8G RAM. Also, the operating system Windows 10 Home and Python 3.6.5 version and TensorFlow 1.15.0 framework were used. The client was implemented in Galaxy Note20 5G and took the pill image used for the deep learning.

MobileNet V2, a pre-trained model, was exploited as the best CNN-based feature extractor for the Faster R-CNN model for medicine recognition and identification.

Fig. 3 Total loss of training



In addition, 100 images of medicines with 300×300 size, which are collected data for learning, were prepared for each type. Thus, a total of 600 images were used for training. The dataset was randomly divided into 70 and 30% for learning and evaluation. Thus, 420 images were used for training and 180 images for evaluation. Since the photographed image may contain several pills, it was manually selected to determine whether the photographed medicine image was appropriate as the data area of the medicine for training.

Figure 3 shows the total loss of the model used as a feature extractor of Faster R-CNN according to the number of training times when learning medicine images. Using MobileNet V2, a tendency to converge was shown quickly as a result of 500 iterative learning progress.

Figure 4 shows the recognition and accuracy rate results of medicines. When a user takes a picture of a medicine through the android camera, it lists up three medicine candidates according to the accuracy. When the texts of the medicine are clearly detected, the reliability is about 98% to 100%.

5 Conclusion

When a prescription was lost, a patient is hard to know the type of pills he/she takes. Also, since the patient is not usually expertise, he has the problem to acknowledge the effects or side effects of the medicine. To solve these problems, we proposed and implemented the medicine recognition system using a camera of a smartphone. The proposed system identified and recognized the medicine the user shoots in real time using the MobileNet V2. The MobileNet V2 is the deep learning module based on the faster R-CNN model. The 3 candidates recognized as the results are listed on the camera screen together. The further research is to improve the reliability of the recognition when the images of the medicine are out of focus or the texts of the medicine are not clear.

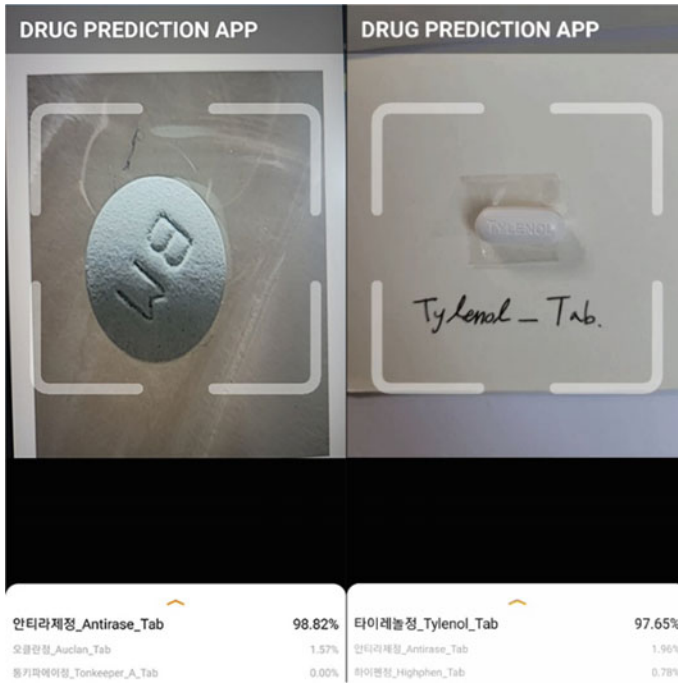


Fig. 4 Result of medicine recognition

Acknowledgements This research was supported by the Ministry of Science, ICT (MIST), Korea, under the National Program for Excellence in SW, supervised by the Institute of Information and communications Technology Planning and Evaluation (IITP) in 2022” (2019-0-01817).

References

1. Soo-Hyoung J, Jin-Goo K, Jung-Hoon K, Sung-Jun L (2019) A study on tensorflow based image processing: focusing by pill classification. In: Proceeding on spring conference of the korea information processing society 2019 26(1):559–561, KIPS, Jeju
2. Gyeongyin Y, Youngjae K, Seongtae K, Hyeon K, Kwanggi K (2019) Comparison and verification of deep learning models for automatic recognition of pills. J Korea Multimed Soc 22(3):349–356, Korea Multimedia Society, Korea
3. Woonki K, Fatemeh D, Seongwon C (2020) Vehicle license plate recognition system using SSD-Mobilenet and ResNet for mobile device. Smart Media J 9(2):92–98, Korea Smart Media Society, Korea
4. Seon-Woo L, Ho-Jun Y, Seung-Yeon O, Mun-Hyung L, Jang-Woo K (2020) Application and performance analysis of double pruning method for deep neural networks. J Converge Inf Technol 10(8):23–34, Convergence Society for SMB, Korea

Development of Ubiquitous Flip Learning Model Based on Backward Design



Kil Hong Joo  and Nam Hun Park 

Abstract Flip learning is an educational model in which learners watch online lectures at home and perform tasks in the classroom by reversing the traditional teaching method of performing tasks at home after class activities in the classroom. Backward is an understanding-oriented curriculum, and it is a curriculum model that allows students to develop a permanent understanding by first designing an evaluation that was at the last stage in the existing curriculum design. The purpose of this study is to analyze the effectiveness of a backward design-based flip learning class model by developing and applying it in a ubiquitous educational environment. In order to examine the effect of this study, the experimental group was subjected to ubiquitous backward design-based flip learning, and the results of effectiveness perception and satisfaction, learner competency and academic achievement were compared and analyzed through experiments.

Keywords Ubiquitous · Flip learning · Backward design

1 Introduction

In the era of the 4th industrial revolution represented by artificial intelligence and big data, we are living in an era of rapid digital transformation due to the development of various software and digital technologies. In order to survive in such a rapidly changing digital environment, a learning attitude is required to create tangible or

K. H. Joo

Department of Computer Education, Gyeongin National University of Education, 155, Sammak-Ro, Manan-Gu, Gyeonggi-Do, Anyang-Si, Republic of Korea

e-mail: khjoo@ginue.ac.kr

N. H. Park (✉)

Department of AI Convergence, Anyang University, 22, Samdeok-Ro 37Beon-Gil, Manan-Gu, Gyeonggi-Do, Anyang-Si, Republic of Korea

e-mail: nmhnpark@anyang.ac.kr

intangible values by successfully acquiring and continuously utilizing knowledge that grows as fast as the speed of thought.

Recently, many teachers are applying flip learning or student-participation-centered classes in their classrooms in order to induce students to become the subject of classes and conduct live classes [1, 2]. What they have in common is that they participate as subjects in the class with problems that students can be interested in and interested in.

In line with this trend, various teaching–learning methods suitable for student participation and problem-solving skills have been introduced recently. Among them, flip-learning, which can utilize videos in the learning process by reflecting students’ preferred media tendencies, is a very effective method [3, 4]. The focus is on solving related problems by examining the contents learned through the pre-learning task and applying the contents. In this process, students take part in solving problems. In addition, it is possible to have new types of learning opportunities through various forms of cooperative learning. In this case, the teacher acts as a facilitator to help the learner and helps the learner in the learning process.

In this study, flipped learning based on the backward design model is applied to improve students’ self-directed ability and analyze the learning effect. The learning effect of students is analyzed to find out the effectiveness and to derive implications.

2 Development of Ubiquitous Flip Learning Teaching and Learning Model Based on Backward Design

The stages of the backward design-based flip learning teaching and learning model go through the stages of analysis, design, understanding, pre-learning evaluation, problem solving, and post-reflection. Focusing on the goal of evaluation, which is the core of the backward curriculum, analysis and design are made, and based on this, flip learning is applied to pursue maximization of learning (Fig. 1).

In the first stage, review whether the flipped learning teaching and learning method is suitable for achieving the learning goal of the subject. At this time, it is necessary to analyze whether there are any learning materials to provide for prior activities. The reason for analyzing the subject is to use it as a basic data for the instructor to determine whether flip learning is appropriate in consideration of the overall subject matter.

In order to operate flip learning, it is necessary to analyze the educational environment. This is because flip learning is an educational method that utilizes lecture contents based on Information and Communications Technologies (ICT) [5, 6]. In order to smoothly operate the activities in class, various environmental requirements must be supported, and it is necessary for the instructor to closely check the inspection and preparations for the operation of flip learning. Therefore, the educational environment analysis includes classroom environment analysis and teaching/learning system environment analysis.

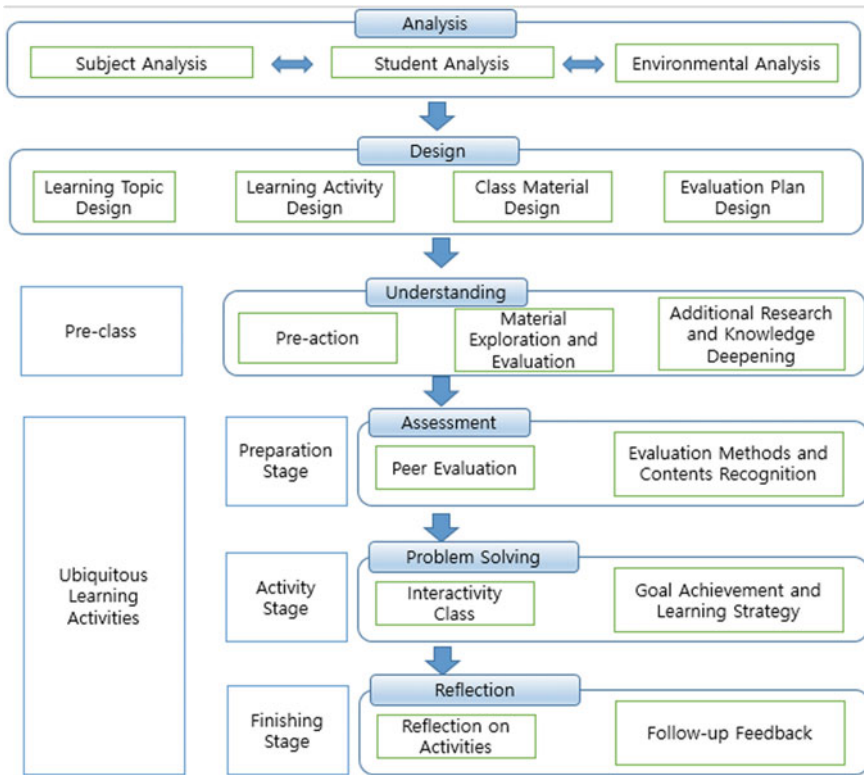


Fig. 1 Backward design-based ubiquitous flip learning model

In the second stage, learning topics, learning activities, instructional materials, and evaluations should be designed based on the backward design model. At this time, in order to increase students’ lasting understanding of the core knowledge of the subject, topics are selected and evaluation is devised, and students’ learning activities and instructional materials are designed in the designed evaluation.

In the third stage, the understanding stage corresponds to the pre-learning stage of the flip learning. Based on the data provided by the teacher in the pre-learning stage, it induces students to learn a wide range of things. The key here is for students to explore the materials in anticipation of evaluation and learning activities based on the materials provided by the teacher. In this stage, self-directed learning occurs, and since it is difficult for students to learn alone, the teacher must set up an environment so that students can interact with each other and explore knowledge. In order to overcome the shortcomings of the backward curriculum, the acquisition of minimal limited knowledge, students should be guided to investigate and explore knowledge in a wide range of materials other than those presented. Through this stage of understanding, students can step on to a lasting understanding of core knowledge.

In the fourth stage, the pre-learning assessment consists of peer evaluation discussion and familiarization with evaluation methods and contents. In the stage of discussing while checking the evaluation expected by the students and the evaluation presented by the teacher, it is necessary to check how much the students approached the evaluation contents and methods through prior learning. If the evaluation expected by the students is different from the content presented by the teacher, feedback is given to make corrections, but the students themselves can give feedback. Through this course, students can clearly understand the next step, the problem-solving step.

In the fifth stage, class activities should be operated to reach the evaluation goal of important content drawn based on the core concepts and principles of the discipline. The learning experience of students must incorporate the elements to explain, interpret, apply, have a point of view, empathize, and have self-knowledge so that the aspect of understanding is possible. In other words, the teacher should focus on 'confirming the desired results'. To this end, keep in mind the template form of the backward design model.

In the sixth stage, learners reflect on what they have learned and felt in class, where they can apply their learning, and how much they have contributed to cooperative learning. It focuses on improving self-directed learning ability, such as will, ability, and attitude to practice and take responsibility for learning in future learning through peer and teacher feedback on educational activity content and evaluation. In other words, mastery goal orientation focuses on learning and mastering based on one's own standards and self-development, improving and developing competency, and striving to achieve challenging tasks.

3 Teaching and Learning Model Application and Analysis

The subjects of this study were 296 students in grades 3–6 of S Elementary School in Gyeonggi-do, who had no experience in software education as in Table 1. The results of recognition of effectiveness and satisfaction with learning using the backward design-based flip learning teaching–learning model are shown in Table 2.

In Fig. 2, a group room was created so that data could be organized while exchanging opinions with each other, and data was shared with each group so that knowledge could be expanded. In Fig. 3, as a result of analyzing the perception of effectiveness for digital textbooks, almost all areas from codes A-1 to A-10 showed

Table 1 Composition of personnel by year of study group

Total number of students		296	100.0 (%)
Grade	Third grade	80	27.03
	Fourth grade	82	27.70
	Fifth grade	66	22.30
	Sixth grade	68	22.97

Table 2 Effectiveness results

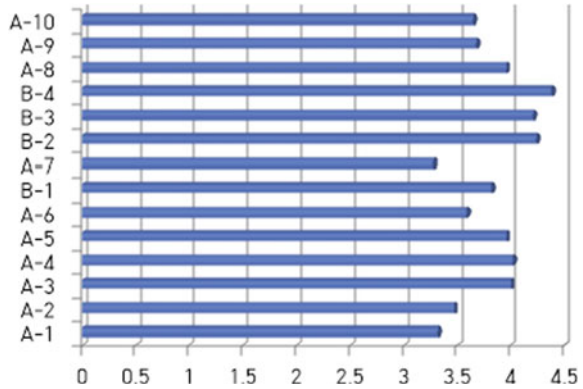
Items (max: 5points)	Code	Points
Self-learning attitude is nurtured	A-1	3.32
Your grades and academic abilities have improved	A-2	3.47
I can understand the class content more easily	A-3	4.00
I learned a lot of new things outside of class content using various materials	A-4	4.02
Increased interest in class	A-5	3.96
Increased concentration in class	A-6	3.59
There are more cases of using smart devices to do things other than learning (reverse coding)	B-1	3.82
The proportion of using smart devices for studying rather than playing games has increased	A-7	3.28
My eyesight got worse (reverse coding)	B-2	4.24
Bad posture (reverse coding)	B-3	4.21
I seem to be addicted to smart devices (reverse coding)	B-4	4.38
I think it is helpful for learning	A-8	3.96
I think the ability to use smart devices will be necessary for future life	A-9	3.68
I think it will help in the future	A-10	3.65

a positive response of 3 points or more. In particular, they showed a positive reaction to the specific effects related to the class, and it was found that they recognized the high effectiveness in the items of ‘understanding class content, interest in class, improvement of concentration, helps in learning’. Satisfaction results for the use of digital textbooks also showed a positive response of more than 3 points, and Code A was relatively low at 3.32, but the self-learning attitude also came out as a positive result value. It can be judged that the motivation for learning is interlocked because the pre-learning by prior learning and the clarity of the evaluation in this class are linked in the teaching–learning model stage. However, as a result of analyzing the reverse coding codes B-1 ~ B-4, it was found that the eyesight and posture of the students deteriorated, and they could become addicted to smart devices. Therefore, it can be seen that guidance and learning are necessary for the correct use of smart devices.



Fig. 2 Advertising production activities using smart devices

Fig. 3 The perception of effectiveness for digital textbooks



4 Conclusion

First, a forward-looking ubiquitous educational environment was supported through the backward design-based flip learning teaching and learning, and it was possible to develop core competencies for creative convergence that collect and share information using digital media to solve problems independently. Second, through a curriculum design centered on student understanding, students engage in interesting classes and learn outside the constraints of time and space, and teachers progressively improve teaching methods and provide effective learner-tailored education to students. I came to the conclusion that it could be done.

In addition, through the development, application, and generalization of the backward design-based flip learning teaching–learning model, self-directed learning in which the individual learner becomes the subject is possible by first performing an understanding-oriented performance task has been shown to be feasible. Second, it was confirmed that the core competency of creative convergence was enhanced in the process of students sharing information through digital textbooks and Widurang (SNS) and solving problems by collaborating with each other offline. Finally, the effectiveness of the national policy, digital textbooks, was verified, and the teaching model and application strategy were generalized and provided to parents and the local community, thereby demonstrating that trust in schools and public education can be increased.

References

1. Zainuddin Z et al (2021) Synchronous online flip learning with formative gamification quiz: instruction during COVID-19. *Interact Technol Smart Educ*, 302–306
2. Parra-González, Elena M et al (2020) Active and emerging methodologies for ubiquitous education: potentials of flipped learning and gamification. *Sustainability* 12(2):602

3. Capella JV (2020) Cloud services in higher education: improving flip teaching and u-learning environments. CIVINEDU 2020:470
4. Srilaphat E, Jantakoon T (2019) Ubiquitous flipped classroom instructional model with learning process of scientific to enhance problem-solving skills for higher education (UFC-PS Model). High Educ Stud 9(1):76–85
5. Lin C-J (2019) An online peer assessment approach to supporting mind-mapping flipped learning activities for college English writing courses. J Comput Educ 6(3):385–415
6. Mertala P (2020) Young children's perceptions of ubiquitous computing and the Internet of Things. Br J Educ Technol 51(1):84–102

KoBART-Based Hybrid Literacy Improvement System for Each Level for Efficient Reading



Soobin Cho  and Seungmin Park 

Abstract The digital literacy of 15 years-old students in Korea is the lowest among OECD countries. This is not a decrease in student's literacy, but an increase in the number of students who cannot read. In this paper, we proposed a system that identifies the number of user summary keywords, presents step-by-step summaries, and provides improvement to learn and grow. First, receives the original text, KoBART identifies the number of key keywords. After that, the user's summary is input, and the number of key keywords is determined by comparing it with the summary summarized by artificial intelligence. We present a summary of three levels (Levels 1, 2, 3) using KoBART and TextRANK algorithm. ROUGE demonstrates the reliability of the model by performing a performance measurement of the developed model. In addition, it compares the artificial intelligence summary with the user's summary to identify the user's levels. Finally, an AI summary is presented according to the level identified by the user summary. In other words, it induces interest in reading through step-by-step text summaries. It is possible to learn the ability to grasp the core of the text and to expect improvement in literacy and learning ability.

Keywords KoBART · TextRANK · ROUGE · Natural language processing · AI

1 Introduction

In the report “Program for International Student Assessment (PISA) 21st Century Reader: Literacy Development in the Digital World” released by the OECD in May 2021, Korea's 15-year-old students' digital literacy was the lowest among OECD countries [1]. According to the PISA results, Korea fell from No. 1 in 2006 to No. 9 in 2018. The phenomenon that children do not know difficult words also has different experiences between generations.

S. Cho · S. Park (✉)
Dongseo University, 47 Jurye-Ro, Sasang-Gu, Busan 47011, Korea
e-mail: sminpark@dongseo.ac.kr

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_88

653

The focus of the OECD results is not on the decline in literacy for all students, but on the increase in the number of students who cannot read. The problem is that the proportion of students who have difficulty reading the minimum information in their writings, let alone critical thinking, has increased. According to the “2019 National Reading Survey” released by the Ministry of Culture, Sports, and Tourism on the 11th, the annual reading rate of adults (October 1, 2018 to September 30, 2019) was 52.1%, and the reading volume was 6.1 books, the number decreased by 7.8% points and 2.2 books, respectively, compared to the 2017 survey.

A slow learner is an IQ of 71–84, which is partially impaired in the ability to adapt to standard intelligence tests, but the degree is not as severe as that seen in intellectual disabilities. It is not classified as an intellectual disability, it is not subject to legal special education, and parents do not think that special education is necessary, only thinking that he is a “slow and stuffy child”. However, in the educational field, many of the children suffer from poor learning, refusal to attend school, etc., and many of them are borderline intelligence [7, 8].

We created a system that uses artificial intelligence to judge users’ literacy and to support them by using a large language model to bridge the digital gap. In this paper, we developed a KoBART-based literacy improvement system for efficient reading. Section 2 introduces related research, Sect. 3 proposed methods, Sect. 4 experimental results, and Sect. 5 conclusions and future works.

2 Related Works

2.1 GPT-3

GPT-3, a “large language model” produced by OpenAI, an artificial intelligence research institute, is an algorithm using deep learning, which learns numerous texts and connects words and phrases to create texts. It is capable of generating complex sentences that appear to have been written by humans. In GPT-3, it can create texts that include cultural context. In the process of developing the GPT-3, almost the same method was used as the previous model, the GPT-2. The difference is that GPT-3 utilized a much larger neural network than GPT-2 and used much larger learning data. GPT-2 had 1.5 billion parameters GPT-3 used much more data when learning with 175 billion parameters. It turns out that training a larger model with more data can produce better overall results. OpenAI decided to increase the size of the model and the learning data size when creating the GPT-3, creating the largest of all language models released by then.

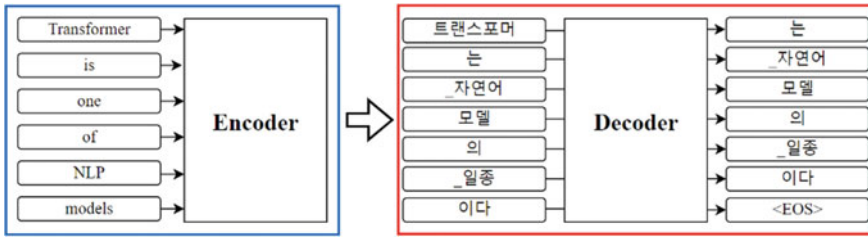


Fig. 1 KoBART: Korean encoder–decoder (Seq2Seq structure)

2.2 KoGPT

KoGPT is a Korean-specific version of Kakao Brain’s Open Application Interface (API) language model “GPT-3”. Overcome the existing insufficient Korean language performance, 6 billion parameters and 200 billion Korean language data were learned and constructed. “GPT-3” is a technology equipped with emotion analysis so that humans and AI can communicate based on natural language. “KoGPT” understands Korean dictionary and contextually and performs tasks such as judging positive and negative sentences, summarizing long sentences, predicting conclusions, and answering questions by understanding the context [3]. KoGPT has been trained with Korean text and is best suited for classifying, searching, summarizing, or generating these texts.

2.3 KoBART

Bidirectional and auto-regulatory transformers (BART) are learned in the form of an autoencoder that adds noise to a portion of the input text and restores it to the original text [4]. KoBART is a Korean encoder–decoder language model that uses the text infilling noise function used in the paper to learn more than 40 GB of Korean text. In BERT, each masked token is independently predicted, making it difficult to use for productive tasks, and GPT family models have limitations that it is difficult to learn full (two-way) contexts because words rely only on left-direction contexts. In this paper, we used a BART model that complements the shortcomings by combining the structural features of the BERT and GPT models (Fig. 1).

2.4 TextRank Algorithm

Graph-based ranking algorithms are essential methods for determining the importance of vertices within a graph based on global information recursively derived throughout the graph [5]. TextRank builds a word or sense graph and then selects

each keyword and key sentence using PageRank, a graph ranking algorithm. They are used to summarize a given set of documents.

TextRank creates a co-occurrence graph between words to select keywords. To select a key sentence, create a sentence similarity graph based on the similarity between the sentences. Then, each graph learns PageRank to calculate the ranking of each bar (word or sentence). Keywords and critical statements will be in the order of their high ranking. It extracts essential words from a document, which identifies the connection between words and words and assumes that the more words are related to many words, the more influential the word is in the document.

2.5 *Rouge*

Recall-oriented understudy for gisting evaluation (ROUGE) is a performance evaluation indicator of a text summary model. This is an indicator for evaluating the performance of natural language generation models such as automatic text summarization and machine translation. The performance score is calculated by comparing the summary or translation generated by the model with the reference pre-made by humans [6, 9].

3 Proposed Method

3.1 *Development Environment*

In order to implement an artificial intelligence system, IDE must be built, and the version of the development environment is fundamental. The configuration of the software is as follows. The software environments used in this paper are Python (3.7), Anaconda (4.9.2), Pytorch (1.10.0), Transformers (4.8.2), Streamlit (1.1.0) PyTorch-lightning (1.5.2), torchmetrics (0.6.0), and CUDA (11.1) were used. The specifications of the computer that ran this program are as follows. GPU (NVIDIA GeForce GTX 1050 Ti), CPU (Intel® Core™ i7-8750H), RAM (16.0 GB), and Window10 64bit operating system were used.

3.2 *KoBART-Based Hybrid Literacy Improvement System*

The overall configuration of the system is shown in the Fig. 2.

On the left side, there is a sidebar where you can select a level. If you click choose level, there are estimate, Levels 1, 2, and 3 that evaluate and recommend the user's levels. On the right side, the original text block is in which artificial intelligence

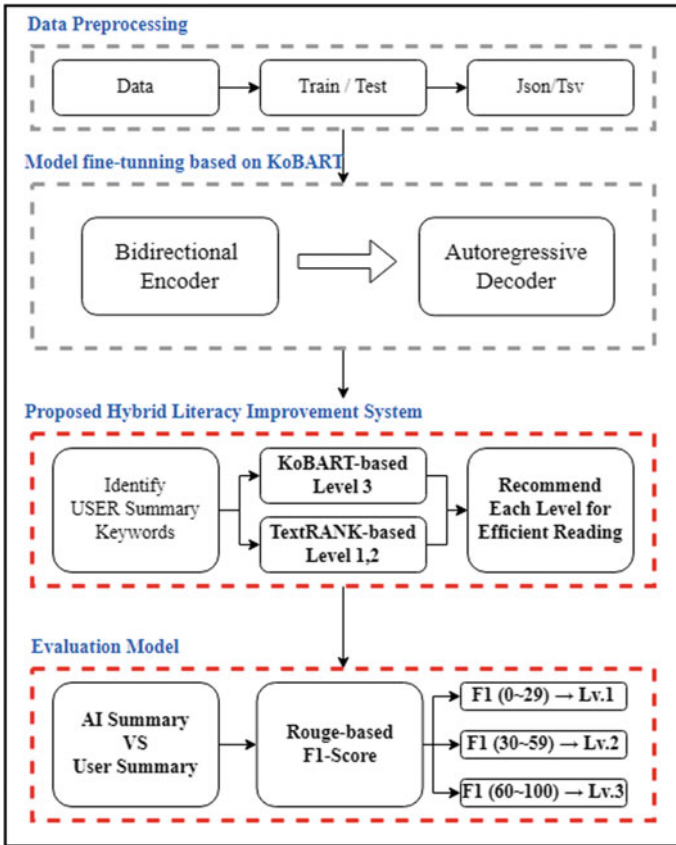


Fig. 2 Proposed literacy improvement system configuration diagram

receives the original text for reading and summarizing. The USER summary block is input to determine the current level of the user’s ability before recommending a level to the user. The final summarization is a block in which artificial intelligence detects the user’s literacy level and recommends levels.

The measurement method is as follows. Subjects are given a total of 5 min to read the original text and summarize it. First, put the original text in the original text block. Second, read the text for 5 min, and input a summary in the USER summary. Third, recommends a summary by dividing the user’s levels with an F1-score score.

In Level 3, KoBART was used, and in Levels 1 and 2, TextRANK was used. Level 3 utilizes KoBART, which is more complex and well-structured, because it requires identifying key words and simplifying. In Levels 1 and 2, TextRANK, which is relatively simple to calculate because it does not need to be summarized in detail as Level 3, was used.

Table 1 Level classification criteria

Level	Level 1	Level 2	Level 3
F1-score	0 ~ 29	30 ~ 59	60 ~ 100

Table 2 Subject experimental results

	aa	ab	ac	ad	ae
Precision	94.4	95.8	95.7	100	92.9
Recall	77.8	91.7	95.7	77.8	85.7
F1	48.3	72.1	94.7	63.6	82.8
Level	2	3	3	3	3

Bold indicates algorithm proposed in this paper has the best performance among the five subjects

3.3 Evaluation Literacy-Based ROUGE

In the last summarization block, a score representing the level of the current user is displayed. The score used to recommend the level is the F1-score. If the F1-score is 60 points or higher, recommended Level 3, when it is 30 points to less than 60 points, recommended Level 2, and the F1-score is less than 30, recommended Level 1 (Table 1).

4 Experimental Results

The experiment was conducted with 5 subjects (aa, ab, ac, ad, ae). The subjects who conducted the experiment were between the ages of 20 and 30. Each subject summarized different article to measure the subject's level. The results of the experiment are shown in Table 2. Referring to Table 2, the aa subject had F1-score assigned Level 2 at 48.3, and the remaining 4 (ab, ac, ad, ae) all assigned Level 3.

Raw data (text) is one of the Korean articles. Summary OUTPUT (Levels 3, 2, 1) expresses the summary of the original article by AI divided into three levels (Levels 1, 2, 3). Level 1 summarizes about 20 to 30% of the original text, and Level 2 summarizes about 10 to 15% of the original text. Level 3 can be seen summarizing it to about one sentence.

5 Conclusions

Literacy is the ability to reconstruct information on its own, not passively using information. Literacy, which reinterprets in its own language after reading, is an

important part of our lives. In this paper, we proposed a system that identifies the number of user summary keywords, presents step-by-step summaries, and provides improvements for learning and growth. The raw data (text) is input, and the key keywords are identified using the KoBART algorithm. The user summary text is input. The level of the user is determined by comparing the AI summary with the user's summary. We present a summary of three levels (Levels 1, 2, 3) using KoBART and TextRANK, depending on the user level we identified. It induces interest in reading through step-by-step text summaries. It is possible to acquire the ability to grasp the core of the text and to improve literacy and learning ability.

Considering the production of easy content for slow learners, it makes it easy for young students to access it. The possibility of a contract with a publisher can also be expected by writing a book introduction article and presenting an introduction using a summary algorithm. It is expected to improve the learning ability of more people by using it so that it can be applied to English papers and articles.

Acknowledgements “This research was supported by the Korean MSIT (Ministry of Science and ICT), under the National Program for Excellence in SW (2019-0-01817) supervised by the IITP (Institute of Information and communications Technology Planning and Evaluation)” (2019-0-01817).

References

1. OECD (2021) 21st-century readers: developing literacy skills in a digital world, PISA, OECD Publishing, Paris
2. Brown TB et al (2020) Language models are few-shot learners. Arxiv
3. Kogpt (2021) Kakao-brain korean(hangul) generative pre-trained transformer, <https://github.com/kakaobrain/kogpt>. Last accessed 2021
4. Yang K (2021) Transformer-based Korean pretrained language models: a survey on three years of progress
5. Mihalcea R, Tarau P TextRANK: bringing order into texts
6. Lin C-Y ROUGE: a package for automatic evaluation of summaries. In: Text Summarization Branches Out
7. Chauhan S (2011) Slow learners: their psychology and educational programmes. *Int J Multi Res* 1(8):279–289
8. Park S, Moon J, Kim S, Cho WI, Han J, Park J, Song C, Kim J, Song Y, Oh T, Lee J (2021) Klue: Korean language understanding evaluation. arXiv preprint [arXiv:2105.09680](https://arxiv.org/abs/2105.09680)
9. Ganesan K (2018) Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. arXiv preprint [arXiv:1803.01937](https://arxiv.org/abs/1803.01937)

Extracting Low-Power Code Pattern Method Through Power Measurement of Software Code



Bo Kyung Park and Su Nam Choi

Abstract In the 4th Industrial Revolution, embedded systems were composed of high specification hardware for high performance and low power consumption. The software operating in such an environment should be capable of stable operation in an environment where resources are limited, such as limited hardware and memory, while maintaining original performance. This increase in power consumption leads to a reduction in the available time and an increase in heat generation. Software using high-end hardware increases the power consumption of the device. Eventually, these problems lead to SW/HW errors and shortened device life. This paper proposes a method to minimize power consumption through low power consumption code patterns in the core control structures (loop, branch, modulization, parameter passing). To make this possible, we define some low power code patterns. We can optimize the core code's performance and power efficiency by improving the most complex areas of the software code.

Keywords Low power consumption · Low-power code pattern · Performance

1 Introduction

The use of smart embedded devices has been rapidly spreading in the 4th Industrial Revolution fields such as drones, autonomous robots, and smartphones. The software operating on these embedded systems must maintain the original high performance developed in the existing development environment. Furthermore, software must perform reliable behavior even in environments with limited hardware resources, such as limited power and memory [1]. This increase in power consumption leads

B. K. Park (✉) · S. N. Choi
Chinju National University of Education, Jinju 52673, Gyeongnam, Korea
e-mail: parkse@cue.ac.kr

S. N. Choi
e-mail: csnpower@cue.ac.kr

to a reduction in the available time and an increase in heat generation. Eventually, these problems lead to malfunctioning the device and shortened life.

In order to solve the problem of stable operation of the embedded system, studies are conducted to lower the amount of power consumed and increase the efficiency of energy use. In particular, there are many studies on low-power systems. Research on low-power software design techniques, such as operating systems and application programs, which consume as little power as possible and maintain performance quality, has been conducted in the recent years. However, the design and development of low-power software depend on the developer's skill and intuition without software quality standards [2].

To solve this problem, this paper proposes a method for minimizing power consumption through the power measurement of the software code. This method measures the power used by each code in the high-level step. Then we extract a low-power code pattern that can reduce power. We apply this improved code with the newly defined low power pattern in this paper to reduce the amount of power consumed by the software code.

The composition of this paper is as follows. Section 2 introduces code-based low-power research. Section 3 explains the newly defined low-power code pattern through power measurement experiments. Finally, Sect. 4 describes the conclusions of this study and future research directions.

2 Related Works

Software optimization research for low power has techniques to reduce power consumption at the code level. Researchers define it as Energy Bug or Energy Code Smell. This coding technique can reduce power consumption by analyzing codes constituting program logic.

The Energy Code Smell is a pattern that is likely to reduce power consumption [3]. This pattern is classified into nine different patterns (Parameter by value, Self-Assignment, Mutual exclusion OR, Switch Redundant assignment, Dead local store, Repeated conditionals, Non-short circuit, Useless control flow). These classified patterns reduce power consumption through refactoring. This method can reduce energy consumption, but it can increase the opposite. Therefore, all patterns are not effective at low power. Table 1 gives Energy Bad Smell [4] defined in other research that makes up for defects.

Table 1 Energy bad smell

Energy Bad Smells	Definition	Pattern and definition
Complex expression	Code containing very complex expressions	<pre>if((a == b) && (b == c) (a + d > 0)){ //Statement }else if(!(a == b) && (b == c) (a + d < 0)){ //Statement }else{ //Statement }</pre>
Common sub-expression	Multiple codes containing the same task	<pre>or (If Statement){ a = x + y + 1; b = (x + y) + z; }</pre>
Tail recursion	Code that is called recursively	<pre>int decrease(int a){ if(a > 0){ return decrease(a - 1); }else{ return 0; } }</pre>
Loop structure	A loop that can be optimized by adjusting the structure	<ul style="list-style-type: none"> - Multiple Index Variable - Double-loop structure - Declaration of global variable in loop
Dead code	Code that does not run under any conditions	Codes that are not used in entire execution

3 Low-Power Code Pattern to Minimize Power Consumption

We mention about low-power software code patterns to minimize power consumption. In order to propose a method of minimizing power consumption through source code power measurement, we measure the power consumption of code patterns within the basic control structures using ARM’s ULINK Plus module [5]. We then classify the patterns based on the language paradigm to present a low-power code pattern.

3.1 Low-Power Pattern

① First, we configure the environment for source code power measurement. To do that, we use ARM’s ULINK Plus module and Keil MCU Eval Board to measure the power of the source code [6]. ② After building the source code measurement environment, we run the Keil uVision5 IDE to create a new project. Keil uVision5 IDE is software that supports code measurement of software. ③ Enter the source code to measure power in Keil uVision5. ④ When the environment is configured, run the project. Projects are executed in the order of Build, Run, and Debug. ⑤ When entering the debugging mode, the power consumption of the input source code can

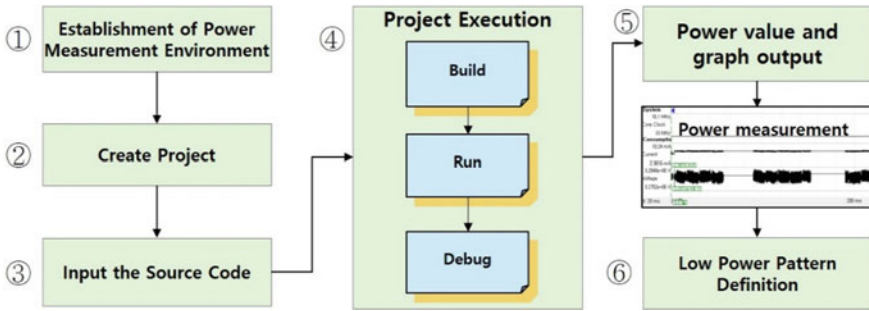


Fig. 1 Procedure of the low power pattern extraction through the power measurement of the source code

be measured. Finally, ⑥ Define the low power pattern by referring to the measured power value and graph (Fig. 1).

3.2 Power Measurement Environment

We mention how to provide a power measurement environment. Connecting the ULINK plus module to the MCBSTM32F200 Board makes software development possible with the Keil tool, and the code result can be checked on the LCD screen. Keil is power measurement software provided by ARM. The connection method is as follows. First, connect the 25 mA module to the ULINK plus module, then connect to the 3V3uC pin of the MCBSTM32F200 Board. Then, power is supplied through the ground. It is connected to the USB port of the PC to supply power. Finally, connect the board and module to the PC and create a new project in the Keil uVision5 IDE. When everything is done, create a c file. We can measure power consumption by writing code (or loading code) and running in debugging mode (Fig. 2).

In ①, it indicates the start and end points of the program. To measure the power during program execution, we must measure the current used up to this start and end point. In ②, the voltage is constantly supplied with 3.3 V. Actually, there is

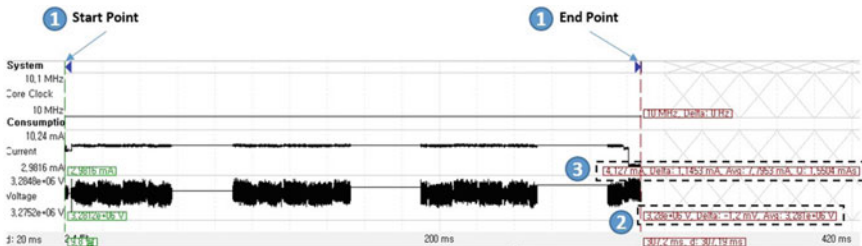


Fig. 2 Graph of the measurement results

a difference on the graph, but it is measured at 3.3 V because the difference is a small amount that changes in 1/1,000,000 units. When we click the start point with a mouse, a green line appears. When we place the cursor on the desired point, various information is displayed as shown on the right. The data of ③ shows the current amount, the delta current amount between the start point (green line) and the end point (red line), the average current amperage amount, and the cumulative current amperage amount from the start point to the end point. By substituting this into Eq. (1), the amount of power can be measured. Equation (1) is the power formula.

$$P = V \times I \tag{1}$$

3.3 Low-Power Code Pattern Definition

To measure the power of software code, we analyze the code power pattern of the procedural language paradigm. Even the code that prints the same result is written differently depending on the developer’s coding tendency. For example, multiple conditions can be added at one time by using the AND or OR operator on one condition in a grade program. On the other hand, we can write code using multiple if statements. Alternatively, we can use if ~ then ~ else and switch statements. In this case, power consumption may be reduced or increased depending on the conditional statement. Through these comparisons, this study defined a total of four procedural code patterns and measured power. Each pattern measures power at least 50 times and averages it (Table 2).

“Double if control statements (①)” uses two if statements to operate on two conditional statements. && operation (②) is performed in the conditional statement. The current average value of ① is 1.591362 mA. The current average value of ② is 1.591362 mA. Equation (1) is used for power consumption. Then, the power consumption of ① is 5.2514946 mW. Since the average current value of ② is 1.584462 mA, the power consumption of ② is 5.2287246 mW. Therefore, it is possible to reduce power consumption by using the && operator in one conditional expression rather than multiple if statements.

In ③ Multiple if the else, we judge various conditions with one variable. In ④ Switch ~ Case, we judge various conditions as switch case statements like ③. The current average value of ③ is 1.802542 mA. ③’s power consumption is 5.9483886 mW. The average value of ④ is 1.725534 mA, so the power consumption is 5.6942622 mW. Therefore, it is possible to reduce power consumption using a switch case statement rather than multiple if-then-else.

“Loop Down Count (⑤)” outputs A while i decreases from 100 to 0 by one. “Loop Up Count (⑥)” outputs A 100 times as i increases from 0 to 100 by one. The average current value of ⑤ is 1.594938 mA, so the power consumption is 5.2632954 mW. The current average value of ⑥ is 1.598812 mA, so the power

Table 2 Measurement results of low-power code pattern

Item	Pattern name
Code pattern	<p>1) Double if control statements</p> <pre>int main() { SystemClock_Config(); int a = 80, b = 90; if (a >= 80) { if (b > 90) { GLCD_DrawString(0,0,"A");} return 0;} Average value: 1.591362 (mA)</pre>
Measured value	Average value: 1.591362 (mA)
Item	3) Multiple if the else
Code pattern	<pre>int main() { SystemClock_Config(); int a = 80; if (a == 80) { GLCD_DrawString(0,0,"A"); }else if (a == 70){ GLCD_DrawString(0,0,"B"); }else if (a == 90){ GLCD_DrawString (0,0,"C"); }else if (a == 60){ GLCD_DrawString (0,0,"D");} return 0;} Average value: 1.802542 (mA)</pre>
Measured value	Average value: 1.802542 (mA)
Item	5) Loop down count
Code pattern	<p>2) if statement that includes && operation</p> <pre>int main() { SystemClock_Config(); int a = 80, b = 90; if (a >= 80 && b > 90) { GLCD_DrawString(0,0,"A");} return 0;} Average value: 1.584462(mA)</pre>
Measured value	Average value: 1.584462(mA)
Item	4) Switch ~ Case
Code pattern	<pre>int main() { SystemClock_Config(); int a = 80; switch (a) { case 80: GLCD_DrawString (0,0,"A"); break; case 70: GLCD_DrawString (0,0,"A"); break; case 90: GLCD_DrawString(0,0,"A"); break; case 60: GLCD_DrawString(0,0,"A"); break;} return 0;} Average value: 1.725534 (mA)</pre>
Measured value	Average value: 1.725534 (mA)
Item	6) Loop up count

(continued)

Table 2 (continued)

Item	Pattern name	
Code pattern	<pre>int main() { SystemClock_Config(); int i; for (i = 100; i > 0; i--) { GLCD_DrawString(16 * i, 0, "A"); } return 0; }</pre>	<pre>int main() { SystemClock_Config(); int i; for (i = 0; i < 100; i++) { GLCD_DrawString(16 * i, 0, "A"); } return 0; }</pre>
Measured value	Average value: 1.594938 (mA)	Average value: 1.598812 (mA)
Item	7) For loop	9) Do-While loop
Code pattern	<pre>int main() { SystemClock_Config(); int i = 0; for (i = 0; i < 100; i++) { GLCD_DrawString (16 * i, 0, "A"); } return 0; }</pre>	<pre>int main() { SystemClock_Config(); int i = 0; do { GLCD_DrawString (16 * i, 0, "A"); } while(i < 100); i++; return 0; }</pre>
Measured value	Average value: 1.802542 (mA)	Average value: 1.740036 (mA)

consumption is 5.2760796 mW. Therefore, the use of Loop Down Count can reduce power consumption.

The power consumption of “For Loop (⑦)” is 5.2741656 mW. The power consumption of “While Loop (⑧)” is 5.2287246 mW. The power consumption of “Do-While Loop (⑨)” is 5.2587678 mW. Using the Do-While statement can consume the least amount of power, but the use of the While statement is recommended because the statement must be executed first due to the nature of the Do-While statement.

4 Conclusion

In this study, we propose a method to minimize power consumption through low power consumption code patterns in key control structures (loop, branch, modulation, parameter passing) to develop high quality software. We presented a code that consumes high power consumption through power measurement of high-level codes. However, the difference in power measurement values is insufficient. Therefore, we will measure the power of more complex and larger codes. We will also need to produce reliable results through various experimental data, systematic experimental scenarios.

References

1. Hyun Sik A, Bo Kyung P, Chul RY, Kim, Du K, Kim (2020) Code visualization approach for low level power improvement via identifying performance dissipation. KIPS Trans. Comp. Comm. Sys. 9(10), 213–220
2. Bo Kyung P, Byungkook J, Kim RYC (2019) Improvement practices in the performance of a CPS multiple-joint robotics simulator. Appl Sci 10, 185–198
3. Vetro V, Ardito L, Procaccianti G, Morisio M (2013) Definition, implementation and validation of energy code smells: an exploratory study on an embedded system. In: The third international conference on smart grids, pp 33–39
4. Jae-Wuk L, Doohwan K, Jang-Eui H (2016) Code refactoring techniques based on energy bad smells for reducing energy consumption. KIPS Tr. Software Data Eng 5(5):209–220
5. ARM ULINK Plus. <http://www.emthink.com/ulink/plus>
6. ARM Keil MCU Eval Board. <http://www.emthink.com/mcb>

A Study on Improved Pet Behavior Monitoring Through Data Imputation Using Multiple Data



Jinah Kim , Hyungju Kim , Chan Park , Jeong-Hyeon Park ,
and Nammee Moon 

Abstract In multiple-data-based pet behavior monitoring-related studies, data may be frequently missing because the area where data can be collected simultaneously is limited. This paper is a study on pet monitoring to supplement this. From the image data, the features of the joint movement of the pet are extracted through DeepLabCut-based learning, and from the sensor data, features according to time such as the direction and speed of the actual movement are extracted. When a pet is out of the shooting area and missing in the image data, the missing data are imputed by augmenting generative adversarial network (GAN)-based behavior data using sensor data. Finally, the behavior is monitored by recognizing the behavior of pets through a convolutional neural network (CNN) and long short-term memory (LSTM)-based network model. The proposed method is meaningful in that it is augmentation between behavioral data, and performance improvement for multiple-data-based behavior monitoring can be expected.

Keywords Behavior monitoring · Behavior recognition · Multi-modal learning · Data imputation · Data augmentation

1 Introduction

Behavior monitoring research is mainly focused on image data and sensor data. Image-based monitoring studies collect data by setting a camera in a specific space using CCTV or IP cameras. Data are collected by attaching sensors to objects or environments (walls, floors, and furniture) or wearing wearable devices [1].

Recently, multiple-data-based studies using both image and sensor data are being conducted. In the case of using only single data, problems with data quality such

J. Kim · H. Kim · C. Park · J.-H. Park · N. Moon (✉)
Hoseo University, Chungcheongnam-Do, 31499 Asan-Si, Republic of Korea
e-mail: nammee.moon@gmail.com

J. Kim
e-mail: kkim.jinah00@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_90

as noise or missing data can be supplemented by combining other types of data. Therefore, multiple-data-based studies show higher performance [2].

However, since multiple-data-based studies deal with high-dimensional data, high-performance computational processing power is required. For this purpose, deep learning-based monitoring studies such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) that automatically extract features are being conducted [3].

Multiple-data-based monitoring should be well synchronized based on the time of data collection. However, there is a difference between the area collected by the image and the area collected by the sensor. A limited space such as inside a car is not a problem, but in a slightly larger indoor space, the camera illuminates only a part of the space. Therefore, image data loss occurs as soon as it exceeds the shooting area of the camera.

This paper is about monitoring the daily behavior of pets indoors, such as inside a house. Due to the privacy problem, it is difficult to collect data without blind spots in the entire area of the room. And due to the nature of the pet, it is difficult to predict the behavior, so image data are often missing. To improve this, we propose a method of imputing the image data through data augmentation based on generative adversarial network (GAN) from the sensor data collected by a wearable device. Finally, monitoring is carried out by recognizing behavior by fusing features extracted from all data.

2 Related Works

2.1 Generative Adversarial Network

GAN is a neural network in which a generator generates data and a discriminator evaluates data to oppose each other and learn to improve performance. Various studies based on GAN are being conducted mainly to generate data. It has been confirmed to be effective in data generation tasks such as generating data with different styles or restoring noisy data to improve data bias [4].

GAN-based research for monitoring mainly consists of studies to generate skeleton data for motion from 2D RGB images [5]. However, in multiple-data-based research, image data may be missing. Therefore, augmentation from other types of data is required rather than image-to-image. In this study, we intend to impute the missing in image data through data augmentation to sensor-to-image.

2.2 Pet Behavior Monitoring

According to the data type, it can be divided into image-based and sensor-based. Image-based research analyzes movement by attaching a camera to the ceiling to see as many movements as possible [6]. Since only the top view is photographed, it is difficult to confirm specific movements. In addition to this, studies are underway to estimate the posture in more detail by extracting the joints [7, 8]. Sensor-based research is mainly conducted with wearable devices [8]. Activities are tracked using commercially available devices such as Petbark and petbit, and acceleration and gyroscope sensors are embedded in the manufactured wearable device.

Most of the current behavior monitoring studies on pets are being conducted based on a single data. Considering the anomalous behavior of pets, analysis of multiple data is better than single data for feature extraction. Based on existing research, this study intends to monitor pet behavior using image data and sensor data (3-axis acceleration and 3-axis gyroscope).

3 The Proposed Pet Behavior Monitoring

The process for pet monitoring proposed in this paper is shown in Fig. 1. The goal of this paper is to increase recognition accuracy by maximally reducing data loss in the learning process of behavior recognition for behavior monitoring. To this end, it is divided into data collection, data preprocessing, and behavior monitoring stages.

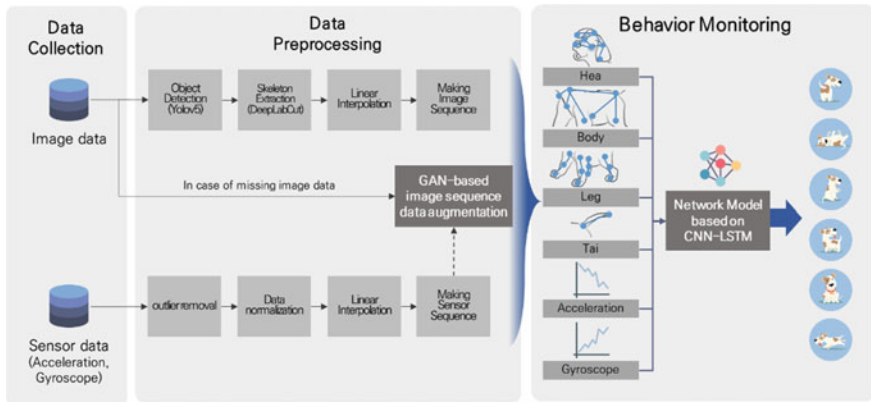


Fig. 1 Process of pet behavior monitoring

3.1 Data Collection

A camera, a wearable device, and a data collection device are used to collect multiple behavioral data. The data collection device is connected to the camera and wearable device to store and manage all data. The camera uses a webcam and is set in a position where the pet’s whole body can be seen. If the image is taken too high, the movement of the pet’s legs may not be seen well, so it was taken between + 30 and - 30° from the front of the pet. It was collected at 20 FPS with 1920 × 1080 resolution. A wearable device is manufactured and is worn by being attached to a neckline. It collects data from the built-in acceleration and gyroscope sensors at a cycle of 50 Hz. The wearable device communicates data with the data collection device through Bluetooth communication.

3.2 Data Preprocessing

The data preprocessing stage is a stage of generating data in the form of a sequence to learn behavior recognition. Pre-processing of the image proceeds as shown in Fig. 2. First, the skeleton is estimated to extract the features of the pet’s behavior. For an accurate skeleton, first, recognize the pet using YOLOv5 and then delete the background. DeepLabCut-based skeleton extraction is performed for the detected pet’s bounding box. After that, for frames in which no object is detected, data are replaced through linear interpolation with intermediate values of the preceding and following frames, and a sequence for the entire length of the time series is generated.

Next, preprocessing the sensor data removes noise and normalizes the sensor data. Outliers are removed using inter quartile range (IQR) and converted to a standard normal distribution through Z-score normalization. As with image data, for missing values, data is replaced through linear interpolation with intermediate values, and finally, a sequence is generated.

The generated image and the sequence data of the sensor are used as input in the next step, but it is impossible when the image data is missing. As shown in Fig. 3, a GAN-based data augmentation model that outputs image sequence data by inputting

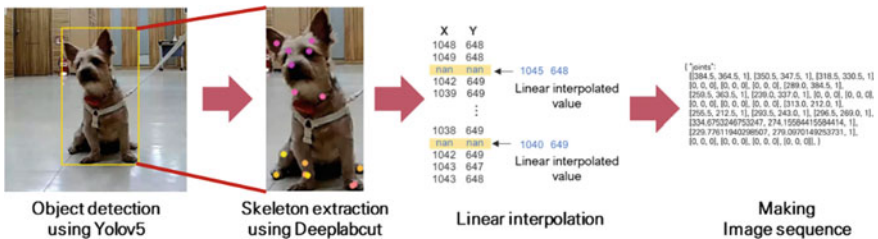


Fig. 2 Process of the image data preprocessing

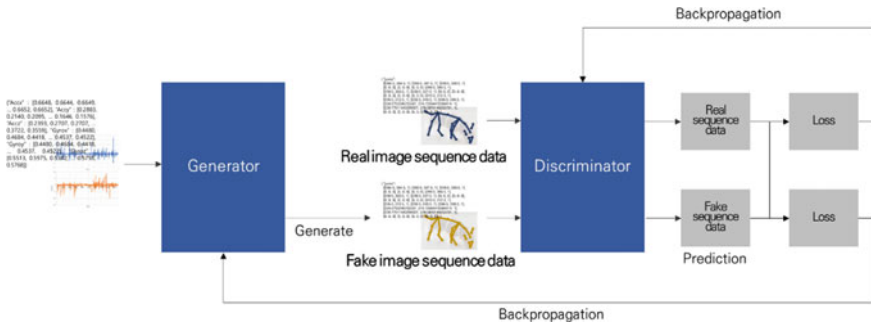


Fig. 3 Process of GAN-based image sequence data augmentation for missing image data

the acceleration and gyroscope sensor sequence data at the time of the missing is learned. Finally, the augmented data replaces the missing data of the image.

3.3 Behavior Monitoring

The previously extracted sequence data recognize the behavior of the pet through the behavior monitoring network. The structure of the monitoring network is a hybrid of CNN and LSTM and is configured as shown in Fig. 4. In the image sequence data, coordinates are divided into head, body, legs, and tail, and features are extracted through 1D-CNN every time, and LSTM is performed on the extracted features over the entire time. In sensor sequence data, 1D-CNN is performed based on time series and 3-axes to extract the main features for each sensor, and LSTM is performed on the extracted features for the entire time. Then, the results of all features are added up, and the pet’s daily behaviors such as sitting, standing, lying down, and eating are classified through softmax.

4 Conclusion

This paper is about monitoring the behavior of pets in an indoor environment through data imputation using multiple data. From the image data, features about skeleton movement are extracted through DeepLabCut learning. And from the sensor data, features such as direction and speed for actual movement are extracted. The extracted features are monitored by recognizing their behavior through a CNN-LSTM-based network model. However, to improve the problem of missing image data in the process of collecting data, image-based behavior data were augmented using sensor-based behavior data through GAN-based learning.

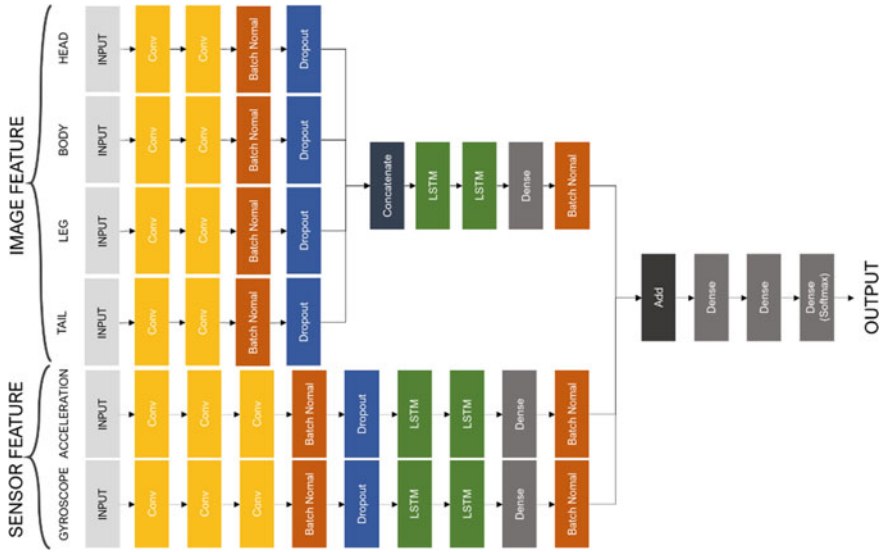


Fig. 4 Structure of network model based on CNN-LSTM for behavior monitoring

The proposed method can be applied in various ways when monitoring an unpredictable object in future, and it is expected that it will solve the problem of the decrease in the number of training data due to the existing missing data.

Acknowledgements This paper was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2021R1A2C2011966).

References

1. Dang LM, Min K, Wang H, Piran MJ, Lee CH, Moon H (2020) Sensor-based and vision-based human activity recognition: a comprehensive survey. *Pattern Recogn* 108(107561)
2. Kim J, Moon N (2022) Dog behavior recognition based on multimodal data from a camera and wearable device. *Appl Sci* 12(6):3199
3. Wang Y, Cang S, Yu H (2019) A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst Appl* 137:167–190
4. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA (2018) Generative adversarial networks: an overview. *IEEE Signal Process Mag* 35(1):53–65
5. Yan Y, Xu J, Ni B, Zhang W, Yang X (2017) Skeleton-aided articulated motion generation. In: *Proceedings of the 25th ACM international conference on multimedia*, pp 199–207
6. Bleuer-Elsner S, Zamansky A, Fux A, Kaplun D, Romanov S, Sinitca A, Masson S, van der Linden D (2019) Computational analysis of movement patterns of dogs with ADHD-like behavior. *Animals* 9(12):1140
7. Kearney S, Li W, Parsons M, Kim KI, Cosker D (2020) Rgb-dog: predicting canine pose from rgb-d sensors. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 8336–8345

8. Tsai MF, Huang JY (2020) Predicting canine posture with smart camera networks powered by the artificial intelligence of things. *IEEE Access* 8:220848–220857
9. Chambers RD, Yoder NC, Carson AB, Junge C, Allen DE, Prescott LM, Bradley S, Wymore G, Lloyd K, Lyle S (2021) Deep learning classification of canine behavior using a single collar-mounted accelerometer: real-world validation. *Animals* 11(6):1549

Deep Learning SPIN Pattern Outlier Detection for Integrated Dynamic Rotary Machine



Jieun Kang , Subi Kim , and Yongik Yoon 

Abstract Recently, the intelligent and advanced IT technology such as IoT, sensor, network and computer vision developed the rotation machinery outlier detection and condition diagnosis technology from vibration sensor data for various industrial environment. However, the rotary machine consists of complex system with various parts and operates under the dynamic environment, almost anomaly detection is not focused on the in–out combined information of rotary machine. With not according to multiple information, anomaly detection doesn't process in the fluent way and be difficult to instantaneous decision making. This paper suggests Spectrogram Power Integrated Pattern (SPIN Pattern) Outlier Detection available to detect outliers based on integrated and multiple frequency patterns of rotary machines. SPIN Pattern extracts vibration frequency patterns from spectrogram image (Spectrogram Pattern), inside vibration attributes and then rotary capacity power frequency patterns (Power Pattern) which is external information. Considering integrated vibration frequency pattern of inside information and power pattern for outside information at the same time, SPIN Pattern is to derive subdivided pattern for fluent outlier causes. After deriving SPIN Pattern, CNN multi-classification model performed outlier detection based on SPIN Pattern and resulted in 85% high accuracy which is confirmed to stable outlier detection and a cause derivation.

Keywords Deep learning · Frequency pattern detection · Outlier detection

J. Kang · Y. Yoon (✉)

Department of IT Engineering, Sookmyung Women's University, 100, Chungpa-ro 47 Gill, Yongsan-gu, Seoul 04310, Korea
e-mail: yiyeon@sookmyung.ac.kr

J. Kang

e-mail: kje9209@sookmyung.ac.kr

S. Kim

Big Data Analysis Convergence, Sookmyung Women's University, 100, Chungpa-ro 47 Gill, Yongsan-gu, Seoul 04310, Korea
e-mail: sbkimdr@sookmyung.ac.kr

1 Introduction

The first rotary machines in various industrial environments are operated with many kinds of parts (bearing, axis, rotating body, belt, etc.) based on various power capacities. When rotating, vibration is occurred by rotational force and various vibration types are detected according to the rotary machine's states. Therefore, defects of rotary machines such as rotation, revolution, sag, and tremor induce natural vibration characteristics patterns [1, 2]. However, if rotary machine is continuously operating under defects occurring with not detecting accurate vibration pattern and attributes, there may happen unpredictable accidents and this results in not only fatal society and economic loss but even human costs [3]. So, it is essential to construct a stable outlier detection system considering the integrated and multiple conditions of possible outliers from rotary machines, which is to detect possible and predictable outliers of rotary machines in a highly accurate way and make stable operation.

Recently, with the development of intelligent and advanced IT technology such as IoT, sensor, network, and computer vision, many industrial environments perform rotary outlier detection and status diagnosis from the vibration signal data attached to rotary machines. There are outlier detection and status diagnosis using statistical values such as mean, variance, skewness, and kurtosis of vibration signals and extraction of feature values through physical values such as root mean square (RMS) and peak-to-peak [4–6]. Others consider vibration frequency attributes information and frequency patterns based on Fourier transform [4–11]. In addition, there are binary classification methods which classify normal and abnormal from frequency patterns based on machine learning and artificial neural networks. Also, recently with developing computing vision technology, deep learning multi-classification is more improved using frequency patterns images for the detection of pattern and outlier causes.

However, nowadays many kinds of outlier detection and status diagnosis studies merely consider various rotating machine parts in a complex way and different vibration types under various operating conditions. Therefore, this paper suggests Spectrogram Power Integrated (SPIN) Pattern Outlier Detection. SPIN Pattern focuses on the inside rotary frequency patterns from various parts and outside information, power capacity patterns at the same time and then integrates both information patterns for induced detailed patterns. The second section describes SPIN Pattern Outlier Detection Model and then the third section is SPIN Pattern experiment and the outlier detection based on SPIN Pattern for validation. After that, the fourth section is the result and finalizes this paper.

2 Spectrogram Power Integrated Pattern (SPIN Pattern) Outlier Detection Model

Pattern method is executed in order of (1) signal data acquisition, (2) STFT-based frequency conversion, (3) spectrogram-based frequency pattern extraction and rated capacity-based frequency pattern extraction, (4) SPIN Pattern extraction, (5) CNN-based SPIN Pattern learning and SPIN Pattern-based outlier detection, as shown in Fig. 1.

2.1 Raw Data and STFT

Raw data is vibration signal data based on time series, which has specific frequency values depending on vibration types. Due to reflecting on time series, STFT is performed to transform frequency. The characteristic of frequency is available to distinguish types of vibration and is identified from how to increase the values of frequency in comparison with standard values on the same frequency. A spectrogram is an image that expresses one dimension in two dimensions through STFT that divides a one-dimensional signal into pieces of a certain length and performs Fourier transform to express time information of the pieces on the x -axis the frequency components of the pieces on the y -axis. Outlier detection by spectrogram enables learning-based pattern extraction with time information unlike FFT and is efficient because it can be learned by removing insignificant pixels and noise in the transformed image. Spectrogram image transform is performed for detecting frequency patterns to identify the rotary inside information frequency pattern. Spectrogram represents frequency amplitude change inside the frequency domain as the depth of color and the spectrogram patterns are derived from the depth of color pattern. After detecting the spectrogram patterns process, the power pattern which is the outside rotary machine information frequency pattern inside the spectrogram pattern is detected and then the frequency patterns are detailed and subdivided.

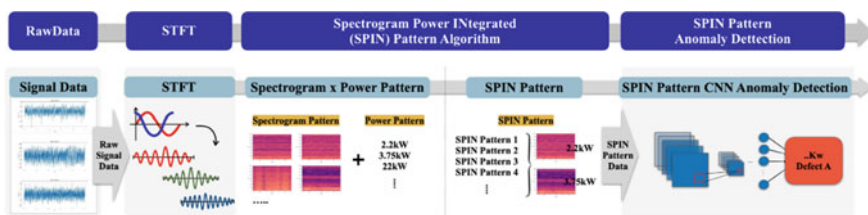


Fig. 1 Spectrogram power integrated pattern (SPIN pattern) outlier detection model

2.2 STFT-Power Integrated Pattern

By integrating the spectrogram pattern and power pattern, the SPIN Pattern is detected. SPIN Pattern performs more detailed and flexible pattern detection by simultaneously integrating the frequency pattern based on inside and outside rotary machine information. Therefore, the SPIN Pattern-based outlier detection is possible to identify the complex and fusion dynamic outlier causes.

3 SPIN Pattern Outlier Detection

For the SPIN Pattern Outlier Detection Algorithm experiment, the vibration data was from AI hub which was machine facility vibration data. Data was collected from acceleration sensors attached to machine facilities in the three metro stations according to possible 5 types (normal, bearing defect, rotating body imbalance, axial misalignment, belt slackness) of rotary machines under 12 kinds of power capacities environments. Each vibration signal data was constructed time and acceleration values collected at 0.00025 s intervals.

3.1 SPIN Pattern

Raw signal data was transformed to time and frequency domain based on STFT. For STFT frequency pattern detection, the frequency was transformed to spectrogram images and spectrogram patterns of each rotary machine states were induced according to vibration nature frequency pattern information. The spectrogram pattern results are Fig. 2 and this shows different patterns in the same rotary machine states under the other power capacities.

Therefore, inside of spectrogram pattern (Sp), it is essential to detect power pattern (Pp) for discriminating power capacities information patterns at the same time and then by integrating the spectrogram patterns and power patterns, SPIN Pattern is constructed. Equation (1) describes the way of detecting SPIN Pattern.

$$\{Sp1, Sp2, \dots\} \otimes \{Pp1, Pp2, \dots\} = \{\text{SPIN Pattern } a, \text{ SPIN Pattern } b, \dots\} \quad (1)$$

Therefore, SPIN Pattern is available to consider both inside frequency patterns able to diagnose each rotary machine's parts and outside rotary machine frequency pattern attributes in a complex way. The integrated and dynamic pattern detection is able to discriminate outliers' causality and detailed outlier detection even though under the various operating environments and unlimited kinds of rotary machine parts.

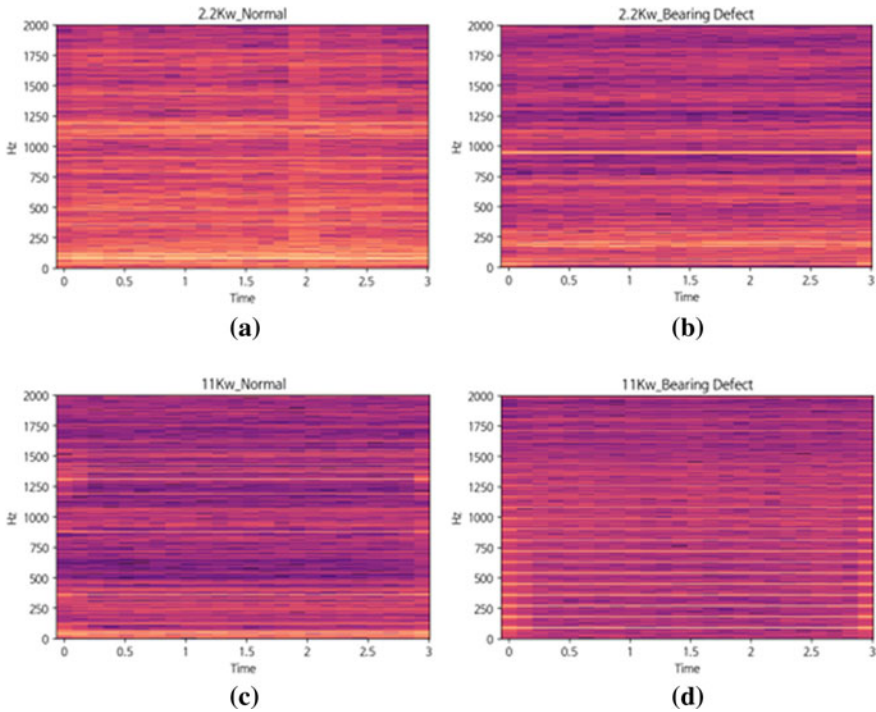


Fig. 2 a, b are both in the rotary normal state but under the other power capacity both patterns are different. c, d in the bearing defect state but both patterns also different even though under the same rotary machine state

3.2 SPIN Pattern CNN Outlier Detection

The multi-classification model is constructed based on CNN to evaluate SPIN pattern-based outlier detection performance. The input spectrogram image size is $150 \times 150 \times 3$ and power and state of rotary machine states are the output. CNN model architecture consists of three convolution layers and two output layers. Convolution layers are stacked with 16, 64, and 256 to maintain the frequency features. Since spectrogram images of input data have features of time series and need to consider time step, the stride is added and set 1 for step-by-step pixel movements. The output layers consist of 512, 2056 for escaping the pixel information disappearance. Each convolution layer consists of ReLU and Max Pooling, and the final output layer consists of SoftMax for multiple classification. The optimizer is Adam, the batch size and epochs is set to 128 and 5 and the learning rate is set to 0.001. This section describes a comparison multi-classification model of spectrogram pattern not according to the integrated pattern and SPIN Pattern which suggested this paper. Figure 3 shows the accuracy of spectrogram pattern and the accuracy of SPIN Pattern is 0.8515. The

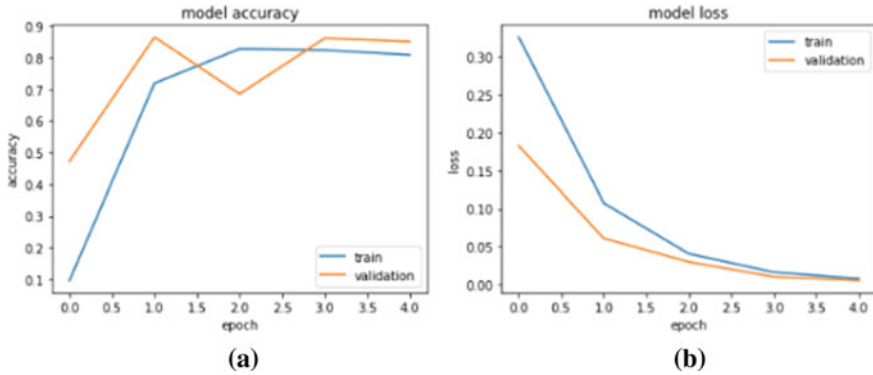


Fig. 3 **a** is the prediction results of the existing method and **b** is the proposed method using spectrogram

result shows that SPIN Pattern is appropriate to classify the rotary machine states based on integrated patterns of capacity and rotary inner states.

4 Result

This paper suggests SPIN Pattern Outlier Detection for rotary machine outlier detection and evaluate the SPIN Pattern performance through CNN multi-classification model. SPIN Pattern consider both spectrogram patterns from STFT frequency images patterns based on various parts of rotary machines inside information and Power Patterns which are rotary each power capacity patterns. Since SPIN Pattern outlier detection induce both outlier patterns from various parts of rotary machine and power pattern according to the capacity information in the complex and integrated way, it is available to detect more detailed and subdivided outlier cause and then more make an accurate decision process. This detailed and dynamic outlier detection helps not only professionals but also people who are in the industrial environment be instantaneous decision making. After that, for more general adaption to all kinds of rotary industrial environments, we process intelligent and more fluent frequency pattern outlier detection studies available to consider more various inside and outside rotary machine information. Then a more robust SPIN Pattern is possible to classify and diagnose the detailed states of rotary machines by developing an outlier detection model.

References

1. Fourier transform. Wikipedia, https://en.wikipedia.org/wiki/Fourier_transform
2. Jeongho J, Jeonguk K (2008) Equipment diagnosis through vibration measurement. *Fire Protect Technol* 44:30–43
3. Jiao J, Zhao M, Lin J, Liang K (2019) Hierarchical discriminating sparse coding for weak fault feature extraction of rolling bearings. *Reliab Eng Syst Saf* 184:41–54
4. Namcheol P, Jeongseo K (2021) A study on vibration parameters for the diagnosis of air compressor bearing failure of urban railway vehicles. *Korean Soc Railw* 9(4):1065–1074
5. Seung-il K, Yoojeong N, Youngjin K, Seonhwa P, Byungha A (2021) Fault classification model based on time domain feature extraction of vibration data. *Comput Struct Eng Inst Korea* 34(1):25–33
6. Altaf M, Akram T, Khan MA, Iqbal M, Ch MMI, Hsu C-H (2022) A new statistical features based approach for bearing fault diagnosis using vibration signals. *Sensor* 22(5):2012
7. Kafeel A, Aziz S, Awais M, Khan MA, Afaq K, Idris SA, Alshazly H, Mostafa SM (2021) An expert system for rotating machine fault detection using vibration signal analysis. *Sensors* 21(22):7587
8. Eraliev O, Lee K-H, Lee C-H (2022) Vibration-based loosening detection of a multi-bolt structure using machine learning algorithms. *Sensors* 22(3):1210
9. Li Y et al (2018) Planetary gear fault diagnosis via feature image extraction based on multi central frequencies and vibration signal frequency spectrum. *Sensors* 18(6):1735
10. Wang L-H, Zhao X-P, Wu J-X, Xie Y-Y, Zhang Y-H (2017) Motor fault diagnosis based on short-time fourier transform and convolutional neural network. *Chin J Mech Eng* 30(6):1357–1368
11. Sungmok J, Woojin C (2022) A study on deep learning-based fault diagnosis using vibration data of wind generator. *J Korean Inst Inf Technol* 20(6):129–136
12. Hasan MJ et al (2021) Bearing fault diagnosis using multidomain fusion-based vibration imaging and multitask learning. *Sensors* 22(1):56
13. Liu C, Meerten Y, Declercq K, Gryllias K (2022) Vibration-based gear continuous generating grinding fault classification and interpretation with deep convolutional neural network. *J Manuf Process* 79:688–704
14. Li J, Liu Y, Li Q (2022) Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method. *Measurement* 189:110500
15. Kyungwon K, Kyeongmin L (2020) CNN-based automatic machine fault diagnosis method using spectrogram images. *J Korea Inst Conver Sig Process* 21(3):121–126

Optimized Vehicle Fire Detection Model Based on Deep Learning



ByoungGun Park , Ji Su Park , and YounSoon Shin 

Abstract Early fire detection is essential to prevent serious problems such as fire disasters and human casualties. These systems should be able to identify fire disasters and send alarms quickly. Existing sensor-based systems tend to be identified after a fire increases because they detect smoke or heat. This paper's proposed vision-based fire detection system can immediately detect fires from cameras and send alarms faster than sensor-based fire detection systems. To this end, we used deep learning to detect fire in real situations easily. The proposed model can achieve high performance even when catching a vehicle fire by improving the backbone based on YOLOv5. The backbone was enhanced based on Facebook AI Research's RegNet, and unlike detecting other general objects, it was configured to distinguish and recognize streetlights that can be confused with fires. In addition, various methods such as data enhancement, model ensembling, and transition learning have been added to improve the accuracy of this model. The proposed model has significantly improved by about 11% compared to the average precision of the existing model (mAP).

Keywords Deep learning · YOLOv5 · Transfer learning · Model ensembling · RegNet

B. Park · Y. Shin (✉)

Department of Computer Science and Engineering, Dongguk University, Seoul 04620, South Korea

e-mail: ysshin@dongguk.edu

B. Park

e-mail: okharry1@dgu.ac.kr

J. S. Park

Department of Computer Science and Engineering, Jeonju University, Jeonju-si 55069, South Korea

e-mail: jisupark@jj.ac.kr

1 Introduction

According to a study by the National Fire Agency, the damage caused by fires increases yearly [1]. Since fire is a vital disaster threatening life and property, it is essential to find out and inform about fire accidents early. In particular, as climate change and carbon neutrality have recently become global hot topics, and the status and demand of E-mobility are rapidly increasing, electric vehicle fires are also on the rise, but the fire response system has not yet been stabilized. Early suppression of automobile fires is very important because many casualties may occur due to vehicle explosions if not extinguished at the initial stage. In particular, if electric vehicles fail to respond in the early stages, the only way to extinguish the fire is to wait for the car to burn down for more than 12 h due to an electric battery issue [2]. The smoke detectors commonly used today are sensor-based and are primarily designed for indoor use. Many studies have designed fire detection and management systems with sensors in buildings or enclosed environments [3, 4]. However, there are many cases where no sensors or risk management is complex in facilities with outdated environments or outdoors. Using CCTV images rather than sensors is a good approach. Detecting a fire in a vehicle through CCTV is critical, especially at night or in a dark parking lot; it is not easy to distinguish between headlights and fires due to light spread. Therefore, we want to improve the accuracy of our model for object recognition and detect vehicle fires with great precision by determining whether it is a vehicle fire or other lightings such as vehicle lights and street lights.

In addition, with the development of robotics, the number of robots used in security and disaster environments is increasing. The proposed model specialized in fire detection can be actively applied to various fields, such as disaster robots and CCTV.

The fire detection system proposed in this paper uses deep CNN that facilitates fire detection under various conditions. An open-source-based deep learning framework YOLO [5] also detects objects in images with excellent accuracy, as in these examples.

However, this is advantageous for detecting various objects, but it is still lacking in terms of accuracy for recognizing a particular thing for special purposes. In this paper, to see fires in vehicles more accurately than in existing models, the backbone in the YOLOv5 model has been transformed from CSPDarkNet to RegNet [6]-based architectures to improve its performance.

The proposed “vehicle fire detection model” recognizes three classes. It detects the vehicle to check if the car is on fire. In addition, the model detects both smoke and fire to detect a fire early. Udacity autonomous vehicle datasets are used to train this model to detect vehicles. The Udacity dataset is mainly composed of video frames taken from urban roads. It provides a total number of 404,916 video frames for training and 5614 video frames for testing. And to detect smoke and fire, we used the fire detection dataset provided by the MIVIA Lab. This dataset has been used to test our fire and smoke detection methods. It is composed of 31 videos; both acquired in real environments. The dataset consists of two main parts: the first 14

videos characterized by the presence of the fire and the last 17 videos, which do not contain any event of interest.

2 Related Work

2.1 Surveillance System Technology Using Artificial Intelligence on CCTV

CCTV has become essential equipment in the city center today. In order to solve various accident accidents and traffic problems, the need for CCTV is increasing day by day, and the image quality is getting better and deployed in various places.

As the demand for CCTVs continues to increase, there is also a very lack of people who play a role in monitoring them. In order to solve this problem and compensate for the shortcomings that humans cannot detect 24 h a day, deep learning is introduced into CCTV so that the police can quickly recognize and solve necessary situations or emergencies. Many approaches using CCTV are being introduced in the study of early fire detection [7].

2.2 Deep Learning Model

Various studies on fire recognition using deep learning have been published [8–10]. YOLO is a network model that learns from labeled data and can estimate the type, location, and size of objects within the frame in real-time. This model is high-speed compared to conventional models such as R-CNN [11]. In particular, YOLOv5 is architected by CSP DarkNet53's [12, 13] backbone structure, which is edited to copy a feature map of the base layer, then send one copy through a dense block and move the other to the next step to separate it.

In addition, the head plays a role in finding the object's position based on the extracted feature map, and initially sets Anchor Box and uses it to derive the bounding box as a result. In particular, there are nine anchor boxes because three are used at each scale depending on the object's size to be recognized.

Table 1 Performance of each model (transfer learning)

Detection model	AP	FPS
Without transfer learning	0.85	32
With transfer learning	0.88	32

3 Optimized Vehicle Fire Detection Model

3.1 Transfer Learning

It takes a lot of time to learn the desired object recognition model through 10,000 units of images and GPU. Therefore, transfer learning can be performed for efficient model learning by reusing models pre-learned through a lot of learning data to generate high-accuracy models with relatively little data.

The CPU used the Intel(R) 12,700 K for model learning and performance evaluation, and the GPU used the NVIDIA GeForce 3060. Therefore, the AP and FPS for the model that performed the transfer classification and the model that recognized the fire without proceeding are shown in Table 1.

3.2 Model Ensemble

Model boosting process divides the train dataset and the test dataset then makes another dataset with the train dataset. After learning the model with this dataset, validation is performed with the entire train dataset in the learned model. In this way, proposed model applied the boosting method in model ensemble, and Table 2 derived the model's AP and FPS.

Table 2 shows that the model that conducted model ensemble has higher accuracy for all classes than the model that only worked with transfer learning. Although the FPS was reduced by 1 when proceeding with model ensemble, this improvement method was also practical because it did not affect the accuracy.

Table 2 Performance of each model (model ensemble)

Detection model	AP	FPS
Without transfer learning	0.85	32
With transfer learning	0.88	32
Transfer learning and model ensemble	0.93	31

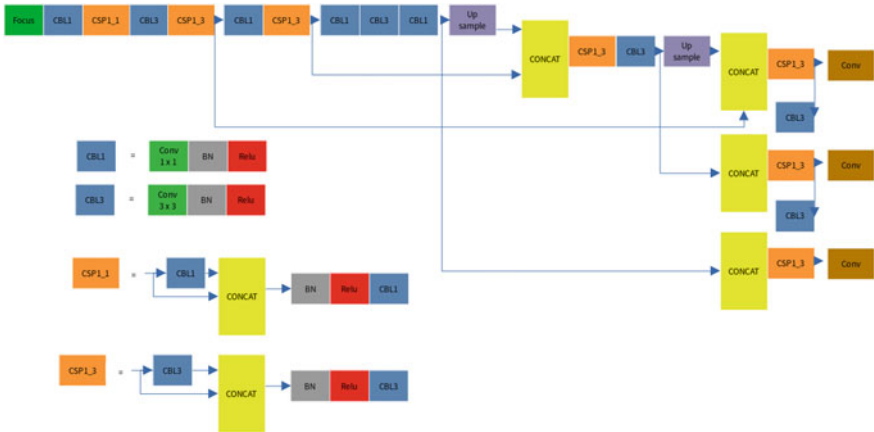


Fig. 1 Improved model architecture

3.3 Improved Fire Recognition Accuracy Through RegNet-Based Architecture Changes

To construct this architecture, RegNet compared network structures such as VGG [14], ResNet [15], and ResNeXt to create AnyNet with relatively free design space and found a low-dimensional design space consisting of simple networks. RegNet’s design space has fewer computations and fewer epochs. It also shows stability at close to 20 blocks without the use of a bottleneck or inverted bottleneck.

RegNet is more effective in mobile environments and performs better than ResNe(X)t. In addition, it is five times faster than EfficientNet [16] and has high generalization performance. Figure 1 is the architecture of proposed model using RegNet.

In this design, RegNet is divided into the input in the stem, followed by the body and head parts that predict class. Here, the calculation amount and accuracy were determined by fixing the stem and the head and changing the body of the network. The body consists of four stages and continues to reduce the resolution. Each stage has individual blocks, and each stage has a block number, a block width, and a block parameter. Here, we investigated the common choice of deep network design. It is used as $N = 100$, epoch = 25, and learning rate = 0.1, and the depth is optimized at around 60 layers, in contrast to the general theory that performance increases as depth increases. It was also found that the input image showed the best performance at 224×224 .

Table 3 Performance of each model (proposed model)

Detection model	AP	FPS
Without transfer learning	0.85	32
With transfer learning	0.88	32
Transfer learning and model ensemble	0.93	31
Proposed model (transfer learning and model ensemble and RegNet-based)	0.95	31

4 Experiment Result

The above studies used transfer learning, model ensemble, and RegNet-based methods to create models optimized for fire detection from existing models. Therefore, the improved APs by sequentially applying the above techniques are expressed in the following table.

Table 3 shows that the model that conducted transfer learning and model ensemble and RegNet together had about 11.7% higher accuracy for all classes than the general model that did nothing.

5 Conclusion

This study conducted a study to detect vehicle fires by applying the YOLOv5 neural network and the RegNet structure among object recognition algorithms. General vehicles, vehicles with smoke before a fire, and vehicles with fire were used as images. The detailed model is an improvement on the YOLOv5s model for rapid object recognition among YOLOv5, and learning adds one performance improvement method for each, confirming and studying performance improvement.

To improve fire detection performance, transfer learning, model assembly, and regNet-based architecture were used to improve performance by 11.7%, which could reduce accidents of vehicle fires through more effective recognition.

Acknowledgements This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the High-Potential Individuals Global Training Program (RS-2022-00155054) (50%) and under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01789) (50%), supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation).

References

1. e-NARA index, https://www.index.go.kr/potal/main/EachDtIPageDetail.do?idx_cd=1632
2. SBS News, https://news.sbs.co.kr/news/endPage.do?news_id=N1006919079&plink=ORI&cooper=NAVER. Last accessed 09 Oct 2022

3. Yar H, Hussain T, Khan ZA, Koundal D, Lee MY, Baik SW (2021) Vision sensor-based real-time fire detection in resource-constrained IoT environments. *Comput Intell Neurosci* 2021:1–15. <https://doi.org/10.1155/2021/5195508>
4. Kim JG, Choi YS, Cho KR, Lee ES (2022) Study for design of defect management to improve the quality of IoT products. *KIPS Trans Softw Data Eng* 11(6):229–236
5. Bochkovskiy A, Wang CY, Mark Liao HY (2020) YOLOv4: optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934)
6. Ilija R, Raj Prateek K, Ross G, Kaiming H, Piotr D (2020) Designing network design spaces. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 10428–10436
7. Sridhar P, Sathiyaraj RR (2021) *J Phys Conf Ser* 1916:012027
8. Jiao Z, Zhang Y, Jing X, Mu L (2019) A deep learning based forest fire detection approach using UAV and YOLOv3. In: *2019 1st international conference on IAI*. <https://doi.org/10.1109/ICIAI.2019.8850815>
9. Nematullo R, Paul A, Saeed F, Seo H (2021) Realtime fire detection using CNN and search space navigation. *J Real-Time Image Proc* 18:1331–1340. <https://doi.org/10.1007/s11554-021-01153-4>
10. Majid S, Alenezi F, Masood S, Ahmad M, Gündüz ES, Polat K (2022) Attention based CNN model for fire detection and localization in real-world images. <https://doi.org/10.1016/j.eswa.2021.116114>
11. Ross G, Jeff D, Trevor D, Jitendra M (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on CVPR*, pp 580–587
12. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. arXiv preprint [arXiv:1804.02767v1](https://arxiv.org/abs/1804.02767v1)
13. Gao H, Zhuang L, Laurens M, Kilian QW (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE CVPR*, pp 4700–4708
14. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
15. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
16. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th international conference on machine learning*, PMLR, vol 97, pp 6105–6114

Smarteria: Intelligent Cafeteria Using Multimodal Data



Yuri Seo, Teh-Jen Sun, Seol Roh, In-hun Choi, Gi-dae Hong, Hong-Ju Jeong, Seung-Woo Jeong, Choong Seon Hong, Seong-Bae Park, and Eui-Nam Huh

Abstract Convergence technologies including the Internet of Things, Big Data, and Artificial Intelligence can detect and respond to dangerous situations through real-time monitoring. Meanwhile, the cafeteria environment has required safety management as it is exposed to various accident risks. Therefore, we conducted a study that applied multimodal data and artificial intelligence for the real-time safety management of the cafeteria environment. We show that multimodal data can improve accuracy in real-time detection. Thus, combining Artificial Intelligence with a multimodal sensors environment will be effective in various real-time safety managements.

Y. Seo · T.-J. Sun

Department of Artificial Intelligence, Kyung Hee University, Yongin-si 17104, South Korea
e-mail: yuri0329@khu.ac.kr

T.-J. Sun

e-mail: dlwmznz11@khu.ac.kr

S. Roh · I. Choi · G. Hong · H.-J. Jeong · S.-W. Jeong · C. S. Hong · S.-B. Park · E.-N. Huh (✉)
Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104,
South Korea

e-mail: johnhuh@khu.ac.kr

S. Roh

e-mail: seven800@khu.ac.kr

I. Choi

e-mail: inhun321@khu.ac.kr

G. Hong

e-mail: kidae92@khu.ac.kr

H.-J. Jeong

e-mail: sub06038@khu.ac.kr

S.-W. Jeong

e-mail: jyng2227@khu.ac.kr

C. S. Hong

e-mail: cshong@khu.ac.kr

S.-B. Park

e-mail: sbpark71@khu.ac.kr

Keywords Multimodal data · Object detection · Intelligent cafeteria

1 Introduction

The Internet of Things (IoT), Big Data, and Artificial Intelligence (AI), the core technologies of the fourth industrial revolution, are leading the paradigm across industries and society [1]. Recently, research is applied convergence technologies including the IoT, Big Data, and AI which is increasing in real-time safety management environments such as smart food safety management systems [2], helmet-wearing at construction sites [3], and natural disaster response management systems [4]. On the other hand, the cafeteria environment is exposed to various accident risk factors including slip, contact with abnormal temperatures, and collision with objects [5]. Most of the safety management in the cafeteria can be improved through real-time notifications including accurate detection. Therefore, in this study, we examined to develop and implement Smarteria technology that prevents cafeteria accidents and improves the environment by multimodal data and artificial intelligence using IoT devices such as thermal imaging cameras and IoT sensors.

The aims of this study are as follows: (1) object detection and context collection to prevent safety accidents and to assist food quality; (2) real-time detection with lightweight architecture; (3) multimodal data and container scheduling system; (4) IoT control management architecture for cafeteria safety and environmental improvement.

2 Related Work

Fire disaster at the cafeteria is more likely to occur when there are flammable objects or non-person around the flame. Therefore, detect fire and flammable object, person around the countertop is important. However, if the fire is covered by kitchen utensils or a person, accurate detection is difficult. Also, slipping is the most common accident type in cafeteria. To prevent slipping issue, safety management system is necessary to detect water on the floor.

2.1 Person Detection

The real-time deep learning-based person detection model has been applied in various businesses including intelligent vehicles [6] and behavior monitoring [7]. Also, many studies have focused on comparing the performance and speed of each model when detecting person. Kim et al. showed that among various models including You Only

Look Once v3 (YOLOv3), Single-Shot Detector (SSD), Faster Regions with Convolutional Neural Networks' features (R-CNN), and Region-based Fully Convolution Networks (R-FCN) of deep learning, YOLOv3 is the best in person detection accuracy and speed [8]. Khalfaoui et al. showed that the YOLOv5 model, the latest model of YOLO, is the most accurate person detection results in real time [9].

2.2 Flammable Object Detection

YOLO which is our base deep learning architecture has both low latency and high accuracy also in the flammable object detection. Tornero et al. presented a real-time detection and location system for recognizing waste in outdoor environments using YOLOv3 and YOLOv5 with a robot-mounted camera [10]. Maharjan et al. used the YOLOv5 model to detect plastic in rivers imaged by unmanned aerial vehicles [11].

2.3 Water Detection

Water is a liquid, a mixture of oxygen and hydrogen. Water scatters light and reduces saturation or color homogeneity, making it blurry. Therefore, it is difficult to detect a small amount of water by image analysis alone. The deep learning-based method to detect water is being used to solve this problem. Santana et al. proposed a method using segmentation to detect water on lakes and roads, where a lot of water exists [12]. Morris et al. improved the performance of the water detection model by determining the road using Detectron2 and converting that part into HSV format [13].

2.4 Utilize Multimodal Data

Previous detection studies include the person and the flammable objects that used camera data only. These methods had a limitation in that detection performance is degraded when an object is covered by another object. Multimodal data can utilize complementary characteristics through the fusion of various sensors. Netinant et al. used multimodal sensors for human activity recognition [14] and, in addition, performed Raspberry Pi 4 and LWIR microthermal imaging camera module measure the real-time temperature of the human body [14].

3 Proposed Systems

3.1 Conceptual Framework

Figure 1 shows the Smarteria framework that utilizes multimodal context data including thermal imaging cameras and various sensors to detect the cafeteria environment. It supports real-time detections using a lightweight neural network model.

Our study fused a thermal imaging camera and sensors to detect the cafeteria environment to gain more accuracy. Multimodal data divide the image into 10×10 areas through preprocessing and create each cell. The temperature data of each cell were calculated as the average value of four points (left-up, left-down, right-up, and right-down), and the preprocessed data were saved in local storage.

Data were transmitted from local storage to the object detection model container through the Convolutional Neural Network's (CNNs) model. Object detection was performed to prevent safety accidents, and the result was saved in the local host. The main process uses object detection results to control hardware such as alerts, smart outlet, display, ventilator, and warning light logo.

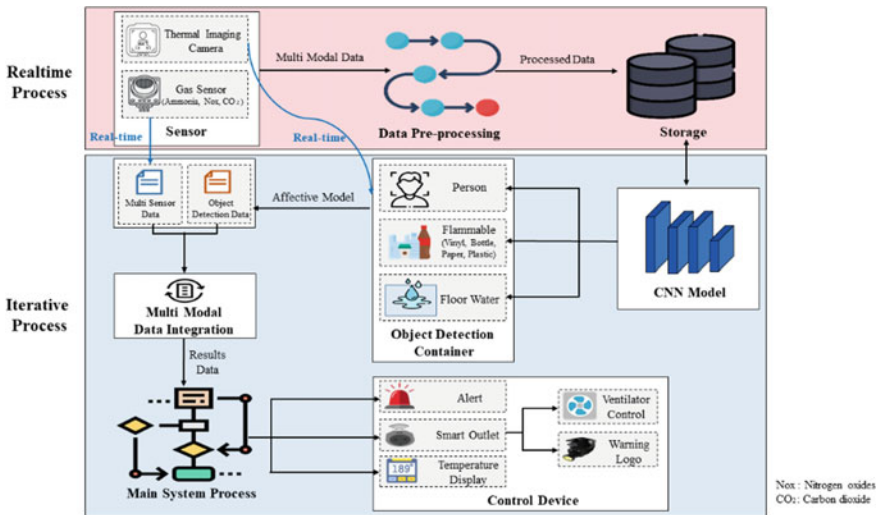


Fig. 1 Conceptual framework of Smarteria using multimodal data



Fig. 2 Flow chart for AI-driven multimodal data processing

3.2 System Process

Figure 2 shows how AI-based multimodal data are implemented. The safety management system starts the operation and processes the measurement of the fused multimodal sensor. When a problem such as a fire or wet floor occurs, the alarm is provided to the user by control devices such as an alarm lamp, fan, and display.

3.3 Object Detection Data Analysis

A. The Person and the Flammable Object Detection.

Image Dataset Collection and Preprocessing

Table 1 shows the detail of dataset what used in the training process. The person dataset has images of cooks in a cafeteria and ordinary people looking from top to

Table 1 Available datasets

Class name	Total images	Images for training (before data augmentation)	Images for training (after data augmentation)	Images for validation	Images for testing
Person	8467	1992	5976	253	246
Flammable object	70,133	56,107	–	7009	7017

bottom [15]. The flammable object dataset has four types: plastics, paper, plastic bottles, and vinyl [16]. The train set, validation set, and test set were each composed of the ratio of 8:1:1 for both the person and the flammable object. In the end, the person detects dataset which has 7968 images including the authentic 1992 images and the augmented 5976 images and 56,107 images for the flammable object were used as the train set.

Object Detection

CNN-based object detection algorithm YOLO can achieve faster inference speed and accuracy than Faster R-CNN. So, it is efficient for real-time object detection [17]. The YOLOv5 was the most accurate and had a better ability to recognize small objects than YOLO's other four series. The YOLOv5 architecture includes YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The safety management requires sufficient accuracy to identify the person, cooking pot, and the flammable object lightweight for real-time performance. Therefore, YOLOv5s is preferred in our implementation. The temperature of cooking pot detected by thermal image camera is used to prevent overcooking and assist to be a better food quality as well as safety for flammable objects.

B. Water Detection

Since it is difficult to detect a small amount of water on the floor only with the image of the cafeteria input through the camera, we used both image and temperature data. First, we used a segmentation model to detect the floor part. Changqian et al. showed Bilateral Segmentation Network V2 (BiSeNet V2) that achieved fast inference speed and high accuracy by considering both low-level detail and high-level semantics [18]. Therefore, we used BiSeNet V2 to achieve real-time performance and high accuracy in water detection. Second, after calculating the histogram by extracting the temperature data of the floor part, the pixel ratio of degrees or less was calculated. In this study, if the pixel ratio of 25° or less was threshold or more, we determined that moisture was present.

4 Experiments' Evaluation

4.1 Hardware

For experiment, we used hardware resource as follows: Arduino Uno, Raspberry Pi 4B, Logojector, Gas air quality sensor module (MQ-135), Thermal imaging camera (FLIR ONE Pro), LED display, Warning light, and Smart outlet.

4.2 Training and Testing YOLOv5 Models

The pre-trained YOLOv5's model was trained with the person and the flammable object datasets. Figure 3 shows the results of the validation dataset derived by setting the network to batch size 16, image size to [640, 640], and learning 300 epochs. The final mean Average Precision (mAP) was 0.5 for all classes, 0.936 for the (a) 'person', and 0.969 for the (b) 'flammable object'. The mAP 0.5:0.95 for all classes is 0.739 for the 'person' and 0.924 for the 'flammable object'. Recall which is the ratio of that identified as true of actual truth is important for the safety system. The 'person' scored 0.892 and the 'flammable object' was 0.934. The 'person' and the 'flammable object' are recorded at approximately over 90%. Also, the precision was 0.972 for the 'person' and 0.970 for all classes of the 'flammable object' dataset. The precision–recall (PR) curves of our YOLOv5 models are shown in Fig. 4.

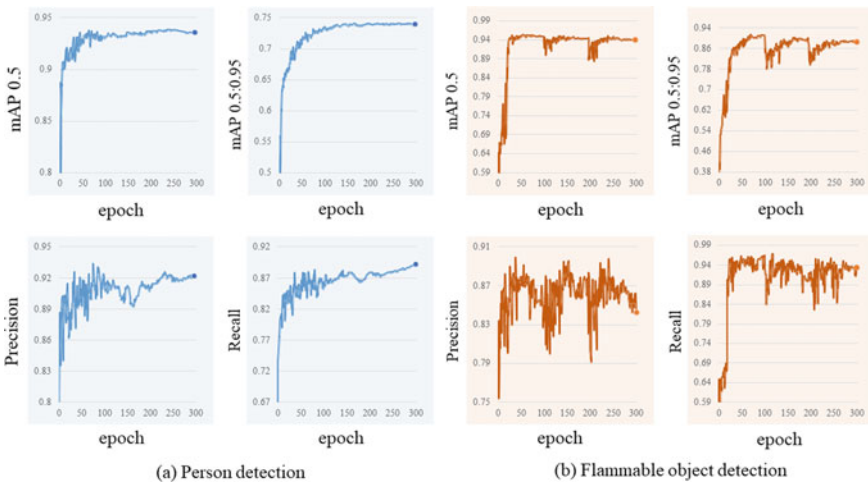


Fig. 3 Person and the flammable object detection training results with YOLOv5

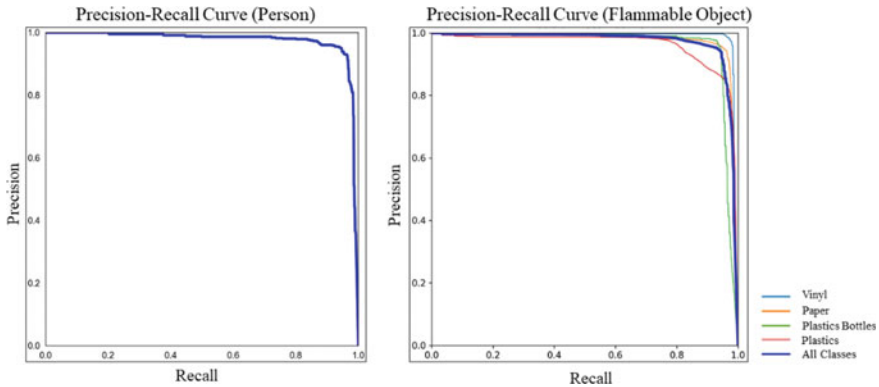


Fig. 4 PR curve of the person and the flammable object test dataset

Table 2 Alert for each situation

Alert	Situation
White alert	• Normal
Red alert	• Fire detected without person
	• Abnormally high temperature
Purple alert	• Fire detected and flammable object detected
Logojector	• Wet detected on the floor

4.3 Real-Time Dangerous Situation Detection Test

This section shows the result of testing our Smarteria safety management system. To test Smarteria, we defined alert for each situation in Table 2.

Figure 5 shows the result of detecting a dangerous situation based on the multi-modal sensor: (a) non-dangerous situation; (b) fire detected, but the person did not detect (red alert); (c) both fire and person detected; (d) even though the person detected, the temperature is abnormally high (red alert); (e) fire, person, and the flammable object detected (purple alert); (f) water detected on the floor (logojector).

5 Conclusion

In conclusion, AI technology with multimodal sensors can provide more accurate real-time detection and alert to users in the cafeteria environment. Therefore, our study adds the result that the convergence of AI and multimodal methods is effective, so it can be helpful in environments that require real-time safety management. In this study, the case with the highest accident rate is considered and tested in the cafeteria

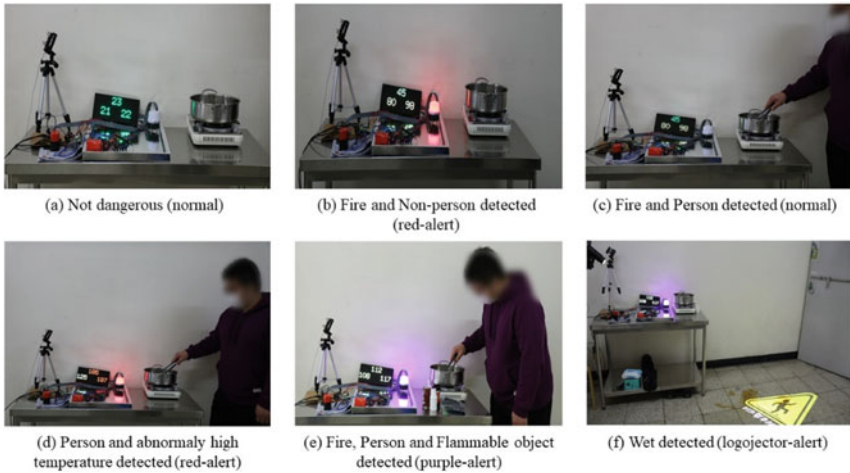


Fig. 5 Experimental result of Smarteria abnormal detection system

environment. In future research, it will be applied to various situations where safety accident prevention is required.

Acknowledgements This work was supported by Institute for Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00818, Machine Learning Based Low Bandwidth Image Communication Edge Computing System for Proactive Anomaly Detection on Smart Plant Environment) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A4A1018607). Professor Eui-Nam Huh is the corresponding author.

References

1. Saadia D (2021) Integration of cloud computing, big data, artificial intelligence, and internet of things: review and open research issues. *Int J Web Based Learn Teach Technol (IJWLTT)* 16(1):10–17
2. Gwon S (2021) The current status of smarter food safety management. *Food Sci Ind* 54(3):124–131
3. Zhou F, Zhao H, Nie Z (2021) Safety helmet detection based on YOLOv5. In: 2021 IEEE international conference on power electronics, computer applications (ICPECA). IEEE, pp 6–11
4. Sun W, Bocchini P, Davison BD (2020) Applications of artificial intelligence for disaster management. *Nat Hazards* 103(3):2631–2689
5. Ko DH, Jeong BY (2019) Work-related injuries of educational support staff in schools. *Int J Occup Saf Ergon* 25(4):568–574
6. Braun M, Krebs S, Flohr F, Gavrilu DM (2019) Eurocity persons: a novel benchmark for person detection in traffic scenes. *IEEE Trans Pattern Anal Mach Intell* 41(8):1844–1861
7. Park J, Chen J, Cho YK, Kang DY, Son BJ (2019) CNN-based person detection using infrared images for night-time intrusion warning systems. *Sensors* 20(1):34

8. Kim CE, Oghaz MMD, Fajtl J, Argyriou V, Remagnino P (2018) A comparison of embedded deep learning methods for person detection. arXiv preprint [arXiv:1812.03451](https://arxiv.org/abs/1812.03451)
9. Khalfaoui A, Badri A, Mourabit IE (2022) Comparative study of YOLOv3 and YOLOv5's performances for real-time person detection. In: 2022 2nd international conference on innovative research in applied science, engineering and technology (IRASET). IEEE, pp 1–5
10. Tornero P, Puente S, Gil P (2022) Detection and location of domestic waste for planning its collection using an autonomous robot. In: 2022 8th international conference on control, automation and robotics (ICCAR). IEEE, pp 138–144
11. Maharjan N, Miyazaki H, Pati BM, Dailey MN, Shrestha S, Nakamura T (2022) Detection of river plastic using UAV sensor data and deep learning. *Remote Sens* 14(13):3049
12. Santana P, Mendonça R, Barata J (2012) Water detection with segmentation guided dynamic texture recognition. In: 2012 IEEE international conference on robotics and biomimetics (ROBIO). IEEE, pp 1836–1841
13. Morris C, Yang JJ (2021) A machine learning model pipeline for detecting wet pavement condition from live scenes of traffic cameras. *Mach Learn Appl* 5:100070
14. Netinant P, Vasprasert P, Rukhiran M (2021) Evaluations of effective on LWIR micro thermal camera IoT and digital thermometer for human body temperatures. In: 2021 the 5th international conference on E-commerce, E-business and E-government, pp 20–24
15. National Information Society Agency, <https://aihub.or.kr/unitysearch/list.do?kwd=%EA%B3%B5%ED%95%AD+%EC%9D%B4%EC%83%81%ED%96%89%EB%8F%99>. Last accessed 21 Aug 2022
16. National Information Society Agency, <https://www.aihub.or.kr/aihubdata/data/view.do?curMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=140>. Last accessed 23 Aug 2022
17. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, vol 28
18. Yu C, Gao C, Wang J, Yu G, Shen C, Sang N (2021) Bisenet v2: bilateral network with guided aggregation for real-time semantic segmentation. *Int J Comput Vision* 129(11):3051–3068

Post-training with Data Augmentation for Improving T5-Based Question Generator



Gyu-Min Park , Seong-Eun Hong , Choong Seon Hong ,
and Seong-Bae Park 

Abstract Post-training is known to be effective for boosting the performance of a pre-trained language model. However, in the task of question generation, question generators post-trained with a well-designed training objective show poor performance without sufficient training examples. To handle this problem, this paper proposes a novel post-training for question generation which adopts a data augmentation technique to increase the number of training examples as well as post-training objectives. As post-training objectives, this paper introduces a new training objective, *wh-words deletion*, in addition to the well-known *question infilling*. Moreover, this paper employs back-translation techniques to increase the number of instances for post-training. To prove the effectiveness of the proposed method, this paper applies the post-training strategies to T5, a large-scale pre-trained language model, on SQuAD-QG. The experimental results demonstrate that the proposed post-training is helpful for enhancing the performance of answer-aware question generation.

Keywords Question generation · Deep learning · Natural language processing · Post-training

G.-M. Park · C. S. Hong · S.-B. Park (✉)
Department of Computer Science and Engineering, Kyung Hee University, Yongin, South Korea
e-mail: sbpark71@khu.ac.kr

G.-M. Park
e-mail: pgm1219@khu.ac.kr

C. S. Hong
e-mail: cshong@khu.ac.kr

S.-E. Hong
Department of Software Convergence, Kyung Hee University, Yongin, South Korea
e-mail: zen152@khu.ac.kr

1 Introduction

Questions play a critical role in acquiring new knowledge and confirming information. Therefore, question generation (QG) is used in diverse applications to construct question answering systems or dialogue systems, and to augment data for machine reading comprehension [1–4]. It is important to not only capture information relevant with a question but also generate appropriate interrogative sentences for generating good questions automatically. There are two types in text-based question generation tasks: answer-aware and answer-agnostic question generation. This paper focuses on answer-aware question generation.

In recent years, with development of deep neural networks (DNNs), language models pre-trained with a large-scale corpus have achieved prominent performance in natural language understanding and generation tasks [5, 6]. Although question generation models fine-tuned with a pre-trained language model (PLM) and a data set for the target task also have shown meaningful performances even in this task, fine-tuning is not enough to optimize PLMs for question generation due to the gap of pre-training and fine-tuning objectives and low resources to adapt to a target domain (question generation).

One simple solution to this problem is to augment qualitative training data for boosting performance on the target task. However, augmenting data using human annotation is time-consuming and expensive. An alternative for human annotation is to generate training data set automatically. However, it is difficult to gain better performance in the target task by simply adding augmented data at the fine-tuning stage. Therefore, a well-designed training method is required to enhance the performance of question generator.

Another solution is the post-training which helps a pre-trained model adapt to a target task with new training objectives and target training data. The post-training is applied between pre-training and fine-tuning stages. For example, in the dialogue response selection task, since vanilla BERT is not trained to experience dialogue data, post-training BERT with domain-specific data can improve the dialogue response selection [7, 8]. Although existing post-training algorithms have shown prominent improvements in diverse tasks, they require a large volume of training examples to adapt a language model to a target domain with post-training objectives. Moreover, it is nearly impractical in the practical environments to increase training instances with external data.

To handle these issues, this paper proposes a new post-training method with data augmentation to enhance the performance for English answer-aware question generation. The proposed method adopts *question infilling* [9] and *wh-words deletion* as its post-training objectives. *Question infilling* is a traditional post-training objective to learn the important parts of a context and is known to be relevant to question generation. *Wh-words deletion* is a new post-training objective for improving the performance of question generation and it helps a model learn appropriate interrogative pronouns considering a context and an answer. To augment training instances

for post-training, the back-translation is adopted which translates a back-translated text into a text in the original language.

T5 [10] is adopted as a pre-trained language model since it shows prominent performances in many natural language generation tasks. The effectiveness of the proposed method is verified through the experiments on SQuAD-QG [11]. The experimental results show that the proposed post-training algorithms are plausible and effective to boost the performance in question generation.

The rest of this paper is organized as follows. Section 2 introduces previous work about the limitations of pre-trained language models and post-training methods, and Sect. 3 formulates the answer-aware question generation based on fine-tuning T5. Section 4 explains the proposed post-training objectives, *question infilling* and *wh-words deletion*, and data augmentation with back-translation data, and Sect. 5 shows the experimental results of the proposed post-training method. Finally, Sect. 6 draws the conclusions.

2 Related Work

Early work on QG mostly studied template-based and rule-based approaches. Recently, sequence-to-sequence neural models have been adopted to generate diverse and fluent questions and have shown better performance than rule-based approaches. Moreover, most recent studies use large-scale language models pre-trained with a large-scale text-based corpus for QG and the models achieve the state-of-the-art (SOTA) performance in QG task. For example, T5 shows a prominent performance not only in natural language understanding tasks but also in generation tasks by adopting the text infilling objective and converting the data format for text-to-text. ERNIE-GEN [12] adopts the multi-flow attention architecture to incorporate infilling generation and a noise-aware generation, and then achieved the SOTA performance in SQuAD-QG. However, these pre-trained language models have a fundamental issue that they did not acquire the knowledge about question generation from the pre-training stage due to the lack of training examples and objectives for question generation and thus could not be optimized fully for the task.

One solution to this issue is to inject the knowledge for question generation to a pre-trained language model. However, constructing the knowledge is time-consuming and cost-ineffective in the real world. Another solution is to generate a new large-scale pre-trained language model for question generation. Unfortunately, generating a new language model is also a challenge that requires a huge training examples and resources to pre-train the language model.

3 Answer-Aware Question Generation

Answer-aware question generation is a task that generates natural language questions based on a context and an answer. This paper adopts T5 as a question generator. T5 has an encoder-decoder architecture of the transformer and is pre-trained with the text infilling objective that fills a missing part of a text. It is fine-tuned by minimizing the negative log likelihood

$$L = - \sum_{i=1}^N \log P(q_i | C_i, A_i; \theta), \quad (1)$$

where C_i is a context, A_i is an answer span, q_i is a question, and θ is the parameters of T5.

4 Post-training with Data Augmentation

This paper proposes two training objectives to post-train T5 for enhancing the performance of question generation. The first objective is *question infilling* that corrupts a question q and makes T5 capture the important parts of a context C considering the answer A to reconstruct the original question q from corrupted question \hat{q}' . Learning *question infilling* is formulated with a loss function

$$L_{qi} = - \sum_{i=1}^M \log P(q_i | C, A, \hat{q}'_{qi}; \theta), \quad (2)$$

where the size of post-training set is M . The second is a new post-training objective *wh-words deletion* which considers a characteristic of English interrogative sentences. Wh-words are widely used to generate questions and directly related an answer type. Thus, it forces T5 to learn which wh-word is proper for a given input. *Wh-words deletion* is formulated with a loss function

$$L_{wd} = - \sum_{i=1}^M \log P(q_i | C_i, A_i, \hat{q}'_{wd}; \theta), \quad (3)$$

where \hat{q}'_{wd} is a corrupted question without wh-words. Then, the final loss function to post-train the question generator becomes

$$L_{\text{post}} = L_{qi} + L_{wd}.$$

Large-scale language models based on the transformer architecture are sensitive to the number of training examples. To supplement the post-training instances, the

back-translation technique is employed. That is, the questions in the original data set are first translated to Chinese, and then the translated questions are re-translated back into a source language. The augmented questions are added to a set of original questions as new training examples. As a result, the number of post-training examples with data augmentation is much larger than that of the original training data set.

5 Experiments

5.1 Experimental Settings

The proposed post-training method is verified with the SQuAD-QG dataset transformed from SQuAD 1.1 dataset which contains 21,135 contexts. Each context has multiple answers and questions. For answer-aware question generation, a context and an answer adopted as input of a question generator, and a question is the output of the question generator. In the post-training stage, a context, an answer, and a corrupted question are given as an input for T5 and an original question is used as an output. SQuAD-QG has 75,722 examples for training data, 10,570 examples for validation data, and 11,877 examples for testing data. Back-translation is used to make additional post-training examples and the target language is Chinese. The number of total examples for post-training is then 151,144.

T5 is post-trained with the batch size of one and the sequence length of 256. It is fine-tuned with same batch size and sequence length. For inference, a beam search algorithm is used with the beam size of five. For both post-training and fine-tuning, AdamW optimizer with the cosine warm-up scheduler is used. The initial learning rate is $3e-5$. All experimental results are evaluated by automatic evaluations that are BLEU-4, ROUGE-L, and METEOR.

5.2 Experimental Results

Table 1 shows the effects of post-training. The T5 post-trained with *question infilling* (QI) and *wh-words deletion* (WD) objectives achieves 22.619 of BLEU-4, 51.398 of ROUGE-L, 26.285 of METEOR, while pre-trained T5 achieves just 22.289 of BLEU-4, 51.069 of ROUGE-L, and 26.126 of METEOR. These results prove that the proposed post-training method is effective to boost up the performance of question generator.

This paper employs the objective of *question infilling* (QI) and proposes *wh-words deletion* (WD) as an additional objective and data augmentation (DA) with back-translation for T5. Without data augmentation, the T5 post-trained only with the two objectives shows lower performance than vanilla T5. This result implies that the number of examples affects the performance of question generation.

Table 1 The performances measured with automatic evaluation metrics on SQuAD-QG

Model	BLEU-4	ROUGE-L	METEOR
Pre-trained T5	22.289	51.069	26.126
+ QI	21.622	50.145	25.385
+ QI, WM	21.698	50.328	25.512
+ QI, WM, DA (proposed model)	22.619	51.398	26.285

6 Conclusions

This paper proposed a novel post-training strategy to boost up question generation. The proposed training objectives, *question infilling*, and *wh-words deletion* allow the question generator to learn the structure of questions and to experience question generation indirectly. Data augmentation with back-translation to increase post-training examples is also shown to be important to improve the effectiveness of the proposed post-training objectives. The experimental results demonstrate that the proposed post-training method is plausible and effective for question generation.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A4A1018607) and Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-00155911), Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University).

References

1. Du X, Shao J, Cardie C (2017) Learning to ask: neural question generation for reading comprehension. In: Proceedings of the 55th annual meeting of the association for computational linguistics, pp 1342–1352
2. Duan N, Tang D, Chen P, Zhou M (2017) Question generation for question answering. In: Proceedings of the 2017 conference on empirical methods in natural language processing, pp 866–874
3. Ghanem B, Coleman L, Dexter J, Ohe S, Fyshe A (2022) Question generation for reading comprehension assessment by modeling how and what to ask. In: Findings of the association for computational linguistics, pp 2131–2146
4. Gu J, Mirshekari M, Yu Z, Sisto A (2021) ChainCQG: flow-aware conversational question generation. In: Proceedings of the 16th conference of the European chapter of the association for computational linguistics, pp 2061–2070
5. Chan Y, Fan Y (2019) A recurrent BERT-based model for question generation. In: Proceedings of the 2nd workshop on machine reading for question answering, pp 154–162
6. Dong L, Yang N, Wang W, Wei F, Liu X, Wang Y, Gao J, Zhou M, Hon H (2019) Unified language model pretraining for natural language understanding and generation. *Adv Neural Inf Process Syst* 32:13042–13054
7. Han J, Hong T, Kim B, Ko Y, Seo J (2021) Fine-grained post-training for improving retrieval-based dialogue systems. In: Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics, pp 1549–1558

8. Whang T, Lee D, Lee C, Yang K, Oh D, Lim H (2020) An effective domain adaptive post-training method for BERT in response selection. In: Proceedings of the annual conference of the international speech communication association, pp 1585–1589
9. Park G, Hong S, Park S (2022) Post-training with interrogative sentences for enhancing BART-based Korean question generator. In: Proceedings of the 2nd conference of the Asia-Pacific chapter of the association for computational linguistics and the 12th international joint conference on natural language processing
10. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu JP (2019) Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint [arXiv:1910.10683](https://arxiv.org/abs/1910.10683)
11. Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) Squad: 100,000+ questions for machine comprehension of text. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 2383–2392
12. Xiao D, Zhang H, Li Y, Sun Y, Tian H, Wu H, Wang H (2020) ERNIEGEN: an enhanced multi-flow pre-training and fine-tuning framework for natural language generation. In: Proceedings of the 29th international joint conference on artificial intelligence, pp 3997–4003

A Memory Replay-Based Continual Learning Utilizing Class Representative and Class Boundary Data



Jun-Beom Lee , Choong Seon Hong , and Seong-Bae Park 

Abstract It is known that deep learning models are performing beyond human beings if sufficient size of training data is provided. However, this is usually true only if a model is trained for a single task. If a model is continuously trained on multiple tasks, the model abruptly forgets the previously learned information. To overcome this phenomenon, various continual learning methods have been proposed. In particular, in the class-incremental learning scenario in which no task identification information is given at the inference time, the memory replay method has shown good performances. In general, the memory replay method stores class representative samples to recall previously learned tasks, but this leads to poor performance when a sample can be a representative of multiple classes. Therefore, this paper proposes a method that stores class boundary data as well as class representative data in the memory buffer to improve the performance of the memory replay. For this, a class representative sample is defined as the one of which feature is close to the mean of a certain class, and a class boundary sample is defined to be located between the means of two different classes. The experiments confirm that the proposed method shows higher performance than existing methods, which proves the importance of utilizing class boundary samples in continual learning.

Keywords Memory replay · Class representative data · Class boundary data

J.-B. Lee · C. S. Hong · S.-B. Park (✉)
Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104,
South Korea
e-mail: sbpark71@khu.ac.kr

J.-B. Lee
e-mail: jblee9410@khu.ac.kr

C. S. Hong
e-mail: cshong@khu.ac.kr

1 Introduction

If the sufficient size of training dataset is given and the model capacity is sufficient, artificial neural network models are known to show higher performance than human beings. For instance, AlexNet [1] firstly achieved lower image classification error rate than human performance in the ImageNet classification task. However, the artificial neural network models including deep learning tend to perform poorly when learning multiple tasks sequentially. For example, human beings can classify cats and dogs without difficulty, even if they learn to classify cars and trees after learning to classify cats and dogs. However, if the artificial neural network model is trained at the same order, the model forgets to classify cats and dogs. This is because the parameters optimized for the previous task are completely changed for a new task. This phenomenon is known as *catastrophic forgetting* [2], and various continual learning methods have been studied to alleviate it [3].

Continual learning methods are roughly classified into memory replay-based, dynamic architecture-based, and regularization-based methods. First, the memory replay-based method is storing the data of a previous task in a memory buffer and training both the stored data and the data for a new task together when the model learns a new task. Second, the dynamic architecture-based method is a method in which the structure of a model is modified according to the task. The last is a regularization-based method, which adds a regularization term to the loss function to force the difference between the values of parameters for the previous task and those for a new task to be small. In addition, continual learning scenarios are categorized into two types by the use of task identification information [4]. The first is the task-incremental learning scenario in which the identification information for a task is given both at the training and inference stages, and the second is the class-incremental learning scenario in which the identification information for a task is not given at the inference stage.

This paper deals with memory replay method that has recently shown good performance in class-incremental learning scenario. The challenge of the memory replay method is that as the number of tasks increases, the size of the memory buffer also increases. To solve this problem, existing methods commonly fix the size of the memory buffer and store only a few class representative data to the memory buffer. The iCaRL [5] extracts the features of data belonging to a specific class through a feature extractor, sets their mean as a class representative feature, and stores the data to the memory buffer in the order of how close a data instance is to the class representative feature.

On the other hand, some research on memory replay without storing data from the previous tasks also has been conducted. Deep generative replay [6] used a generative adversarial network (GAN)-based data generator that can mimic data from the previous tasks. It is designed to apply continual learning to the situations where it is impossible to store data from the previous tasks. When a classifier learns a new task, it learns pseudo old data generated by the trained data generator together to prevent forgetting the old tasks. Deep generative replay has a strong advantage of

not having to store data from the previous tasks, but training a data generator is another challenge, and nevertheless, the catastrophic forgetting can still occur.

This paper proposes a method for determining whether or not a data instance should be stored in a memory buffer for memory replay with a limited-sized memory buffer. First of all, in order to mitigate the catastrophic forgetting in sequential learning, a classifier is re-trained with class boundary data that are located between classes. For this, the average of two class representative features is regarded as a class boundary and the data of which feature is close to the class boundary are regarded as class boundary data. In addition, the proposed method improves the classification performance by utilizing the class boundary data even in the test stage. That is, in the test stage, the class prediction of a test sample is determined by voting of the model prediction for the test sample and the classes of its k nearest class boundary neighbors. The experiments on the split-MNIST dataset consisting of five binary classification tasks confirm the plausibility of the proposed method.

2 Proposed Method

This paper proposes a memory replay method that utilizes both class representative and class boundary data under the learning scenario in which binary image classification tasks are given sequentially. The proposed memory manages two memory buffers. One is a representative memory buffer for storing class representative data, and the other is a boundary memory buffer for storing class boundary data. The size of each buffer is fixed as k . The image classifier which was used for this method has three convolutional layers and a fully connected layer. The convolutional layers extract features from input images, and a fully connected layer performs classification based on the features. It is assumed that a total of n image classification tasks exist, the training data of the i -th task are represented as $D_i^{\text{train}} = (X_i^{\text{train}}, Y_i^{\text{train}})$, and the test data of the i -th task are represented as $D_i^{\text{test}} = (X_i^{\text{test}}, Y_i^{\text{test}})$.

In the training stage, the image classifier is first trained using the D_i^{train} . After training the image classifier, the convolutional layer of the trained classifier is fixed, and the image features are extracted. Figure 1 shows the training stage of the image classifier and the extraction stage of the image feature for the first task of the split-MNIST dataset. In this figure, the blue dots represent the features of the data of the first class, and the red dots represent the features of the data of the second class.

Thereafter, the average of each class feature is calculated, and it is set as a class representative feature. As shown in the upper left of Fig. 2, the representative feature of the first class of the i -th task is indicated by P_1^i , and the representative feature of the second class of the i -th task is indicated by P_2^i . Then, the L_2 -distance between the class representative feature and the feature of each image in that class is calculated, and the images are stored in the representative memory buffer in the order of small L_2 -distance. In addition, the average of the two class representative features is indicated as $B_{1,2}^i$ as shown in the right side of Fig. 2. After the L_2 -distance between $B_{1,2}^i$ and the feature of each image in each class is calculated, the images are stored in the

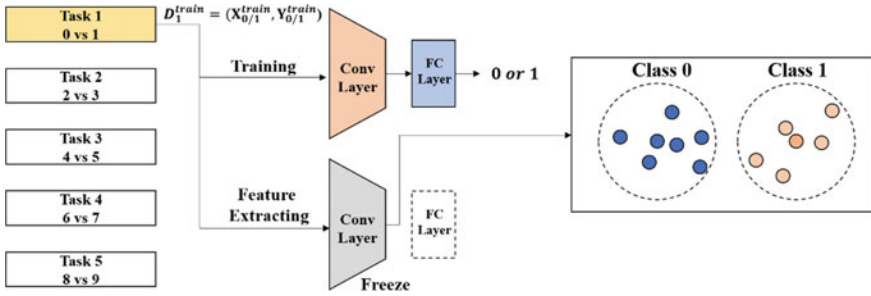


Fig. 1 Training and feature extraction for the first task

boundary memory buffer in the order of small $L2$ -distance. Since the size of memory buffer is k , $k/2i$ data for each class are stored in the memory buffer.

The data in both memory buffer are re-trained during training the classifier for the next task, and the data from the boundary memory buffer are utilized at the test stage of the current task. At the test stage, the features of all data in the boundary memory buffer are extracted using the trained and fixed convolutional layer. After that, the $L2$ -distance between these features and the feature of the data sample to be tested is calculated, and the two data with the smallest distance are selected from the boundary memory buffer. Then finally, the class for the test sample is predicted by majority voting of the class labels of the two selected boundary data and the predicted classification result of the test sample.

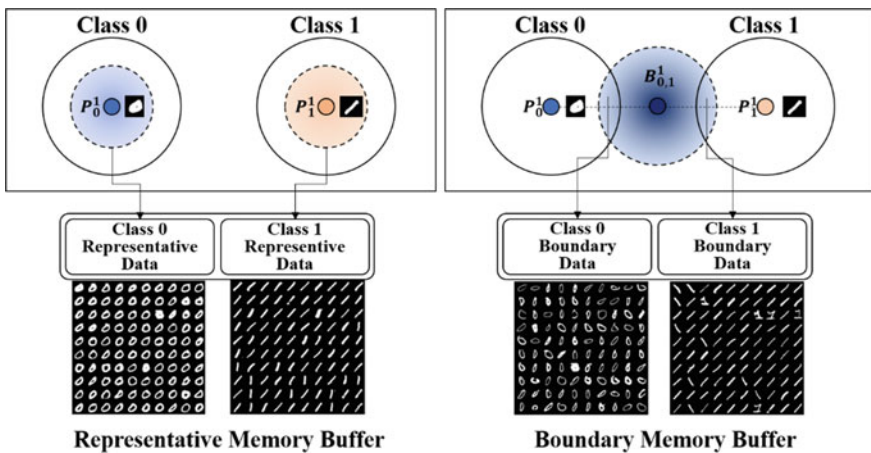


Fig. 2 Class representative data and class boundary data of the first task

Table 1 Data distribution of split-MNIST

Task ID	1	2	3	4	5
Train	12,665	12,089	11,263	12,183	11,800
Test	2115	2042	1874	1986	1983

3 Experiments

We use split-MNIST [7] to compare the performance of the proposed method and the baselines. Split-MNIST is a dataset that separates the original MNIST dataset into five different binary classification tasks for continual learning scenario. The original 28×28 grayscale images were used without any processing. There are about 6000 training images and about 1,000 test images per digit, and the data distribution by task can be found in Table 1.

In this paper, iCaRL, deep generative replay, and random sampling method are used as baselines. iCaRL is a representative method of memory replay-based continual learning. It actually uses class representative data for replay. The deep generative replay uses an image generator to mimic and store the images of old tasks. The random sampling method is a naive memory replay method. It stores randomly selected data in the memory buffer for replay. For fair comparison, all the baselines and the proposed method equally set the number of the replay data as 1000.

The performance of the baselines and the proposed method were compared by measuring the final accuracy. The final accuracy refers to the accuracy of the model on the combined test sets of the five tasks. For the experiments, a simple image classifier consisting of three convolutional layers and one fully connected layer was used.

Table 2 shows the final accuracies of each method. *None* is the result of training the image classifier sequentially on five tasks without using any continual learning method, and thus, its accuracy can be considered as a lower bound. *Joint* is the result of training the image classifier with the entire training set of five tasks, and thus, its accuracy is an upper bound. According to Table 2, the proposed method shows lower performance than *Joint*, but outperforms all baselines. That is, this fact proves empirically that using class representative data as well as class boundary data for memory replay helps alleviate catastrophic forgetting.

4 Conclusion

This paper has proposed a novel continuous learning method based on memory replay. It mitigates the catastrophic forgetting of a learning model in the training scenario of sequential task-training. The existing memory replay methods usually store the class representative data in a memory buffer and use them for training a classifier for the next task. On the other hand, the proposed method does not

Table 2 Experimental results

Method	Final accuracy (%)
None (lower bound)	18.55
Random	91.94
DGR	92.82
iCaRL	92.24
Ours	92.91
Joint (upper bound)	98.95

Bold emphasize the performance of our proposed method, except the upper bound method

store only the class representative data in a memory buffer, but also stores the class boundary data located between the classes. These class boundary data are believed to be difficult for a model to classify correctly. In addition, by utilizing the class boundary data as a member for k -NN classification in the test stage, the classification performance for the data samples near class boundaries is improved. The superiority of the proposed method was confirmed using the split-MNIST dataset that is made by splitting original MNIST dataset into five binary classification tasks.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1A4A1018607) and Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (No. RS-2022-00155911), Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University).

References

1. Krizhevsky A, Sutskever I, Hinton G (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
2. McCloskey M, Cohen J (1955) Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol Learn Motiv* 24(1):109–165
3. De Lange M, Aljundi R, Masana M, Parisot S, Jia Xu, Leonardis A, Slabaugh G, Tuytelaars T (2022) A continual learning survey: defying forgetting in classification tasks. *IEEE Trans Pattern Anal Mach Intell* 44(7):3366–3385
4. Gido V, Andreas T (2018) Three continual learning scenarios and a case for generative replay. In: *NeurIPS workshop on continual learning*. Montreal
5. Rebuffi A, Kolesnikov S, Lampert H (2017) iCaRL: incremental classifier and representation learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2001–2010
6. Hanul S, Jung L, Jaehong K, Jiwon K (2017) Continual learning with deep generative replay. In: *Advances in neural information processing systems*, pp 2994–3004
7. Zenke F, Poole B, Ganguli S (2017) Improved multitask learning through synaptic intelligence. *arXiv preprint arXiv:1703.04200*

Automatic Classification for Representative Spatio-temporal-Based Event Document Using Machine Learning



Byoungwook Kim, Yeongwook Yang, Ji Su Park, and Hong-Jun Jang

Abstract As the use of social media and online news media increases, the amount of text data online is rapidly increasing. As the use of spatio-temporal information increases, research on extracting spatio-temporal information from text data and detecting events is being actively conducted. However, the document also contains insignificant and trivial spatio-temporal information that is not relevant to the subject. Therefore, it is important to extract the spatio-temporal-based events that describe the core topic of the document. In this study, spatio-temporal-based events related to the core topic of the document are defined as 'representative spatio-temporal-based events'. We propose a machine learning-based document classifier to classify documents containing representative spatio-temporal-based events. To train the proposed document classifier, guidelines for building training data were prepared, and 10,000 training data were constructed. Experimental results showed that the accuracy of traditional machine learning was not high.

Keywords Representative spatio-temporal information · Event extraction

B. Kim

Department of Computer Science and Engineering, Dongshin University, Naju, South Korea
e-mail: bwkim@dsu.ac.kr

Y. Yang

Division of Computer Engineering, Hanshin University, Osan, South Korea
e-mail: yeongwook.yang@hs.ac.kr

J. S. Park · H.-J. Jang (✉)

Department of Computer Science and Engineering, Jeonju University, Jeonju-si, South Korea
e-mail: hongjunjang@jj.ac.kr

J. S. Park

e-mail: jisupark@jj.ac.kr

1 Introduction

The amount of text data is increasing exponentially due to the increase in the number of social network service (SNS) users, the rapid increase in SNS data due to the spread of smartphones, and the increase in the number of online news media. In social trend analysis, it is important to extract event sentence from documents composed of text [1]. An event sentence is a sentence that expresses information about a certain event on a specific subject. Information about an event includes the place where the event occurred, who caused the event, and the time the event occurred. In many cases, information about events is used in policy decision-making. For example, a disease prevention policy can be established through information that a disease has occurred in a specific country in a news article [2]. As such, online news articles include important time information or spatial information that can be used to determine the direction of policy by analyzing social trends and to establish marketing strategies for companies.

Existing studies have focused on improving the accuracy of spatio-temporal information extraction from text documents such as online news articles [3–5]. Documents can contain various event information. In this study, among the various events included in the document, spatio-temporal information explaining the core topic of the document, the subject of the event, and the event content are defined as ‘representative spatio-temporal-based events’. Documents containing representative spatio-temporal-based events are defined as ‘representative spatio-temporal-based event documents’. Extracting a large number of general spatio-temporal information as well as representative spatio-temporal information from a single document may reduce the accuracy of key event analysis based on spatio-temporal information. Extracting representative spatio-temporal-based events from a single document is an expensive operation, making it difficult to treat all documents for analysis. To efficiently analyze key events through representative spatio-temporal events, it is important to classify representative spatio-temporal-based event documents.

In this paper, we propose a representative spatio-temporal-based event document classifier through machine learning. To train the document classifier, 10,000 training data were constructed from news articles provided by the National Institute of the Korean Language [6]. Seven workers read the news article and get the gist of the article. In relation to the core topic of the article, four types of event information (who, what, when, where) were found, and labeling was performed. In order to maintain the quality of learning data for seven workers, guidelines for building learning data were produced. A representative spatio-temporal-based event document classifier was trained with three machine learning algorithms (Gaussian naïve Bayes, random forest, and linear SVM). The final goal of our study is to extract representative spatio-temporal-based events from large volumes of documents. In order to extract the representative spatio-temporal-based event, it is necessary to first classify the representative spatio-temporal-based event document from among a plurality of documents. This paper corresponds to the representative spatio-temporal-based event document classification stage.

The contributions of this paper are as follows:

- We defined a novel problem to classify representative spatio-temporal-based event documents that describe the document's key topic.
- We developed 100,000 training data to train representative spatio-temporal-based events documents classifier. In order to maintain the quality of learning data for seven workers, guidelines for building learning data were produced.
- We proposed a representative spatio-temporal-based event document classifier trained on data constructed with three machine learning algorithms (Gaussian naïve Bayes, linear SVM, and random forest).

2 Related Work

2.1 Textual Event Extraction

An event is an occurrence of a happening [7, 8]. It may be related to a specific date or time, place, people, objects, or other events. Event extraction is one of the most important branches of knowledge extraction from textual data. Event-oriented models provide effective requirements to build decision support systems or question answering models [9]. This branch started in the healthcare field for bacterial behavior analysis [10], and with the progressive advancement of big data, the event extraction field gained popularity to enable information extraction from the exponentially growing heterogeneous textual data collections. From social media posts to web pages, news messages, articles, or researchers, event-oriented knowledge discovery enables event identifications along with the related entities that play a role in those events [11, 12].

2.2 Machine Learning for Event Extraction

Document classification is being researched in the direction of utilizing machine learning, in which a computer automatically understands the contents of a document and classifies it into a predetermined category, rather than a person reading and classifying the document. Classifying spam in email is an early attempt at document classification using machine learning.

One of the early machine learning algorithms used to detect spam mail is Gaussian naïve Bayes (GNB). Gaussian naïve Bayes algorithm is a representative probabilistic machine learning classification algorithm. GNB regards data as independent events and assigns these events to Bayesian theory to classify the label with the highest probability. First, stop words are deleted from the entire document and separated into individual words through preprocessing. Calculate the frequency count of words used in the entire document. Words with low frequency are removed, and words with

a certain frequency or higher are analyzed. The spam or non-spam label is then used to calculate the probability of each word being included in the spam mail.

Random forest is an ensemble machine learning model. Form multiple decision trees, pass new data points through each tree simultaneously, and vote on the results classified by each tree to select the result with the most votes as the final classification result. Random forest can produce accurate and good results in document classification compared to conventional techniques using parameters such as the maximum likelihood method and machine learning algorithms such as single decision trees and neural networks.

Support vector machine (SVM) is one of the fields of machine learning and is used for pattern recognition, supervised learning model classification for data analysis, and regression analysis. SVM is different from existing classification methods in its basic principle. Existing methods, including neural networks, are designed for the purpose of minimizing the error rate of classification. SVMs are designed to go a step further and maximize the white space that exists between the two classes.

3 Materials and Methods

3.1 Data Sources

We used the published Korean corpus by National Institute of Korean Language [6] to build data to train a representative spatio-temporal-based event document classifier. The National Institute of the Korean Language provides various Korean corpus for research purposes, and we used the newspaper corpus among them. The newspaper corpus provided by the National Institute of the Korean Language consists of a total of 3,536,491 news articles from 2009 to 2018. Raw data contains article content in the document tag.

3.2 Data Preprocessing

We constructed representative spatio-temporal-based event document data for 10,000 documents among the entire corpus. Workers read news articles directly and labeled news articles containing representative spatio-temporal-based event information. The quality of training data is important because the quality of training data determines the performance of machine learning. Since the quality of learning data can be different for each worker, we conducted a task to increase the consistency of data quality among workers. Consistency was maintained by reexamining the labeled data between workers.

Table 1 Result of three machine learning classifiers

Machine learning	Precision	Recall	F1-score	Accuracy
Gaussian naïve bayes	0.587	0.278	0.348	0.767
Linear SVM	0.643	0.453	0.526	0.793
Random forest	0.738	0.188	0.324	0.781

3.3 Machine Learning

We tried to classify representative spatio-temporal-based event documents through machine learning. The machine learning algorithms used for document classification are Gaussian naïve Bayes, linear SVM, and random forest. For our experiments, we implemented scikit-learn, which is provided as a Python machine learning library.

4 Evaluation

We compared the three machine learning classifiers (Gaussian naïve Bayes, random forest, and linear SVM). The experimental results are presented in Table 1.

The accuracy was derived from a minimum of 0.767 to a maximum of 0.793. The three machine learning accuracies are lower than the average accuracy of general document classification machine learning. Linear SVM is 0.526, indicating the highest *F1*-score. However, the *F1*-score is low in all machine learning. It can be seen that representative spatio-temporal-based document classification is a difficult problem to classify using general machine learning.

5 Conclusion

In this study, we defined spatio-temporal-based events related to the core subject of the document as ‘representative spatio-temporal-based events’. We propose a machine learning-based document classifier to classify documents containing representative spatio-temporal-based events. In order to train the proposed document classifier, guidelines for building training data were prepared, and 10,000 training data were constructed. We trained a representative spatio-temporal-based event document classifier with three machine learning classifiers (Gaussian naïve Bayes, random forest, and linear SVM). Experimental results show that it is difficult to classify representative spatio-temporal-based event documents by traditional machine learning. In order to improve the performance of the document classifier through future research, deep learning such as convolutional neural network (CNN) or recurrent neural network (RNN) should be applied.

Acknowledgements This research was funded by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIT) (No. 2021R1F1A1049387). This result was supported by the ‘Regional Innovation Strategy (RIS)’ through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (1345341782).

References

1. Naughton M, Stokes N, Carthy J (2010) Sentence-level event classification in unstructured texts. *Inf Retr* 13:132–156. <https://doi.org/10.1007/s10791-009-9113-0>
2. Lan R, Adelfio MD, Samet H (2014) Spatio-temporal disease tracking using news articles. In: Proceedings of the HealthGIS’ 14: 3rd ACM SIGSPATIAL international workshop on the use of GIS in public health, Dallas, TX, USA, 4–7 Nov 2014, pp 31–38. <https://doi.org/10.1145/2676629.2676637>
3. Badia A, Ravishankar J, Muezzinoglu T (2007) Text extraction of spatial and temporal information. In: Proceedings of the 2007 IEEE intelligence and security informatics. New Brunswick, NJ, USA, 23–24 May 2007. <https://doi.org/10.1109/isi.2007.379527>
4. Lim C-G, Jeong Y-S, Choi H-J (2019) Survey of temporal information extraction. *J Inf Process Syst* 15:931–956. <https://doi.org/10.3745/JIPS.04.0129>
5. Ferial A, Kholadi M-K (2015) Automatic extraction of spatio-temporal information from arabic text documents. *Int J Comput Sci Inf Technol* 7:97–107. <https://doi.org/10.5121/ijcsit.2015.7507>
6. National Institute of Korean Language [Internet]. Available online: <https://www.korean.go.kr>. Accessed on 9 Apr 2022
7. Moens M, Steedman M (1988) Temporal ontology and temporal reference. *Comput Linguist* 14(2):15–28
8. Hogenboom F, Frasinca F, Kaymak U, de Jong F, Caron E (2016) A Survey of event extraction methods from text for decision support systems. *Decis Support Syst* 85:12–22
9. Duan X, Tang S, Zhang S, Zhang Y, Zhao Z, Xue J, Zhuang Y, Wu F (2018) Temporality-enhanced knowledge memory network for factoid question answering. *Front Inform Technol Electron Eng* 19(1):104–115
10. Chen Y, Ding Z, Zheng Q, Qin Y, Huang R, Shah N (2020) A history and theory of textual event detection and recognition. *IEEE Access* 8:201371–201392
11. Mejri M, Akaichi J (2017) A survey of textual event extraction from social networks. *LPKM*
12. Zhou H, Yin H, Zheng H, Li Y (2020) A survey on multi-modal social event detection. *Knowl-Based Syst* 195:105695

Case Monitoring for Secondary Battery Transfer with Enhanced Fire Prevention and Stability



Kihak Lee, Heji Lee, Md. Jahirul Islam, Saetbyeol Kim, Wolyoung Kim, Tae-Young Byun, and Bonghwan Kim 

Abstract The use of batteries today is constantly growing. Therefore, it is crucial to safely charge and manage usable and trash batteries. In this paper, a case monitoring system is designed for secondary battery transfer with enhanced fire prevention stability. A microprocessor and ESP32-CAM (video streaming and face recognition with Arduino IDE) equipped with a Wi-Fi module were installed to enable remote monitoring in real time. In addition, the system was constructed by connecting a temperature and humidity sensor for case environmental monitoring, an infrared flame sensor that can detect flames when a flame occurs, a gas sensor for detecting gas, and a GPS module that can check the transfer path of a powder fire extinguisher and a waste battery in real time. This system is connected to the Wi-Fi or LTE network of the carrier (aircraft, ship, and truck) where the transport case of the waste battery is loaded, so that real-time GPS data and images in the case of flames are transmitted to the telegram messenger in real time.

Keywords Case monitoring · Fire extinguisher · Remote monitoring

1 Introduction

Both homes and offices use batteries more and more frequently. Battery technology is advancing as a result of its effectiveness and absence of greenhouse gas emissions, whether they are used to power industrial robots, drones, phones, tablets, solar storage, electric vehicles, or other devices. These batteries can, however, burn if they

K. Lee · H. Lee · Md. J. Islam · B. Kim (✉)

School of Electronic and Electrical Engineering, Daegu Catholic University, Gyeongsan, Gyeongbuk 38430, South Korea
e-mail: bhkim@cu.ac.kr

S. Kim · W. Kim

Solarlightkorea Co. Ltd., Daegu 41061, South Korea

T.-Y. Byun

School of Computer Software, Daegu Catholic University, Gyeongbuk 38430, South Korea

are overcharged or if there is a problem with how they were made, stored, or handled [1]. In the last five years, the number of structure fires brought on by a battery, power-pack, or charger has nearly doubled [2]. A lithium-ion battery can overheat and ignite a fire that has the ability to completely destroy a home in a matter of minutes if the battery is overcharged, defective, or improperly made. Battery fires present a unique problem and increased risk to individuals who are attempting to put out the fire, in contrast to other fires that may be started by flammable liquids, combustibles, or electrical faults. After the battery fails, there could be an explosion, fire, hazardous gas leak, and jet flames [3, 4]. In the case of waste batteries, since they are transported using ships, aircraft, and trucks, the safety of the network for data transmission must be ensured during the transport process.

2 Design and Experiment

2.1 Monitoring Circuit

Here, the ESP32-CAM board, which is an image capturing microprocessor that can be controlled remotely using an internet network and monitors the environment in a real-time case, is selected as a controller. The ESP32-CAM board is equipped with an OV2640 camera module. It also has the advantage of being able to store data because it has an SD card slot built in. However, due to environmental monitoring, flame, and GPS sensors, the use of SD cards is not applicable here due to the limitations of general-purpose input/output (GPIO) pins. The NEO-7M module was applied to check the transportation path of the secondary battery in real time. In order to monitor environmental changes in a case during transportation, Bosch's BME280 temperature and humidity sensor are applied to monitor temperature conditions in the secondary battery case in real time during transportation, so that an alarm can be raised to a long-distance control room and a transport officer. In addition, a flame sensor and a gas sensor are mounted to sense whether a flame or gas is generated in the secondary battery by an overheating reaction, and a powder fire extinguisher mounted on the case is operated to extinguish the flame in an instant and to deliver the incident to the control room and transportation personnel through messenger. Figure 1a, b shows the monitoring of the secondary battery transport circuit and the monitoring board manufactured for fire extinguisher control.

2.2 Lab Tester for Secondary Battery Carrying Case Monitoring

Figure 2 shows the 3D modeling of the lab tester for case monitoring of secondary batteries.

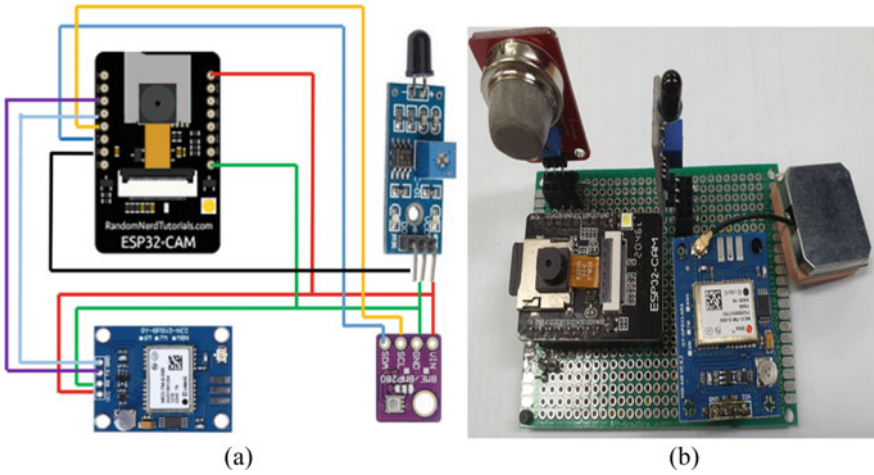


Fig. 1 **a** Secondary battery carrying case monitoring circuit and **b** secondary battery carrying case monitoring board

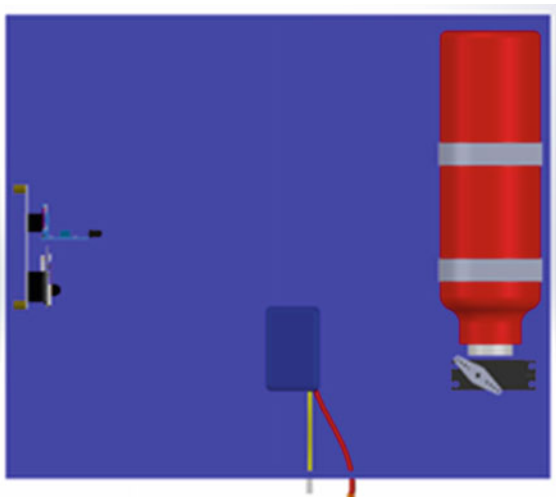


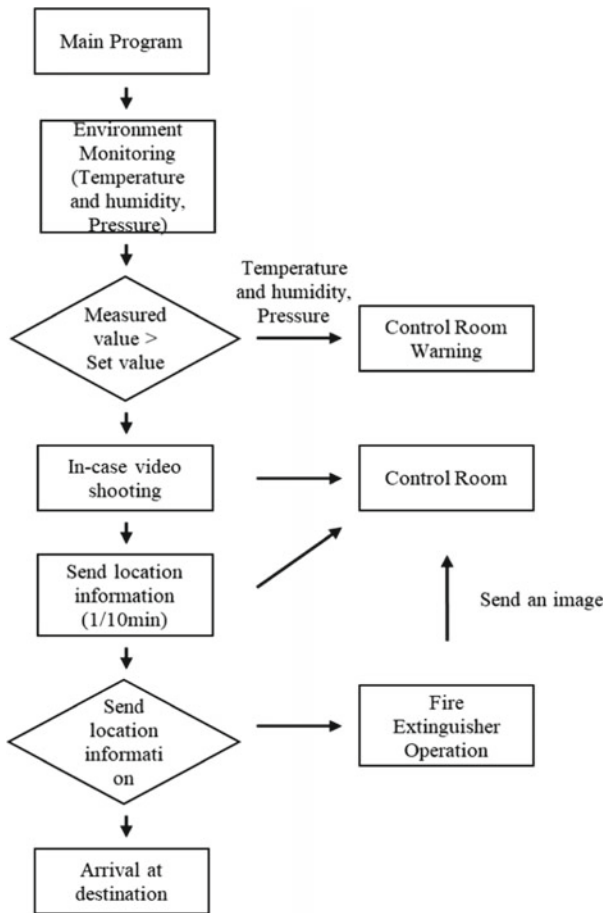
Fig. 2 Secondary battery carrying case monitoring lab tester (3D stereoscopic)

Case configuration.

- Case: Al-Plate (350 mm × 300 mm × 110 mm with plate thickness: 5 mm)
- Monitoring board
- Lithium polymer battery: 700–1500 mAh, 7.4 V
- Powder fire extinguisher: 65 × 210 mm
- Servo motor: MG995.

By connecting voltage-adjustable switch mode power supply (SMPS) to the terminal of the lithium polymer battery and instantaneously applying a voltage (10–15 V) higher than the allowable voltage, the flame denser mounted on the monitoring board detects the flame and rotates the servo motor to eject the extinguishing fluid.

2.3 Secondary Battery Carrying Case Monitoring Process



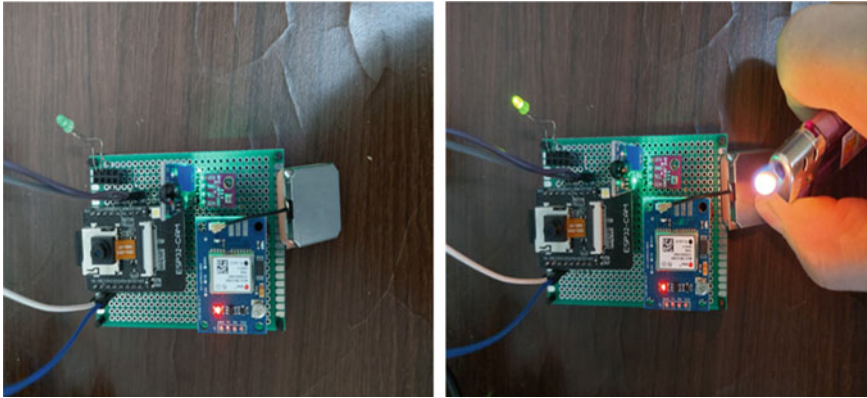


Fig. 3 Comparison of operating conditions before and after flame sensor flame sensing

3 Result and Discussion

In order to confirm the operation of the flame sensor, a flame was arbitrarily generated on the front of the sensor to confirm the response of the sensor. In Fig. 3, the LED was mounted on the servo motor output pin, when there was no flame and the flame occurred, and the on–off state of the LED was checked. When a flame occurs, it is determined that a flame has occurred by sensing a change in the infrared wavelength of the flame sensor, and the power of the general-purpose input/output (GPIO) pins connected to the servo motor is made high (on) to operate the fire extinguisher. Figure 4 illustrates a photograph of a situation in a secondary battery transfer case when a flame occurs and transmits a photo message to a person in charge of the remote-control room. Also, the fact that the flame occurred and the operation of the fire extinguisher were transmitted using the telegram messenger. In addition to regular transmission of sensor and camera images in the case, it has a two-way communication function that allows the remote-control room to request environmental data and images inside the case irregularly as necessary. The temperature, humidity, and gas sensor data are shown in Fig. 5. All the environment data inside the case is saved in MariaDB in real time and can be used for further analysis.

4 Conclusion and Future Research Direction

The overall operating performance of the case monitoring system for transferring the secondary battery resulted in good results, but there was a problem to be improved in the future research process. The experiment is conducted in an open lighting environment, and even if the intensity of the flame is absorbed by the ambient lighting, it is estimated that it will not be affected by the ambient lighting in a sealed case.

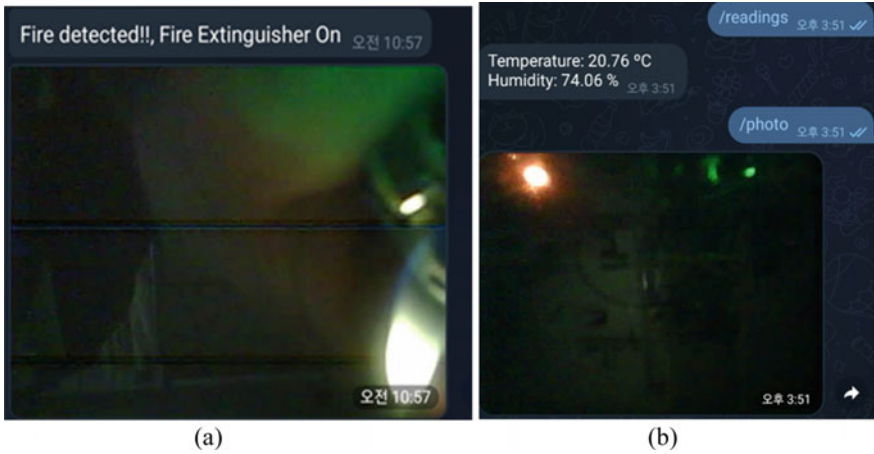


Fig. 4 a Photos sent to remote control room in case of flame and b irregular case monitoring in the remote-control room

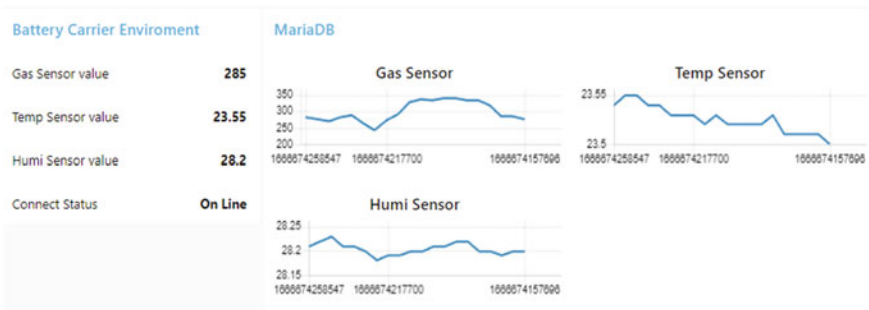


Fig. 5 Temperature, humidity, and gas sensor data

However, as the test case applied in this research has a size of $300 \times 350 \times 110$ mm, it is necessary to check how the effect on the distance between the sensor and the flame generation position affects the flame sensing in the large battery transport case. For this study, a portable powder fire extinguisher was applied. In the case of this fire extinguisher, it was not easy to accurately match the location of the injection port and the location of battery flame generation. To improve these problems, when designing a waste battery transfer case, it is necessary to design a case considering the injection angle of the fire extinguisher.

Acknowledgements This work is financially supported by the Ministry of Trade, Industry and ENERGY (MOTKE) through the fostering project of project for industry-university convergence district build-up.

References

1. Kong L, Li C, Jiang J, Pecht MG (2018) Li-ion battery fire hazards and safety strategies. *Energies* 11(9):2191
2. <https://blog.storemasta.com.au/>
3. Sun P, Bisschop R, Niu H, Huang X (2020) A review of battery fires in electric vehicles. *Fire Technol* 56(4):1361–1410
4. Larsson F, Andersson P, Blomqvist P, Mellander B-E (2017) Toxic fluoride gas emissions from lithium-ion battery fires. *Sci Rep* 7(1):1–13

Effective Use of Knowledge Graphs in a Language Representation Model



Chanwook Min, Hyungsun Na, Yeji Song, Jinhyun Ahn, and Dong-Hyuk Im

Abstract Methods such as knowledge-enabled language representation model (K-BERT) that help train models using external information, such as knowledge graphs, have recently been proposed in the field of natural language processing. However, adding knowledge that does not match the topic when external information is used may hinder training. Hence, a method is required that adds only knowledge matching the topic of the input data by partitioning the external information by topic. Topic-based knowledge graph BERT (TK-BERT), which uses the existing latent Dirichlet allocation (LDA) model to partition the knowledge, does not consider the contextual information. To compensate for this drawback and effectively use external information, this study uses the BERTopic technique to partition the knowledge graph.

Keywords K-BERT · TK-BERT · Knowledge graphs

1 Introduction

With recent advances in deep learning technology, various studies are being conducted on natural language processing [1–3]. Areas of natural language processing include recommendation systems, question and answer tasks, language translation, and text generation. Among deep learning-based natural language processing models, transformer-based pretrained models, such as GPT and BERT, have achieved high performance in various natural language processing fields [4, 5]. The BERT model uses a large amount of text to generate a pretrained model, and the BERT model consists of two pretraining methods: (1) the masked language model (MLM) and (2) next sentence prediction (NSP). The MLM is a training method

C. Min · H. Na · Y. Song · D.-H. Im (✉)
Kwangwoon University, Seoul, South Korea
e-mail: dhim@kw.ac.kr

J. Ahn
Jeju National University, Jeju, South Korea

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_98

731

that masks a certain number of tokens and predicts the masked tokens, and NSP is a training method that takes two sentences as input and predicts the order of the sentences. The BERT model generates a pretrained model that has acquired general language knowledge based on these two pretraining methods. The pretrained model generated by the BERT model is used in various tasks through transfer learning. However, even if general language knowledge is acquired through pretraining, the pretrained model does not perform well in tasks that require specialized knowledge. To overcome this drawback, pretraining should be performed with data containing specialized knowledge about the corresponding field. However, this method requires considerable time and a large amount of specialized knowledge. Hence, applying this method in practice is difficult. Therefore, research has been conducted on using external data such as a knowledge graph to supplement insufficient knowledge about the input data. A knowledge graph is structured knowledge. It expresses the predicate representing the relationship between entities of the subject and object in a triple structure <subject, predicate, object> format. The K-BERT model in [6] employed a method that adds knowledge graph information, which is external data, to compensate for the BERT model's drawback of having poor performance in specialized tasks. This method improved the performance of the BERT model even in specialized tasks that were not general natural language tasks. However, because the knowledge graph has a lot of information, the K-BERT model uses information outside the topic of the input data, which may confuse the training of the model. To compensate for this drawback of the K-BERT model, [7] used the LDA technique, a topic modeling technique, in the knowledge graph. The LDA technique is a statistical technique for determining a topic in a document. This technique was used in the TK-BERT model to divide a vast knowledge graph into topics and infer the topic of the input data. Based on this, only knowledge matching the topic of the input data is added. Consequently, the TK-BERT model achieved a better performance than that of the K-BERT model. However, the LDA technique takes a document-term matrix (DTM) and term frequency-inverse document frequency (TF-IDF) as input. Hence, the LDA technique does not consider the order of the words. That is, the LDA technique does not consider the contextual information. Because the structure of a knowledge graph is constructed in the <subject, predicate, object> triple structure, the order of each word is important. To compensate for this drawback, this study proposes a method that uses a knowledge graph after dividing the knowledge graph into topics considering the contextual information of the knowledge graph using the BERTopic technique.

2 Knowledge Graphs

Knowledge graphs are data structurally constructed to represent the concept of knowledge. Each node of a knowledge graph consists of entities corresponding to the subjects and objects. In addition, the edge connecting each node is expressed as a predicate representing the relationship between subjects and objects. The constructed

graph can be expressed as a triple structure in the form of <subject, predicate, object>. The triple structure can express the relationship between entities clearly and concisely. This simple structure is suitable for adding insufficient knowledge about the input data in natural language processing. In the K-BERT model, which was investigated in a previous study, the lack of knowledge about input data in various fields of natural language processing was overcome using knowledge graphs. However, because the K-BERT model refers to a vast knowledge graph, it refers to knowledge outside the topic of the input data, as well as knowledge about the input data. This is called the knowledge noise problem. The knowledge noise problem is the phenomenon of confusing the training of models, and it occurs because too much knowledge is added from a knowledge graph. To prevent the occurrence of knowledge noise, Min et al. [7] used a knowledge graph by dividing it into topics using the LDA technique, which is a topic modeling technique. In this study, the BERTopic model is used to solve the LDA technique's problem, which is not considering the order of words. The BERTopic model shows a more effective use of a knowledge graph in natural language processing because it partitions the knowledge graph more appropriately according to the topics.

3 Topic Modeling

Topic modeling is a statistical model for estimating abstract topics that are inherent in a set of documents, and the LDA technique is a common topic modeling technique. The LDA technique begins by assuming that a document is composed of a mixture of topics, and the topics generate words based on the probability distribution. Based on this assumption, the LDA technique backtracks the process through which the document was created and infers the topic of the document and words. However, the LDA technique takes a DTM and TF-IDF as input. Hence, it uses the frequency of the words in the document but does not consider the order of the words. If the order of the words is not considered, the contextual information of each document cannot be known, making the identification of the exact topic difficult. To compensate for this problem, research was recently conducted on the BERTopic technique, which is a topic modeling technique that considers the contextual information [8]. The BERTopic technique uses BERT-based embedding and class-based TF-IDF. Figure 1 shows the structure of the BERTopic technique, and it largely consists of three stages. The first stage is the embedding stage, and the BERT model is used to perform embedding for each document. Here, the BERT model used in the embedding process is a pretrained model. The second stage uses UMAP to reduce the dimension of each document vector and performs clustering using HDBSCAN. Here, similar documents are clustered for each document vector. The third stage determines words that well represent each group using c-TF-IDF and makes adjustments using the maximize candidate relevance algorithm such that words representing each group are selected as diversely as possible. The BERTopic model assigns topics considering the contextual information through this process. To assign topics considering the context of the triple

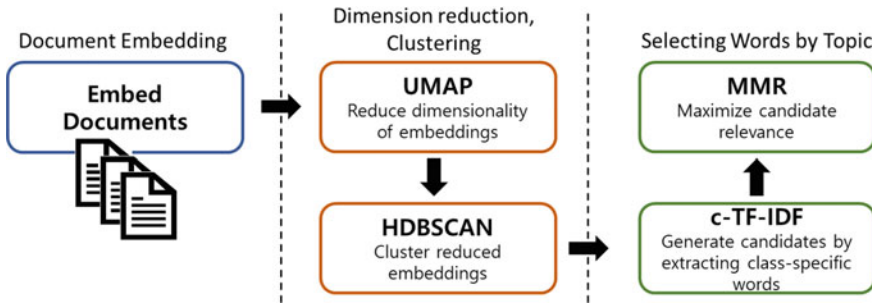


Fig. 1 Structure of the BERTopic model

structure of the knowledge graph, the BERTopic model was used to assign the topics of the knowledge graph, and the knowledge graph was divided and used based on the topics.

4 Method

Figure 2 shows the structure of the overall method of this study. Figure 2 consists of three stages: (1) generating the topic model and partitioning the knowledge graph, (2) inferring the topic of the input sentence, and (3) adding knowledge that matches the topic. In the first stage, a topic model is generated using the knowledge graph and BERTopic technique, and the generated topic model is used to partition the knowledge graph according to the topics. Here, the knowledge that matches the topic of the input data can be determined using the partitioned knowledge graph. In the second stage, the topic of the input data is inferred using the topic model generated in the first stage. Finally, in the third stage, the knowledge of the knowledge graph matching the inferred topic is added. The knowledge added in this stage refers only to the knowledge graph that matches the topic of the input data inferred in the second stage. That is, the BERTopic technique can be used to alleviate the knowledge noise phenomenon, which is a problem of the K-BERT model, and solve the existing LDA technique's drawback of not considering the context.

5 Experiment

In this study, an $F1$ -score comparison experiment was conducted to compare the performances of the K-BERT model, which provides insufficient knowledge through the knowledge graph, the TK-BERT model, which uses the LDA technique to partition the knowledge graph by topic, and our model, which uses the BERTopic technique. Two knowledge graphs, Cn-DBpedia and HowNet, were used to train the

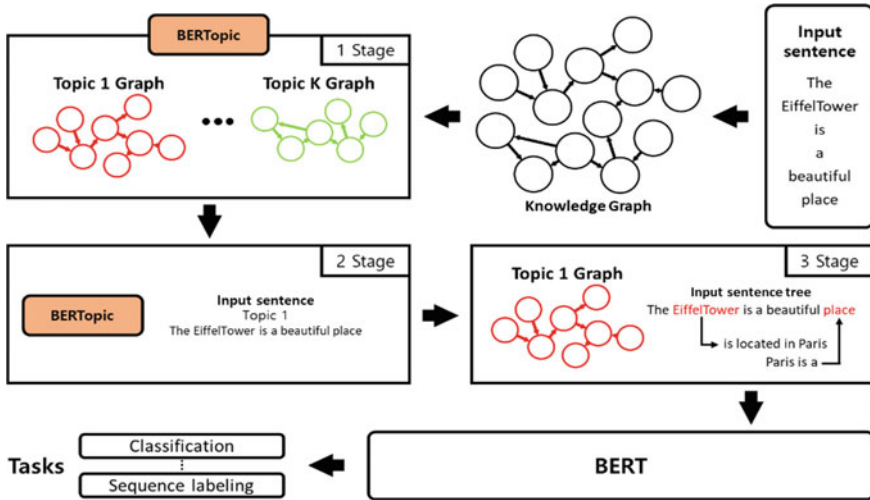


Fig. 2 Structure of the method

K-BERT model. The TK-BERT model was trained using the two knowledge graphs after partitioning them using the LDA technique. Finally, the knowledge graphs were partitioned by topic using the BERTopic technique, and the partitioned knowledge graphs were used to train our model.

5.1 Experiment Environment

Google BERT was used as the pretrained model of BERT in the experiment. This model was pretrained with WikiZh data, which is Chinese Wikipedia data composed of 12 million sentences. In addition, two knowledge graphs, Cn-DBpedia and HowNet, were used in the experiment. The Cn-DBpedia knowledge graph consists of approximately 5.16 million triple data points. The HowNet knowledge graph concerns the Chinese lexicon and consists of approximately 52,000 triple data points. The TopicModel was constructed using the two knowledge graphs and BERTopic technique. The pretrained model of the BERTopic model used models trained with more than 50 languages. Because the knowledge graphs were partitioned by topic using the topic model, the two knowledge graphs were combined and then partitioned by topic. Here, the number of topics was set to 50 because 50 achieved the best performance in an experiment conducted using 50, 100, and 150 as the numbers of topics. In addition, the Book_review, Chnsenticorp, and Shopping datasets were used. The datasets consist of positive and negative reviews. The Book_review dataset consists of 20,000 positive b and 20,000 negative book reviews. The Chnsenticorp dataset is hotel review data, and it consists of 6000 positive and 6000 negative hotel

Table 1 Results of the K-BERT model, TK-BERT model, and our model

Model	Dataset					
	Book_review		Chnsenticorp		Shopping	
	Dev	Test	Dev	Test	Dev	Test
K-BERT (HowNet)	86.8	85.85	94.5	94.6	96.5	96.05
K-BERT (Cn-DBpedia)	87.4	86.25	93.95	94.5	96.45	96.3
TK-BERT (Topic-based HowNet)	87.35	86.45	94.5	95.6	96.75	96.45
TK-BERT (Topic-based Cn-DBpedia)	88.2	86.9	94.5	95.5	96.5	96.35
Our model	88.55	87.1	94.5	95.65	96.75	96.65

Bold represents high performance

reviews. Finally, the Shopping dataset contains online shopping review data, and it consists of approximately 21,000 positive and 19,000 negative shopping reviews.

5.2 Results

Table 1 compares the $F1$ -scores of the K-BERT model, TK-BERT model, and our model; our model performed the best on each dataset. The K-BERT model adds knowledge using the knowledge graph, and the TK-BERT model uses the knowledge graph by partitioning it using the LDA technique. OurModel uses the knowledge graph by partitioning it using the BERTopic model to compensate for the LDA technique's drawback of not considering the contextual information. The results show that when the K-BERT model, which adds knowledge using the knowledge graph, is used; information that benefits learning can be added if the knowledge graph is used after partitioning it by topic. In addition, when the BERTopic technique is used to partition the knowledge graph, the knowledge graph can be divided into topics more effectively than when using the LDA technique to partition the knowledge graph because the topic model is generated considering the contextual information.

6 Conclusion

In this study, we proposed a method for a more effectively implementation of the TK-BERT model, addressing the problem of the existing K-BERT model. The topic model of the TK-BERT model uses the LDA technique. The LDA technique estimates a topic using only the frequency of the words in a document. Hence, it cannot reflect the contextual information. To address this problem, the BERTopic technique was used to implement the topic model because this technique can consider the contextual information. Because the BERTopic model considers the contextual information through document embedding, it can use the knowledge graph by partitioning it

more effectively. The experiment verifies that our model outperformed the TK-BERT model, which uses the existing LDA model. The results indicate that our model can effectively partition even a larger knowledge graph for training.

Acknowledgements This work was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2022-2018-08-01417) supervised by the Institute for Information and Communications Technology Promotion (IITP). This research was also funded by industry-academic Cooperation R&D program funded by LX Spatial Information Research Institute (LXSIRI, Republic of Korea).

References

1. Maulud DH et al (2021) State of art for semantic analysis of natural language processing. Qubahan Acad J 1(2):21–28
2. Zhang Z et al (2020) Semantics-aware BERT for language understanding. Proc AAAI Conf Artif Intell 34(05)
3. Xu W et al (2021) Enabling language representation with knowledge graph and structured semantic information. In: 2021 international conference on computer communication and artificial intelligence (CCAI). IEEE
4. Radford A et al (2018) Improving language understanding by generative pre-training
5. Devlin J et al (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
6. Liu W et al (2020) K-BERT: enabling language representation with knowledge graph. Proc AAAI Conf Artif Intell 34(03)
7. Min C, Ahn J et al (2023) TK-BERT: effective model of language representation using topic-based knowledge graphs. In: International conference on ubiquitous information management and communication (IMCOM) (accepted)
8. Grootendorst M (2022) BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint [arXiv:2203.05794](https://arxiv.org/abs/2203.05794)

The Performance Analysis of Lightweight Neural Networks for Salient Object Detection



Thien-Thu Ngo, Eui-Nam Huh, and Choong-Seon Hong

Abstract Nowadays, salient object detection (SOD) has become a prominent research area in computer vision, which has various applications in real life. With the advancement of convolutional neural networks (CNN), numerous SOD methods have been proposed to imitate the human visual system to identify salient objects in images. Their marvelous performance comes with large models, high computational complexity, and memory consumption, raising many challenges to deploy these models on practical applications and low-resource devices. Under the high demand for developing efficient models for SOD tasks, in this paper we provide a comprehensive performance analysis focusing on existing CNN-based lightweight SOD models in terms of accuracy and efficiency. By discussing recent methods regarding the proposed ideas and main techniques, we analyze the current issues of lightweight SOD models in detail. Various metrics are considered for performance assessment followed by future research directions and conclusion.

Keywords Salient object detection · Lightweight salient object detection

1 Introduction

Salient object detection (SOD) aims to recognize the most noticeable objects in the scene, which has been applied in various surveillance applications such as visual tracking, scene understanding, image segmentation, image editing [1–4], etc. Conventional methods calculated saliency maps based on handcraft low-level

T.-T. Ngo · E.-N. Huh (✉) · C.-S. Hong
Kyung Hee University, Global Campus, Yongin 17104, South Korea
e-mail: johnhuh@khu.ac.kr

T.-T. Ngo
e-mail: thu.ngo@khu.ac.kr

C.-S. Hong
e-mail: cshong@khu.ac.kr

features. Due to lack of high-level semantic information, these methods are less accurate when detecting objects in complex backgrounds. Recently, the advancement of convolutional neural networks (CNN) with the help of large-scale datasets for salient object detection also has huge attractions from the research community and contributes to the outstanding performance of recent state-of-the-art methods [5]. Based on the pre-trained models on the ImageNet dataset for classification tasks, the authors discussed new challenges on salient object detection and proposed new models for solving these problems. In this way, the semantics features learned by CNNs have been re-used in SOD tasks effectively. However, the amelioration of these models triggers extra costs simultaneously, such as heavy networks with the large number of parameters and large model size, heavy computational overhead, slow inference speed, and large memory consumptions. These ponderous aspects make it difficult to deploy on strictly constrained environments and low-resource devices such as mobile and embedded devices. In other words, the demand for developing small-scale models for salient object detection which tradeoff between the cost and performance is highly requisitioned. However, there are very less frameworks specially designed for salient object detection. Existing surveys lack technical details and less focus on shallow frameworks. In this paper, we provide an extensive analysis of promising techniques for lightweight salient object detection, by discussing various existing CNN-based methods based on the motivations and contributions of authors. Based on these reviews, we provide various comparisons between the representative lightweight models in terms of accuracy and efficiency.

2 The Analysis of Existing Lightweight Techniques

2.1 *The Baseline Network*

Building standard lightweight CNNs to solve the above problems draws much attention from researchers recently. Heavyweight models based on FCN [6] using encoder-decoder framework have achieved great performance over the years. In order to reduce the complexity, considering shallow networks as a backbone network is a promising solutions. However, existing lightweight networks such as MobileNet [7], ShuffleNet [8] have fewer layers so naively applying these networks as an encoder can degenerate the network to produce multi-scale features. Moreover, these networks do not take into account the inter-correlations between channels and spatial positions, which may yield poor performance. In Table 1, we summarize the main ideas of existing approaches for lightweight salient object detections.

Table 1 Summary of most representative lightweight salient object detection models

Methods	Year	Backbone	Strategy
AAmulet [9]	2018	VGG16	A simple efficient aggregating module for SOD
PoolNet [10]	2019	MobileNetV2	Efficient pooling-based modules to improve SOD
CPD [11]	2019	VGG16	Partial decoders for reducing the complexity, holistic attention to enlarge coverage area of saliency map
LANet [12]	2020	MobilenetV2	Lightweight bottleneck block, adversarial training, multi-scale contrast
DNM [13]	2020	IRblock	Depthwise non-local module to mining intra-channel and inter-channel correlations
100k [14]	2020	ILblock	Generalized OctConv to explore multi-scale representation and cross-scale features interactions, weight pruning
SAMNet [15]	2021	SAM module	Stereoscopic attention mechanism for efficient multi-scale learning
HVPL [16]	2021	HVP module	Mimicking primate visual hierarchy to improve multiscale learning
CTD [17]	2021	ResNet18	Complementary Trilateral Decoder (CTD) with three branches: Semantic Path, Spatial Path, Boundary Path
PPAM [18]	2022	MobileNetV3 small	Design a position prior attention module, efficient feature fusion module and structural polishing loss

2.2 Enhancing Lightweight Neural Networks Block

Inverted residual (IR) bottleneck is the basic block in MobileNetv2 [19]. IR blocks are based on depthwise separable convolutions, in which each spatial position is expanded to higher dimensional before applying convolutions on channel splitting independently. This convolution significantly reduces computational cost, while inverted residual accelerates training and inference speed. Several works adopted IR blocks into their models in an efficient way. LANet [12] adopted these blocks to build their lightweight adversarial network. DNM [13] efficiently improved IR blocks through channel-wise coherence and contrast learning. SAMNet and HPVL [15, 16] embed DS convolutions in between their proposed module to reduce parameters and computational load. Besides, generalized OctConv [14] is a promising method to reduce the complexity of basic blocks while maintaining multi-scale representations.

2.3 Attention Mechanism

Recently, attention has shown their great potential roles in many SOD models for feature recalibrations. As salient objects are usually disturbed by relative regions, discriminating salient objects from the background helps reduce wrong detection.

Attention mechanism guide the network to focus more on salient objects rather than other areas by eliminating redundant information in both channel and spatial dimensions. Particularly, different channels yield different semantic information, to highlight salient objects and reduce feature redundancy, channel attention assigned larger weights to channel which has more contributions to salient objects. Similarly, spatial attention is applied to filter out non-relevant information from the background. AAmulet [9] constructed a contextual attention pyramid to exploit rich context at all scales. CPD [11] proposed holistic attention module aimed at expanding the coverage area of the initial saliency map. SAMNet [15] proposed the stereoscopic attention mechanism for multi-scale feature learning which has strong intra-channel and spatial-wise dependency. HVPL [16] employed channel spatial attention and residual attention sequentially and iteratively to explore inter-channel dependencies. PPAM [18] optimized attention mechanism by adding position prior based on Euclidean distance between pixels to the image center.

2.4 Feature Aggregation Module

Multi-scale features in both coarse and fine level yield more fruitful representations for SOD tasks. Local information extracted from low-level features preserves object boundaries and enhances details information. The extracted high-level information contains rich semantic information but lacking spatial details. As a results, an efficient feature fusion strategy in SOD provides more sophisticated representation in multi-scale manner. Most of the existing works enrich the fusion process using complicated modules. Therefore, developing lightweight and efficient fusion modules is one of solutions for improving lightweight SOD models. AAmulet [9] proposed a straightforward yet incredibly effective aggregation strategy that converts each level's features into a dimension-reduced tensor with the same spatial resolution as the input images. 100k [14] used gOctConvs to fuse cross-stages multi-scale features from the last three stages of the features extractor. SAMNet [15] used a simple combination of DS convolutions, elementwise summation, and 1×1 convolution to aggregate features at different scales. PPAM [18] introduced a novel context fusion module that combined low-level and high-level features through residual connection.

3 Performance Comparisons

We select 12 existing lightweight models for the performance comparison in terms of accuracy and efficiency as in Fig. 1 and Table 2. We perform quantitative comparison on four state-of-the-art datasets: HKU-IS, ECSSD, DUTS-OMRON, and PASCAL-S. We adopted two metrics, maximum F -measure and MAE, to evaluate the accuracy performance, and conduct the efficiency comparison under number of parameters, number of FLOPS and speed (FPS).

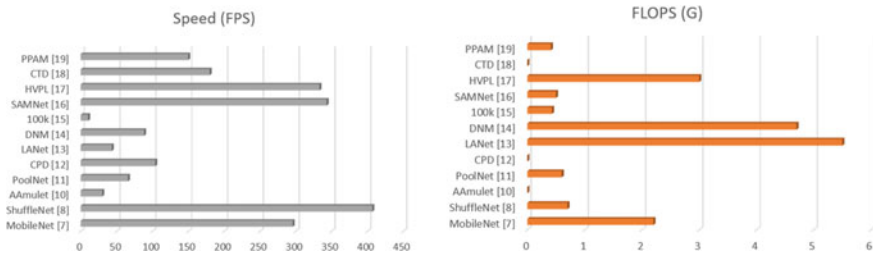


Fig. 1 Efficiency measurements based on speed (FPS) and number of FLOPS (G)

Table 2 Performance comparison based on number of parameters, F-max ($F\uparrow$), MAE ($M\downarrow$)

Methods	#Param (M)	ECSSD		PASCAL-S		DUT-OM		HKU-IS	
		$F\uparrow$	$M\downarrow$	$F\uparrow$	$M\downarrow$	$F\uparrow$	$M\downarrow$	$F\uparrow$	$M\downarrow$
MobileNet [7]	4.27	0.890	0.073	0.771	0.131	0.744	0.087	0.897	0.053
ShuffleNet [8]	1.80	0.876	0.083	0.785	0.121	0.730	0.094	0.891	0.055
AAmulet [9]	–	0.887	0.049	0.794	0.092	0.691	0.076	0.861	0.038
PoolNet [10]	3.0	0.938	0.048	0.864	0.078	0.830	0.056	0.930	0.037
CPD [11]	47.85	0.936	0.042	0.859	0.071	0.796	0.056	0.925	0.034
LANet [12]	2.09	0.917	0.049	0.858	0.076	0.814	0.061	0.922	0.031
DNM [13]	5.3	0.909	0.064	0.822	0.106	0.779	0.077	0.913	0.045
100k [14]	0.94	0.916	0.066	0.835	0.102	0.792	0.080	0.899	0.059
SAMNet [15]	1.33	0.925	0.053	–	–	0.797	0.065	0.915	0.045
HVPL [16]	1.23	0.925	0.055	–	–	0.799	0.064	0.915	0.045
CTD [17]	11.82	0.920	0.037	0.831	0.065	0.767	0.052	0.916	0.028
PPAM [18]	1.04	0.923	0.041	0.862	0.068	0.798	0.053	0.920	0.030

4 Future Research Directions and Conclusion

In this paper, we provide the empirical performance analysis for lightweight salient object detections. The review of existing models has been conducted to further understand key concepts and main ideas of recent proposals. As we have known, this is the first paper to provide reviews for lightweight salient object detections for constrained resource devices. For fair comparison, we perform various experiments results in terms of accuracy (maximum F -measure, MAE) and efficiency [number of parameters, number of FLOPS, and speed (FPS)]. Based on this study, we suggest research directions for future works as in below:

- **Knowledge distillation-based models:** To reduce memory consumption and accelerate model speed, one of the useful techniques is knowledge distillation [KD]. KD [20] provides the transformation to transfer knowledge from the large model (teacher) to the small model (student) without loss of validity. The main challenge in KD-based models is the complexity of transformation module for handling various sizes of feature maps. Moreover, very less paper carefully designed KD techniques for SOD tasks.
- **Binary neural network:** another technique for producing compact models is quantization, which is widely used in binary neural network [21]. Instead of using full precision, this method transforms weight and activation layer to 1 bit layer. The Pioneer model can speed up to $32\times$ memory saving and $58\times$ speed up on CPUs. This technique has been applied in image classifications but has not been fully studied for salient object detections.
- **Others:** the common loss criterion for SOD is binary cross-entropy loss (BCEL). However, BCEL ignores the relationship between pixels, so carefully designed loss function can improve the performance effectively.

References

1. Ma C, Miao Z, Zhang X-P, Li M (2017) A saliency prior context model for real-time object tracking. *IEEE Trans Multimed* 19(11):2415–2424
2. Ren Z, Gao S, Chia L-T, Tsang IW-H (2013) Region-based saliency detection and its application in object recognition. *IEEE Trans Circuits Syst Video Technol* 24(5):769–779
3. Li Q, Zhou Y, Yang J (2011) Saliency based image segmentation. In: 2011 international conference on multimedia technology. IEEE, pp 5068–5071
4. Ardizzone E, Bruno A, Mazzola G (2013) Saliency based image cropping. In: International conference on image analysis and processing. Springer, Berlin, Heidelberg, pp 773–782
5. Wang W, Lai Q, Fu H, Shen J, Ling H, Yang R (2021) Salient object detection in the deep learning era: an in-depth survey. *IEEE Trans Pattern Anal Mach Intell* 44(6):3239–3259
6. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
7. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861)
8. Zhang X, Zhou X, Lin M, Sun J (2018) Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6848–6856
9. Zhang P, Wang L, Wang D, Lu H, Shen C (2018) Agile amulet: real-time salient object detection with contextual attention. arXiv preprint [arXiv:1802.06960](https://arxiv.org/abs/1802.06960)
10. Liu J-J, Hou Q, Liu Z-A, Cheng M-M (2022) Poolnet+: exploring the potential of pooling for salient object detection. *IEEE Trans Pattern Anal Mach Intell*
11. Wu Z, Su L, Huang Q (2019) Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3907–3916
12. Huang L, Li G, Li Y, Lin L (2020) Lightweight adversarial network for salient object detection. *Neurocomputing* 381:130–140

13. Li H, Li G, Yang B, Chen G, Lin L, Yu Y (2020) Depthwise nonlocal module for fast salient object detection using a single thread. *IEEE Trans Cybern*
14. Gao S-H, Tan Y-Q, Cheng M-M, Lu C, Chen Y, Yan S(2020) Highly efficient salient object detection with 100k parameters. In: *European conference on computer vision*. Springer, Cham, pp 702–721
15. Liu Y, Zhang X-Y, Bian J-W, Zhang L, Cheng M-M (2021) SAMNet: stereoscopically attentive multi-scale network for lightweight salient object detection. *IEEE Trans Image Process* 30:3804–3814
16. Liu Y, Gu Y-C, Zhang X-Y, Wang W, Cheng M-M (2020) Lightweight salient object detection via hierarchical visual perception learning. *IEEE Trans Cybern* 51(9):4439–4449
17. Zhao Z, Xia C, Xie C, Li J (2021) Complementary trilateral decoder for fast and accurate salient object detection. In: *Proceedings of the 29th ACM international conference on multimedia*, pp 4967–4975
18. Zhang J, Liang Q, Shi Y (2022) Accurate and efficient salient object detection via position prior attention. *Image Vis Comput* 124:104508
19. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4510–4520
20. Gou J, Yu B, Maybank SJ, Tao D (2021) Knowledge distillation: a survey. *Int J Comput Vis* 129(6):1789–1819
21. Qin H, Gong R, Liu X, Bai X, Song J, Sebe N (2020) Binary neural networks: a survey. *Pattern Recognit* 105:107281

Pendulum: An I/O Efficient Algorithm for Nearest Surrounding Queries in 3D Space



Kyeongseok Hyun  and Jaehwa Chung 

Abstract A Nearest Surrounding (NS) query retrieves a set of all visible spatial objects near a given query point. For many years, a NS query proved useful in handling 2D objects and extended to 3D geospatial environment for location-based services. As recent pandemic has prompted greater interest in virtual reality, a NS querying technique can support many applications processing 3D spatial objects. However, a 3DNS query still adopts conventional ray shooting algorithm, exploring 3D objects in angular order on xy-plane. This search method leads to frequent updates and discarding intermediate NS results caused by a visibility of newly observed objects, incurring more I/O access for object comparisons. Nevertheless, this problem can be prevented by applying best-first search technique with a metric that well captures the 3D geospatial environment properties. Motivated by this weakness, we propose a pendulum algorithm by applying distance from a query point to an object and objects' height through inclination angle as a priority.

Keywords NS query · 3D spatial objects · Search algorithm

1 Introduction

The impact of Covid-19 pandemic affected many aspects of lives, especially including outdoor activities. Consequently, virtual reality (VR) technologies gained immense interests as a substitute for not only outdoor activities but also for various social activities. To support a variety of applications on VR environments, we re-examine spatial query processing of three-dimensional spatial objects with respect to a user location.

K. Hyun
Korea University, Seoul, Republic of Korea
e-mail: ks_hyun@korea.ac.kr

J. Chung (✉)
Korea National Open University, Seoul, Republic of Korea
e-mail: jaehwachung@knou.ac.kr

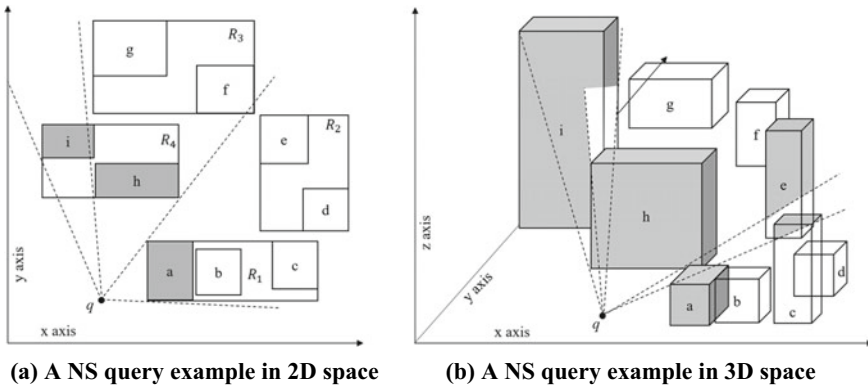


Fig. 1 NS query examples

The notion of Nearest Surrounding (NS) was proposed by Lee et al. [1] initially to retrieve a set of all visible spatial objects around the user location in two-dimensional space. In Fig. 1a, given a query point q , the NS query returns only visible objects a, h , and i from an object set. Objects b, c, d , and e are discarded by invisible region formed by a , while f, g and part of i are discarded by invisible region formed by h . However, in three-dimensional space, objects' heights should also be considered when processing an NS query. We extend Fig. 1a by adding height information which is shown in Fig. 1b. In this figure as seen, the angle formed by a 's height and q does not completely veil c and d , which make them visible. To process NS query in 3D space, Chung et al. [2] extended conventional NS query by defining angular measurements.

For efficient traversal of spatial objects, Lee et al. [3] proposed ripple search algorithm via visiting the closest node first that encloses 2D objects. On the other hand, 3D NS query processing still adopts conventional ray shooting algorithm [4] which visits node in angular order, incurring unnecessary node traversal through R^* -tree index [5]. Motivated by this weakness, we propose a novel algorithm that effectively searches every visible object in 3D spaces by using inclination angle (ϕ) ordering.

Our contributions are summarized as follows

- Address the issue of heuristic for NS query processing 3D spatial objects.
- Propose I/O efficient algorithm considering 3D space properties by applying distance from a query point to an object and objects' height through inclination angle.

2 Related Work

In the literature, Lee et al. [1] proposed the concept of NS query. In later work, Lee et al. [3] proposed sweep algorithm based on the conventional ray shooting query [4] to avoid sequential comparisons for the visibility among the 2D objects. The sweep algorithm uses angular-based distance metric (madist) which returns a distance from a query to an object. When traversing a R^* -tree index, the lower-bound distance (min_madist) of visible objects is used to safely prune nodes that are unnecessary to explore. Although the sweep algorithm decreases the latency of processing NS query, it still visits unnecessary nodes while exploring in angular order from 0 to 2π . To resolve this, Lee et al. [3] also proposed the ripple algorithm that got maximum efficiency out of madist metric applied heuristics. In 2D environment, objects near a query point have heavy influence on the further objects due to the invisible regions they can create. Therefore, instead of traversing the R^* -tree in angular order, the ripple algorithm visits nodes in mindist order (ascending) which can significantly increase the efficiency of traversing a R-tree. Extension to NS query, Zhang et al. [6] proposed Group Visible (GV) NS query in 2D space. The focus of GVNS query is to find the NS of a query set Q , where obstacles are mixed with objects. Nevertheless, GVNS query algorithm also traverses objects incrementally ordered according to their starting angles.

To handle 3D spatial objects, Chung et al. [2] proposed a heuristic for efficient processing of 3DNS query. By using inclination angle, it computes a secluded region created by a query point and an object that could veil other objects in its region. Yet, it still adopts conventional ray shooting algorithm when traversing a R^* -tree. Although angle-based ordering approach can still prune nodes, it does not prioritize the objects that creates invisible region crucial from a query point. Conversely, our pendulum algorithm tries to capture important properties (mindist and inclination angle) of 3D environment and prioritize search order to increase efficiency in exploring 3D spatial objects.

3 Query Processing in 3D Geospatial Environment

In this section, we propose heuristic utilizing inclination angle of 3D spatial objects from a query point and efficient search algorithm on a R^* -tree.

3.1 Angular Measurements of 3D Spatial Objects

The angular measurements of 2D spatial objects from a query point are denoted with viewable angular range $[\theta^{\rightarrow}, \theta^{\leftarrow}]$. The first component of an angular range θ^{\rightarrow} is minimum polar angles measured from the baseline (positive x -axis, where y equals to a query point's y -coordinate) like Fig. 2a. The simplicity of 2D space makes visibility of objects determined only by distance from a query point and angular range.

However, when it comes to 3D space, visibility of an object cannot be determined only by $[\theta^{\rightarrow}, \theta^{\leftarrow}]$ and a distance from a query point. Thus, Chung et al. [2] expanded the viewable angular range representation with inclination angle measured from the xy -plane like Fig. 2b. For an arbitrary point $p = (x_p, y_p, z_p)$ on an object, and a query point $q = (x_q, y_q)$, we can calculate the inclination angle ϕ with following:

$$\phi = \arctan\left(\frac{z_p}{\sqrt{(x_q - x_p)^2 - (y_q - y_p)^2}}\right) = \arctan\left(\frac{z_p}{\text{madist}(q, p, \theta)}\right), \quad (1)$$

$\text{madist}(q, p, \theta)$ refers to the shortest distance from a q to p at θ which is a polar angle formed between q and p on xy -plane. Yet, when processing geo-spatial data, madist function usually takes point q , an object O_i , and an angular range $[\theta_i^{\rightarrow}, \theta_i^{\leftarrow}]$. Accordingly, when we want to calculate the ϕ of an object O_i , $\text{madist}(q, O_i, [\theta_i^{\rightarrow}, \theta_i^{\leftarrow}])$ should be calculated as a denominator of Eq. (1). ϕ is not static and changes according to $[\theta_i^{\rightarrow}, \theta_i^{\leftarrow}]$. For inclination angular range representation, Chung et al. [2] also proposed ϕ_{\min} and ϕ_{\max} shown in Fig. 2b. The definition of inclination angular range (ϕ_{\min}, ϕ_{\max}) can be defined as follows:

Definition 1 (ϕ_{\min}, ϕ_{\max}) Given an object O_i and a query q , ϕ_{\min} and ϕ_{\max} refer the minimum and maximum inclination angles of O_i from q , and defined by the following equations:

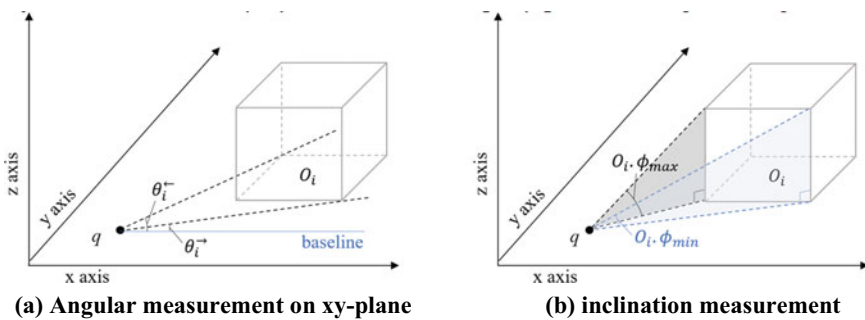


Fig. 2 Angular measurement of a 3D object

$$\phi_{\min} = \arctan(O_i^{\text{height}} / \max_madist(q, O_i, [\theta_i^{\rightarrow}, \theta_i^{\leftarrow}])) \quad (2)$$

$$\phi_{\max} = \arctan(O_i^{\text{height}} / \min_madist(q, O_i, [\theta_i^{\rightarrow}, \theta_i^{\leftarrow}])) \quad (3)$$

With angular measurement θ and ϕ , we can now define 3D nearest surrounders query as follows:

Definition 2 (*3D nearest surrounders query*) Given dataset D and a query point q , 3D nearest surrounders query of q , $3DNS(q)$, returns a set of tuple $O_i : \langle \theta, \phi \rangle \in N$ satisfies that: (i) $N \subset D$, (ii) $O_i \in N$ and O_i is visible to q , and (iii) $\forall O_j \in D - N$, $\text{dist} \leq \text{dist}(q, O_j)$ at angular range $\langle \theta, \phi \rangle$.

3.2 Pendulum Algorithm for 3DNS Query Processing

Searching NS through R*-tree index requires pruning heuristics for an efficient traversal. As Lee et al. [3] used conservative upper bound and lower bound to skip exploration of objects in 2D space by using madist metric, Chung et al. [2] also proposed basic pruning heuristics for 3D space as follows:

Heuristic 1 Suppose that current nearest surrounders, $N = \{O_1, O_2, \dots, O_n\}$. Given an object ϵ , for invisibility condition, if $\exists O_i \in N$ such that (i) $[\theta_\epsilon^{\rightarrow}, \theta_\epsilon^{\leftarrow}] \subset [\theta_i^{\rightarrow}, \theta_i^{\leftarrow}]$, (ii) $\text{madist}(O_i, q) < \text{madist}(\epsilon, q)$, and (iii) $\epsilon \cdot \phi_{\max} < O_i \cdot \phi_{\min}$, then ϵ can be safely excluded from the query results.

Regarding the Heuristic 1, if O_i 's angular range covers ϵ 's angular range, and O_i 's height is tall enough to veil ϵ , we can safely remove ϵ from the result or prune it for further exploration. Although heuristic 1 reduces overall search space, 3DNS query processing algorithm proposed in [2] still adopts conventional ray shooting algorithm which occurs unnecessary index lookup incurring more I/O cost. To improve the search performance, we propose one more heuristic to apply best-first search technique using ϕ metric.

Heuristic 2 Using inclination angle $\phi \cdot \min$ as search priority can guarantee of preventing situation where newly observed object ϵ satisfies following condition:

$$(\epsilon \cdot \min \text{madist} < O_i \cdot \max \text{madist}) \quad \text{and} \quad (O_i \cdot \phi_{\max} < \epsilon \cdot \phi_{\min}),$$

which means NS object O_i could have been pruned or discarded earlier if we have observed ϵ first.

By using $\phi \cdot \min$ as best-first search metric, the pendulum algorithm searches through objects back and forth from a query point, but ensures unnecessary objects

being placed in a NS list saving I/O costs. Algorithm 1 shows the overall pendulum algorithm.

Algorithm 1 Pendulum algorithm

Input:	a query point q , root of R-tree
Output:	NS result N
1.	Initialize priority Q with root; //priority: ϕ_{min} descending order
2.	Initialize NS result N with $\{\emptyset\}$
3.	while (Q is not empty)
4.	$\epsilon \leftarrow \text{dequeue}(Q)$
5.	if (ϵ is node) then
	Let $[\theta_c^+, \theta_c^-]$ be the common angular range of O_i and ϵ ;
	Initialize visible angle VA with $\{[\theta_c^+, \theta_c^-]\}$;
6.	foreach O_i in NS
7.	if $([\theta_c^+, \theta_c^-] \subseteq [\theta_{O_i}^+, \theta_{O_i}^-])$ and $\epsilon, \phi_{max} < O_i, \phi_{max}$ then $VA \leftarrow VA - \{[\theta_c^+, \theta_c^-]\}$;
8.	else if $([\theta_c^+, \theta_c^-] \subseteq [\theta_{O_i}^+, \theta_{O_i}^-])$ and $\epsilon, \phi_{max} > O_i, \phi_{max}$ then break ;
9.	if ($VA = \emptyset$) then prune ϵ ;
10.	if (ϵ is not pruned) then
	explore ϵ and put all its child to Q;
11.	else
12.	$N \leftarrow N \cup \text{3DObjectCompare}(q, NS, \epsilon)$;
13.	Output N ;

Algorithm 2 3D object compare algorithm

Input:	a query point q , NS result N , object ϵ
Output:	NS
1.	foreach tuple in N
2.	Let $[\theta_c^+, \theta_c^-]$ be the common angular range of O_i and ϵ ;
3.	if (O_i and ϵ has $[\theta_c^+, \theta_c^-]$) then
4.	if ($O_i, \text{maxmadist} < \epsilon, \text{minmadist}$ and $O_i, \phi_{max} < \epsilon, \phi_{min}$) then
	$R \leftarrow R \cup \langle \epsilon: [\theta_c^+, \theta_c^-], [O_i, \phi_{min}, \epsilon, \phi_{max}] \rangle$;
5.	else if ($O_i, \text{maxmadist} < \epsilon, \text{minmadist}$ and $\epsilon, \phi_{max} < O_i, \phi_{min}$) then skip ;
6.	else
	update O_i 's angular range to $\langle O_i: [\theta_c^+, \theta_c^-], [\epsilon, \phi_{min}, O_i, \phi_{max}] \rangle$;
7.	if $([\theta_\epsilon^+, \theta_\epsilon^-] - [\theta_c^+, \theta_c^-] \neq \emptyset)$ then
	$N \leftarrow N \cup \{ \langle \epsilon: [\theta_\epsilon^+, \theta_\epsilon^-] - [\theta_c^+, \theta_c^-], [0, \phi_{max}] \rangle \}$;
8.	if $([\theta_\epsilon^+, \theta_\epsilon^-] - [\theta_c^+, \theta_c^-] \neq \emptyset)$ then
	$N \leftarrow N \cup \{ \langle O_i: [\theta_{O_i}^+, \theta_{O_i}^-] - [\theta_c^+, \theta_c^-], [0, \phi_{max}] \rangle \}$;

For comparison, trace of ray shooting algorithm and pendulum algorithm using Fig. 1's examples is shown in Fig. 3. The ray shooting algorithm on Fig. 3a incurs unnecessary node visit of R_3 and its child causing more I/O costs. However, pendulum algorithm visits R_4 node first due to its priority of having the highest ϕ_{min} . Consequently, R_3 node is pruned in later phase of the algorithm. Recalling Eq. 1, shorter the distance from the query point and higher the height of the building gets the higher priority. Therefore, we can conclude that inclination angles well depict important properties of 3D spatial objects.

ϵ	Q	NS
	$\{R_1, R_2, R_3, R_4\}$	\emptyset
R_1	$[a, b, c, R_2, R_3, R_4]$	\emptyset
a	$[b, c, R_2, R_3, R_4]$	$\{a: \{(\theta_a^-, \theta_a^+), [0, a, \phi_{max}]\}\}$
b	$[c, R_2, R_3, R_4]$	Ditto (b removed by a, ϕ_{max})
c	$[R_2, R_3, R_4]$	$\{a: \{(\theta_c^-, \theta_c^+), [0, \phi_a, max]\}, c: \{(\theta_c^-, \theta_c^+), [a, \phi_{min}, c, \phi_{max}]\}\}$
R_2	$[d, e, R_3, R_4]$	Ditto
d	$[e, R_3, R_4]$	Ditto (d removed by a, ϕ_{min})
e	$[R_3, R_4]$	$\{a: \{(\theta_e^-, \theta_e^+), [0, \phi_a, max]\}, c: \{(\theta_e^-, \theta_e^+), [a, \phi_{min}, c, \phi_{max}]\}, e: \{(\theta_e^-, \theta_e^+), [a, \phi_{min}, e, \phi_{max}]\}\}$
R_3	$[f, R_4, g]$	Ditto
f	$[R_4, g]$	$\{a: \{(\theta_f^-, \theta_f^+), [0, \phi_a, max]\}, c: \{(\theta_f^-, \theta_f^+), [a, \phi_{min}, c, \phi_{max}]\}, e: \{(\theta_f^-, \theta_f^+), [a, \phi_{min}, e, \phi_{max}]\}, f: \{(\theta_f^-, \theta_f^+), [0, f, \phi_{max}]\}\}$
R_4	$[h, i, g]$	Ditto
h	$[i, g]$	$\{a: \{(\theta_h^-, \theta_h^+), [0, \phi_a, max]\}, c: \{(\theta_h^-, \theta_h^+), [a, \phi_{min}, c, \phi_{max}]\}, e: \{(\theta_h^-, \theta_h^+), [a, \phi_{min}, e, \phi_{max}]\}, f: \{(\theta_h^-, \theta_h^+), [0, f, \phi_{max}]\}, h: \{(\theta_h^-, \theta_h^+), [0, h, \phi_{max}]\}\}$
i	$[g]$	$\{a: \{(\theta_i^-, \theta_i^+), [0, \phi_a, max]\}, c: \{(\theta_i^-, \theta_i^+), [a, \phi_{min}, c, \phi_{max}]\}, e: \{(\theta_i^-, \theta_i^+), [a, \phi_{min}, e, \phi_{max}]\}, h: \{(\theta_i^-, \theta_i^+), [0, h, \phi_{max}]\}, i: \{(\theta_i^-, \theta_i^+), [h, \phi_{min}, i, \phi_{max}]\}, j: \{(\theta_i^-, \theta_i^+), [0, i, \phi_{min}]\}\}$
g	$[\]$	Ditto (g removed by h, ϕ_{min} and i, ϕ_{min})

ϵ	Q	NS
	$[R_4, R_1, R_2, R_3]$	\emptyset
R_4	$[i, R_1, h, R_2, R_3]$	\emptyset
i	$[R_1, h, R_2, R_3]$	$\{i: \{(\theta_i^-, \theta_i^+), [0, h, \phi_{max}]\}\}$
R_1	$[h, c, R_2, R_3, a, b]$	Ditto
h	$[c, R_2, R_3, a, b]$	$\{i: \{(\theta_h^-, \theta_h^+), [h, \phi_{min}, i, \phi_{max}]\}, j: \{(\theta_h^-, \theta_h^+), [0, i, \phi_{max}]\}, h: \{(\theta_h^-, \theta_h^+), [0, h, \phi_{max}]\}\}$
c	$[R_2, R_3, a, b]$	$\{i: \{(\theta_c^-, \theta_c^+), [h, \phi_{min}, i, \phi_{max}]\}, j: \{(\theta_c^-, \theta_c^+), [0, i, \phi_{max}]\}, h: \{(\theta_c^-, \theta_c^+), [0, h, \phi_{max}]\}, c: \{(\theta_c^-, \theta_c^+), [0, c, \phi_{max}]\}\}$
R_2	$[e, R_3, a, b, d]$	Ditto
e	$[R_3, a, b, d]$	$\{i: \{(\theta_e^-, \theta_e^+), [h, \phi_{min}, i, \phi_{max}]\}, j: \{(\theta_e^-, \theta_e^+), [0, i, \phi_{max}]\}, h: \{(\theta_e^-, \theta_e^+), [0, h, \phi_{max}]\}, c: \{(\theta_e^-, \theta_e^+), [0, c, \phi_{max}]\}, e: \{(\theta_e^-, \theta_e^+), [0, e, \phi_{max}]\}\}$
R_3	$[a, b, d]$	Ditto (R_3 pruned by a, ϕ_{min} , and h, ϕ_{min})
a	$[b, d]$	$\{i: \{(\theta_a^-, \theta_a^+), [h, \phi_{min}, i, \phi_{max}]\}, j: \{(\theta_a^-, \theta_a^+), [0, i, \phi_{max}]\}, h: \{(\theta_a^-, \theta_a^+), [0, h, \phi_{max}]\}, c: \{(\theta_a^-, \theta_a^+), [a, \phi_{min}, c, \phi_{max}]\}, e: \{(\theta_a^-, \theta_a^+), [a, \phi_{min}, e, \phi_{max}]\}, a: \{(\theta_a^-, \theta_a^+), [0, a, \phi_{max}]\}\}$
b	$[d]$	Ditto (b removed by a, ϕ_{min})
d	$[\]$	Ditto (d removed by a, ϕ_{min})

(a) ray shooting algorithm tracing

(b) pendulum algorithm tracing

Fig. 3 Trace of algorithms

4 Conclusion

This paper presents an I/O efficient 3DNS search algorithm using inclination angle of objects from a query point. On future work, we explore algorithms for searching 3D space with less constraints where queries and objects are not fixed to the xy-plane.

Acknowledgements This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2022R1F1A1074727).

References

1. Lee K et al (2007) Round-Eye: a system for tracking nearest surroundings in moving object environments. J Syst Softw 80(12):2063–2076
2. Chung J et al (2015) Efficient processing of nearest surround query for 3D geospatial data. Int J Softw Eng Appl 9(11):207–216
3. Lee K et al (2009) Nearest surround queries. IEEE Trans Knowl Data Eng 22(10):1444–1458
4. Salomon D et al (1999) Computer graphics and geometric modeling. Springer

5. Beckmann N et al (1990) The R*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD international conference on management of data, pp 322–331
6. Zhang H et al (2019) Group visible nearest surround query in obstacle space. In: International conference on computer science and educational informatization. IEEE, pp 345–350

Proposal of Telehealth Service Reference Models as a Prerequisite for Cybersecurity Risk Analysis



In Hye Lee and Jung-ha Jin

Abstract Telehealth, a non-face-to-face service over a network, is increasing the risks of cybersecurity with increasing demands. To properly respond to these risks, it is necessary to understand the types of telehealth and analyze the risks associated with each type. It is important to define a telehealth service models that effectively represent cybersecurity characteristics that differ from a medical perspective. Therefore, this paper identifies and defines telehealth reference models useful for deriving cybersecurity requirements for telehealth services. The scope of telehealth encompasses telemedicine and telecare, and the reference models have been simplified by dividing the primary telehealth activities into three categories: encounter, observation, and intervention. And, this paper proposes that telehealth service reference models should be built up base on the flows of data and services among actors. In the future, it is necessary to apply the telehealth reference models to the real-world telehealth entities including the healthcare delivery organizations to verify the models and discover the improvements. In addition, based on the improved models, it is expected to extend the study to derive cybersecurity risks and appropriate countermeasures.

Keywords Telehealth service model · Telehealth service cybersecurity · Telemedicine model

1 Introduction

Changes in the healthcare environment and ICT development are rapidly increasing the number of telehealth cases. However, telehealth, a non-face-to-face service through networks, is increasing the risk of cybersecurity as well as the increased demand. In order to properly respond to these risks, it is necessary to understand the types of telehealth and analyze the risks associated with each type.

I. H. Lee · J. Jin (✉)

Institute of Cyber Security and Privacy, Korea University, Seoul 02841, South Korea
e-mail: nemoda75@korea.ac.kr

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
J. S. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 1028,
https://doi.org/10.1007/978-981-99-1252-0_101

755

The telehealth types from a cybersecurity perspective have different characteristics from those in a medical perspective. Therefore, in order to identify telehealth cybersecurity threats and vulnerabilities, it will be a prerequisite to define telehealth types that are easy to analyze cybersecurity risks. Accordingly, this paper proposes telehealth reference models useful for analyzing telehealth cybersecurity characteristics.

Following the introduction, this paper focuses on analysis for telehealth cybersecurity research in Sect. 2, proposes telehealth reference models for addressing cybersecurity issues in Sect. 3, and presents considerations for further study in Sect. 4.

2 Related Research

2.1 Telehealth Terminology

Telehealth, a compound word of the prefixes ‘tele-’ which means telephone, and ‘health,’ is originated from medical services for remote patients using telephones. Currently, various types of services such as real-time video conference, analysis of test results, and image reading are provided as well as phone calls. JASEHN [1] defines ‘eHealth’ as the most comprehensive concept that utilizes electronic communication and information technology in the medical field. It places ‘telehealth’ as a subset of eHealth and illustrates telehealth including ‘telecare,’ which is a non-medical preventive activity, and ‘telemedicine’, which is a medical activity. Also, teletriage, teleconsultation, telediagnosis, telesurgery and telescreening, and telemonitoring are explained as types of medical service, and radiology, pathology, dermatology, psychiatry, cardiology, and ophthalmology are mentioned in medical specialism perspective. Since this paper aims to derive a telehealth service models that are enable to apply cybersecurity requirements, telehealth, telemedicine, telecare, and teleconsultation are mainly applied by referring to the terms defined in the JASHN [1].

eHealth. It is the combined use of electronic communication and information technology in the health sector to share, store, and retrieve electronic health data for prevention, diagnosis, treatment, monitoring, educational and administrative purposes, both at the local site and at a distance [1].

Telehealth. It is a subset of eHealth and refers to the delivery of healthcare at a distance. It encompasses the delivery of healthcare services by all healthcare professionals, where distance is a critical factor, using ICT to provide clinical and non-clinical services—preventive, promotive and curative healthcare services, research and evaluation, health administration services and continuing education of healthcare providers [1].

Telemedicine. It is a subset of telehealth and refers to the delivery of healthcare services at a distance only by physicians or healthcare professionals who perform the predefined medical activities by physicians. Although JASHN [1] narrowed

the provider of telemedicine to physicians, this paper includes professionals who have licenses approved for medical activities by jurisdiction regulators. Therefore, telemedicine can include activities by nurses, therapists, or pathologists.

Telecare. Telecare involves the use of ICT, such as alerts and sensing technologies, for the remote monitoring of care needs, emergencies, and lifestyle changes of elderly or vulnerable individuals with physical or mental disabilities in order to provide personalized care services at a distance, supporting patients' self-management and helping them to remain independent in their home environment [1–5].

2.2 Telehealth Service Models

ATA Guidelines. American Telemedicine Association (ATA) guidelines are developed in each clinical specialization and do not present a comprehensive high-level model. Nevertheless, the concepts of 'virtual visits', 'chat-based interactions', 'remote patient monitoring', and 'technology-enabled modalities' defined in ATA [6] have implications applicable to the telehealth service models of this paper. In addition, ATA [7, 8] provides the 'Typical telehealth information network' topology and items of telemedicine cybersecurity capabilities.

ISO 13131 [9]. This standard, which is quality planning guidelines for telehealth, represents the management model for telehealth service provision from the perspective of PDCA (Plan-Do-Check-Act). This approach recommends risk-based thinking of *ISO 31000 Risk management—Guidelines* for effective telehealth service and highlights safety consideration for telehealth quality management.

IEEE [10]. This document, which deals with the cybersecurity concerns involved in the IEEE 11073 series standards, presents the intended actors, exchanged data, mapping to assets, and threat model, focusing on a use case of Physical Activity Monitor (PAM). In particular, the PAM threat model is a good reference for developing a telehealth service models because it represents detailed interactions among the device components, actors, and data flow. Although IEEE 11073 series standards illustrate different characteristics of each personal health devices (PHDs), they help to understand a comprehensive topology, data flow, and cybersecurity threats.

Yeoh [11]. Figure 2 'Benefits and limitations of implementing chronic care models (CCMs) in primary care' represents specifically designed and conducted program for the fulfillment of multiple CCM components. This model shows the CCM service flow based on health system or a health organization (HSO), professional health providers, clinical information system (CIS), patients, and other support systems.

3 Reference Models of Telehealth Services

3.1 Healthcare Service Activities

High level of healthcare services can be represented as follows:

- Encounter: participants meet and initiate clinical activities. The activities include patient enrolment, appointment, patient reception, and entering the consulting room;
- Observation: a healthcare service provider sees the patient health status using medical resources such as EMR, questionnaire, clinical check-up, or test results;
- Intervention: a healthcare service provider makes clinical decisions or performs clinical procedures to improve patient health status such as prescription, treatment, or surgery (Fig. 1).

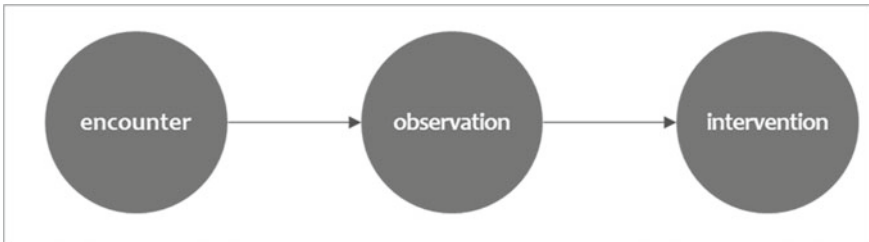


Fig. 1 High level activities of healthcare services

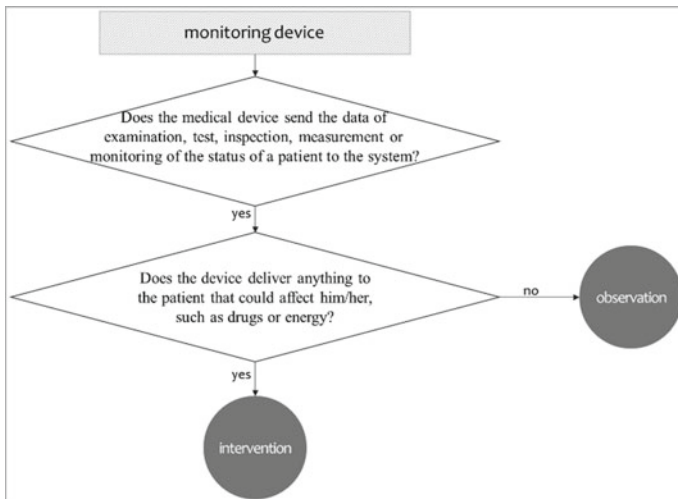


Fig. 2 Telehealth activity using the monitoring device

When a patient visits doctor after enrolment, appointment, and reception (encounter), a physician examines the patient's health status by taking a history, palpation, percussion, or auscultation (observation). If more information is needed, the physician orders additional test such as EKG, X-ray, or MRI (observation). The physician develops a treatment plan based on the observation results and provides clinical treatments, therapy, or prescription (intervention).

In telehealth services, a physician who is a telehealth service provider interacts with various remote counterparts involved in clinical practices. The remote counterparts may be a patient or user directly receiving telehealth services provided by a telehealth provider, or it may be a caregiver or supporter assisting a patient or user who is unable to see himself/herself. Otherwise, it may be a physician who needs a medical opinion or cooperations to properly treat the patient.

For telehealth encounter, each party goes through the steps of subscribing to, scheduling, and accessing a telehealth system based on identification and authorization. Observation activities are performed based on real-time virtual examinations or data collected from the medical devices installed at the remote patient's or user's site. After reviewing the observations, the intervention is made according to the decision of the physician.

In the case of the observation using a monitoring device, if the device simply sends the data collected from the patient to the system, it is a kind of observation. The physician will review the data collected and make the appropriate clinical decision, such as changing medication, dosage, or energy delivered to the patient. In this case, observation and intervention should be performed using separated routes. If the monitoring device could send the data collected from the patient to the system as well as give medication or energy to the patient through the same device, the telehealth activity should be considered as intervention rather than observation (Fig. 2).

For example, if an implantable sphygmomanometer only collects the heart rate data and alert the physician to take emergency action at abnormal values, this telehealth activity is defined as an observation. However, if the device is preset to automatically control the heart rate with a built-in drug on a specific trigger, the telehealth activity is an intervention not an observation.

3.2 Reference Model of Telemedicine Services

Telemedicine is the related to the clinical activities provided by the physicians including observations and interventions. Even in sessions where medical care is provided experts other than physicians, the activities for patients are governed by treatment plan developed by the physicians.

- (observation) Data is collected from the patient's devices or systems and sent to the physician. If the patient cannot properly send or communicate data to the physician, caregivers or supporters can do it for the patient. ① indicates the

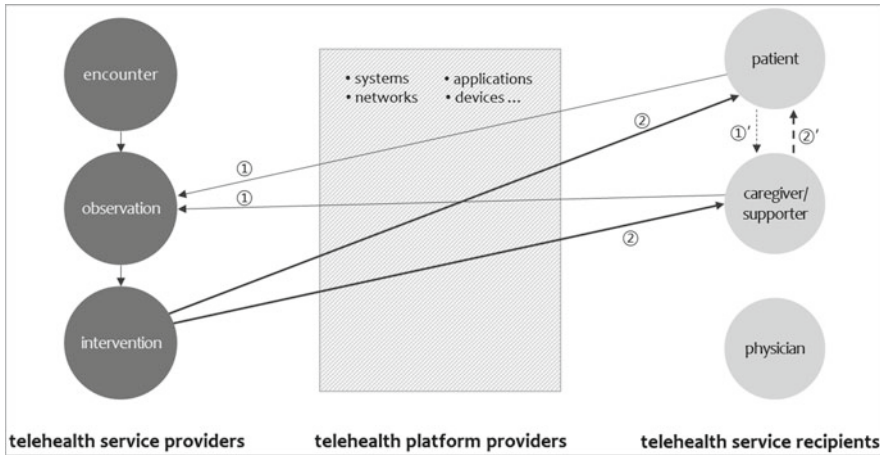


Fig. 3 Reference model of telemedicine

direction of observation data. ①' means data collection activities between the patient and caregiver or supporter in the same physical area.

- (intervention) Based on results of the observations, the physician develops the treatment plan and delivers prescriptions, advice, or treatment to the patient or indirectly to the caregivers or supporters. ② indicates the direction of intervention data. ②' means healthcare activities performing by caregiver or supporter according the physician instruction (Fig. 3).

3.3 Reference Model of Teleconsultation Services

Teleconsultation is defined as “physician–physician remote consultation for improving patient’s condition using ICT”. This service is for connecting physicians at two or more separate areas to discuss and cooperate for the better outcomes such as treatment plans, surgical techniques, or emerging technologies, etc. based on the shared information of the subject of care. Although ISO/TR 14639-1:2012 [12] defines teleconsultation as ‘provision of healthcare via a telehealth service, generally for the purpose of diagnosis or treatment of a subject of care at a site remote from the patient or their primary care provider,’ this document narrows the scope of teleconsultation to the ‘physician-to-physician activities’ for distinguishing it from other telehealth services.

- (observation) The physician is shared data of the patient’s conditions by the remote physician who is charge of the patient. The physician only discusses and shares the opinions with the remote physician for the remote subject of care (Fig. 4).

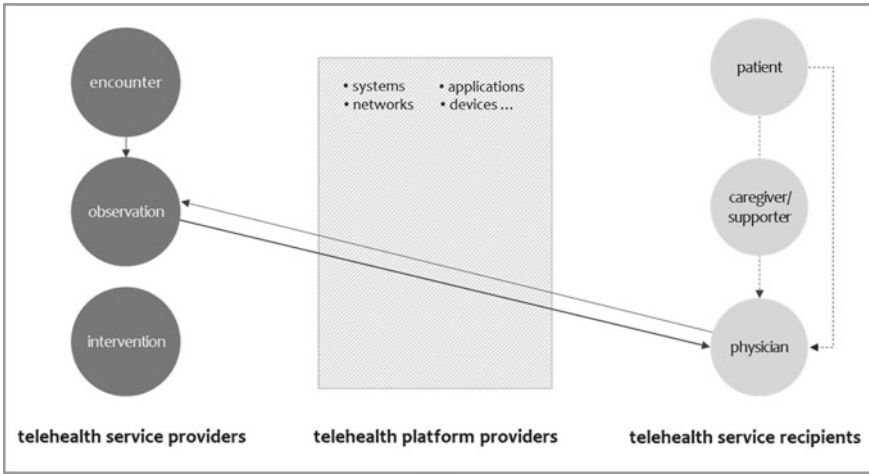


Fig. 4 Reference model of teleconsultation

3.4 Reference Model of Telecare Services

Telecare is placed outside of medical territories but contributes to improve the quality of life as a preventive mechanism by combining with emerging technologies.

- (observation) Telecare is not related to intervention. Telecare service providers analyze data of service users and provide predefined services according to contract. They monitor sensor data, activate emergency alarms, or provide corrective advices (Fig. 5).

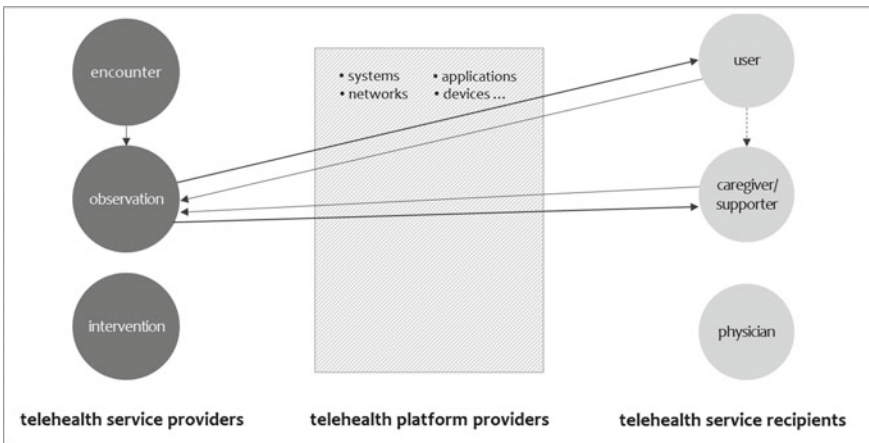


Fig. 5 Reference model of telecare services

4 Conclusion

As a prerequisite for analyzing the risk of telehealth service from a cybersecurity perspective, a comprehensive healthcare model that classifies healthcare service into encounter, observation, and intervention is proposed. Telehealth service reference models based on primary medical activities and data flow were derived from the healthcare service model.

- Telemedicine: Observation and intervention activities take place. Telemonitoring which is only collecting patient data is observation, but telemonitoring that can automatically affect a patient according to a preset by a doctor is an intervention.
- Teleconsultation is an observation because a physician checks the patient's condition and provides medical advice for the remote physician without performing intervention to the patient.
- Telecare is an observation as a non-medical activity.

Although service reference models are defined as a prerequisite for telehealth service risk analysis, risk assessment and classification of each telehealth service are required to derive cybersecurity requirements appropriate for various types of telehealth service. In particular, the cybersecurity of telemedicine services, where information systems and medical devices interact based on the different physical environments, will be a challenge.

In the future, it is necessary to apply the telehealth reference models to the real-world telehealth entities including the healthcare delivery organizations to verify the models and discover the improvements. In addition, based on the improved models, it is expected to extend the study to derive cybersecurity risks and appropriate countermeasures.

Acknowledgements This work was supported by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (1711138615, KMDF_PR_20200901_0272).

References

1. JASEHN (2018) Report on EU state of play of telemedicine services and uptake recommendations v.0.5, D7.1.2. JASEHN
2. Oh H, Rizo C, Enkin M, Jadad A (2005) What is eHealth (3): a systematic review of published definitions. *J Med Internet Res* 7(1):e1. <https://doi.org/10.2196/jmir.7.1.e1>
3. Mitchell J (1999) From telehealth to e-health: the unstoppable rise of e-health. Department of Communications, Information Technology and the Arts
4. Telehealth and telecare aware. What is telecare? Retrieved from: <http://telecareaware.com/what-is-telecare/>. Accessed 29 Mar 2017
5. McLean S et al (2013) The impact of telehealthcare on the quality and safety of care: a systematic overview. *PLoS ONE* 8(8):e71238. <https://doi.org/10.1371/journal.pone.0071238>

6. ATA (2020) Telehealth: defining 21st century care. <https://www.americantelemed.org/resource/why-telemedicine/>
7. ATA (2014) Clinical guidelines for telepathology. American Telemedicine Association
8. Healthcare & Public Health Sector Coordinating Councils (2021) Health industry cybersecurity—securing telehealth and telemedicine
9. ISO 13131 Health informatics—Telehealth services—Quality planning guidelines (2021)
10. IEEE PHD cybersecurity standards roadmap v.1.0 (2019), pp 91–94. ISBN: 978-1-5044-5773-6. STDVA23662
11. Yeoh EK, Wong MCS, Wong ELY, Yam C, Poon CM, Chung RY, Chong M, Fang Y, Wang HHX, Liang M, Cheung WWL, Chan CH, Zee B, Coats AJS (2018) Benefits and limitations of implementing Chronic Care Model (CCM) in primary care programs: a systematic review. *Int J Cardiol* 258:279–288. <https://doi.org/10.1016/j.ijcard.2017.11.057>
12. ISO/TR 14639-1 Health informatics—Capacity-based eHealth architecture roadmap—Part 1: Overview of national eHealth initiatives (2012)

A Challenge of the Representative Spatiotemporal Document Classification and Spatiotemporal Information Extraction



Yeongwook Yang, Ji Su Park, Hong-Jun Jang, Byoungwook Kim, and Jin Gon Shon

Abstract Classification of representative spatiotemporal documents from text data and extraction of representative spatiotemporal information are necessary studies for searching for specific events and accidents, and real-time time and place recognition for events. However, this task is challenging due to the absence of data and difficulties in recognizing representative spatiotemporal information. We are collecting data to solve this problem, and in the process of collecting data, we analyzed the challenges of classifying representative spatiotemporal documents and extracting representative spatiotemporal information through qualitative and quantitative evaluation.

Keywords Representative spatiotemporal information extraction · Representative spatiotemporal document classification

Y. Yang

Division of Computer Engineering, Hanshin University, Osan, South Korea
e-mail: yeongwook.yang@hs.ac.kr

J. S. Park · H.-J. Jang

Department of Computer Science and Engineering, Jeonju University, Jeonju-si, South Korea
e-mail: jisupark@jj.ac.kr

H.-J. Jang

e-mail: hongjunjang@jj.ac.kr

B. Kim

Department of Computer Science and Engineering, Dongshin University, Naju, South Korea
e-mail: bwkim@dsu.ac.kr

J. G. Shon (✉)

Department of Computer Science, Graduate School, Korea National Open University, Seoul, South Korea
e-mail: jgshon@knou.ac.kr

1 Introduction

The development of information technologies has made it possible to produce various and vast amounts of data (Big Data) and has made it possible to conduct research to find high-quality information by processing Big Data. Among the various types of data, information related to space and time provides a specific place and location, and since it provides the time period in which a specific event or accident occurred, various studies using it are being conducted [1].

Specific documents or data may or may not include spatiotemporal information, and if included, there may be multiple spatiotemporal information. Existing researches have extracted and used all spatiotemporal information in documents [2–4]. However, the importance of spatiotemporal information differs according to the purpose of providing information in the document. We define spatiotemporal information (STI) that matches the purpose of providing information in a document or has high importance as representative spatiotemporal information (RSTI). Also, a document containing representative spatiotemporal information is defined as a representative spatiotemporal document (RSTD) [5].

However, it is a difficult problem to distinguish the purpose of information provision and the importance of STI in a document. The reason is that a human should actually check all the contents of the document, and there is a difference in recognizing the purpose of providing information and classifying the importance of STI according to human cognitive and language abilities. Also, as far as we know, there is no dataset for the classification of RSTD. Therefore, it is very challenging to distinguish whether a document contains RSTI or not, and it is more challenging to find RSTI in RSTD.

We are conducting RSTD classification and RSTI extraction using the news data from the National Institute of Korean Language¹ to deal with the aforementioned problem. We analyze and share the challenges of classifying RSTD and extracting RSTI through the verification of the quality of the collected data and the survey contents of the stakeholders who participated in that work.

2 Related Work

Spatiotemporal data mining research includes various subresearches such as clustering, prediction, pattern mining, and visualization [1]. The most data type used for spatiotemporal data mining uses location values including latitude and longitude [1]. However, it is rare that latitude and longitude are included in the text data most used in real life.

In the case of research using text data in spatiotemporal data mining, most of the target data is text data generated from social networks [6]. In this case, spatial information is expressed using latitude and longitude as metadata.

¹ <https://www.korean.go.kr/>.

In the study on news data, the text expressing spatial information was replaced with latitude and longitude through geocoding. A study was conducted to visualize news trends through substituted latitude and longitude values [7, 8]. In this study, spatiotemporal information that can represent documents is not considered.

3 Method

To solve this research problem, we extracted and used a part of the corpus provided by the National Institute of Korean Language and analyzed RSTD classification, RSTI extraction, and verification results.

Among the data of the National Institute of the Korean Language, 1400 news data produced in 2009 were used, and using an identifier it can identify each document, the title of the document, the creation date of the document, and the document content which were extracted. In the step of classifying RSTD and extracting RSTI, labelers were made to judge whether or not they were RSTD after perusing the contents of the documents according to the guidelines prepared in advance. If it is a RSTD, the labeler extracts spatial and temporal information representing the document. One document was assigned to two or more labelers, and cross-check was performed to ensure the accuracy of labeling. After that, the expert confirmed the labeling result. After labeling was performed on 1400 news data sets, four validators analyzed whether the dataset was correctly classified and labeled. Finally, through a questionnaire to labelers, opinions were collected on “How difficult was the task?”, “Why was it difficult?”, and “How much confidence do you have in your work?”.

According to the research procedure, we performed quantitative and qualitative analysis to derive the challenges of RSTD classification and RSTI extraction from text data.

4 Analysis Results

4.1 Quantitative Data Analysis Results

In the total dataset (1400), RSTD accounts for about 26% (359), and non-RSTD accounts for about 74% (1041), indicating that there are more non-RSTD than RSTD. In the task of verifying the accuracy of RSTD classification and RSTI extraction, the ratio of true positives in RSTD and the ratio of false negatives in all non-RSTDs were analyzed according to the number of participants (Table 1).

According to the number of validator who correctly classified the RSTD, when all four people classified correctly, when 3 out of 4 classified correctly, when 2 out of 4 classified correctly, and when 1 out of 4 classified correctly, each was analyzed. The rate that all four validators correctly classified the RSTD was the lowest at 52.7%,

Table 1 Percentage of true positive and false negative by the number of labeler

% of true positive				% of false negative			
1 validator	2 validators	3 validators	4 validators	1 validator	2 validators	3 validators	4 validators
92.5	86.4	69.9	52.7	13	6.3	2.3	0.48

and the rate that 1 out of 4 correctly classified the RSTD was the highest at 92.5%. 3 out of 4 cases and 2 out of 4 cases showed true positive values of 69.9% and 86.4%, respectively.

4.2 Qualitative Data Analysis Results

We asked the following questions to six labelers who participated in the classification of RSTD and extraction of RSTI.

- (1) How difficult is the classification of RSTD and extraction of RSTI?
- (2) How accurately did you extract RSTI?
- (3) Which task is more difficult: representative temporal information extraction or representative spatial information extraction?
- (4) Why is it difficult or easy to classify RSTD and extract RSTI?

To the first question, four labelers answered ‘normal’, 1 labeler selected ‘easy’, and the rest answered ‘difficult’. In other words, the labeler felt that the task was not difficult. As for the result of the second question, five labelers answered that they extracted correctly, and one labeler that it was ‘normal’. In other words, they answered that their work results were accurate. As for the third question result, three labelers answered that it was difficult to extract representative spatial information, one labeler answered representative spatial information was difficult and the rest selected the same and not know, respectively. The result of reason for the difficulty in the fourth question is as follows. One labeler answered, “There may not be representative spatiotemporal information in text data, or there may be several. In this case, it was difficult to understand the context of the entire news.” Another answer was, “In the case of administrative documents, it was difficult because a specific administrative organ could be spatial information.” The reason the task was easy was that the information retrieval range was small because the date format was fixed in the data. They also said that it is easy to extract information even if the document deals with sports or events.

The results of qualitative data analysis are considered by labelers as easy as this task is, but it is difficult for humans to accurately judge this task, as can be seen from the actual results of quantitative data analysis.

5 Conclusion

As a result of qualitative data analysis, it was confirmed that labelers tend to think that it is easy to classify RSTD and extract RSTI. However, the results of quantitative data analysis numerically show that the task is difficult.

The reason why the work is difficult is that the document may not be an RSTD, and it is difficult to determine even an RSTD if there is several spatiotemporal information. Also, it was analyzed that the content of the document could be easy or difficult depending on what category it was for. In the category related to administration, it was difficult to classify the RSTD, and it can be derived that the categories related to sports or events are easy to classify the RSTD.

This study is expected to be able to quickly search time for important events and accidents, and it is expected to provide a framework for representative spatiotemporal research from text data. In our future research, we intend to classify and summarize documents using RSTI.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1C1C2004868).

References

1. Hamdi A, Shaban K, Erradi A, Mohamed A, Rumi SK, Salim FD (2022) Spatiotemporal data mining: a survey on challenges and open problems. *Artif Intell Rev* 55(2):1441–1488
2. Badia A, Ravishankar J, Muezzinoglu T (2007) Text extraction of spatial and temporal information. In: *Proceedings of the 2007 IEEE intelligence and security informatics*, New Brunswick, NJ, USA, 23–24 May
3. Lim C-G, Jeong Y-S, Choi H-J (2019) Survey of Temporal Information Extraction. *J Inf Process Syst* 15:931–956
4. Ferial A, Kholadi M-K (2015) Automatic extraction of spatio-temporal information from Arabic text documents. *Int J Comput Sci Inf Technol* 7:97–107
5. Kim B, Yang Y, Park JS, Jang HJ (2022) A convolution neural network-based representative spatio-temporal documents classification for big text data. *Appl Sci* 12(8):3843
6. Zhang D, Lee K, Lee I (2015) Periodic pattern mining for spatio-temporal trajectories a survey. In: *Intelligent systems and knowledge engineering (ISKE)*, 2015 10th international conference on. IEEE, pp 306–313
7. Kastner JH, Samet H (2020) Visualizing spatiotemporal keyword trends in online news articles. In: *Proceedings of the 28th international conference on advances in geographic information systems*, pp 195–198
8. Bo A, Peng S, Xinming T, Alimu N (2011) Spatio-temporal visualization system of news events based on GIS. In: *Communication software and networks (ICCSN)*, 2011 IEEE 3rd international conference on, pp 448–451

IoT Sensor Predictive Analysis System Using Apache Spark in Edge Cloud Environment



Gi-dae Hong, Young-Min An, Seol Roh, and Eui-Nam Huh

Abstract The Internet of Things (IoT) environment, which enables networking and computing in everything, is rapidly spreading. IoT environments cause bottlenecks and service delays to process data and provide services to users through a cloud-based central processing structure. In order to solve this problem, edge computing, which provides services to users by directly frosting data from IoT nodes and access networks, is rapidly spreading to cloud environment, where research continues to efficiently provide low delay intelligent services to users. In this paper, the response time of the service required to complete LSTM training is mainly measured by comparing to the locality of centralized data in the core cloud. In addition, the accuracy of the existing LSTM and the improved algorithm is compared on various reliability conditions.

Keywords Edge-computing · Apache Spark · LSTM

1 Introduction

With the recent Fourth Industrial Revolution, advanced information and communication technologies such as artificial intelligence, the Internet of Things, big data, and mobile have been converging into the economy and society, leading innovative

G. Hong · Y.-M. An · S. Roh · E.-N. Huh (✉)

Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea

e-mail: johnhuh@khu.ac.kr

G. Hong

e-mail: kidae92@khu.ac.kr

Y.-M. An

e-mail: aem9704@khu.ac.kr

S. Roh

e-mail: seven800@khu.ac.kr

changes. In particular, a vast amount of data is being collected in real time from numerous IoT devices as IoT technology is widely distributed [1]. And as limited resources of devices require external resources, IoT cloud collaborative services based on the convergence of IoT and cloud computing are emerging as key models for implementing ICT-based creative and diverse future-oriented application services [2].

However, existing centralized cloud server-based services must go through a central network (Network Backbone), resulting in latency, and many requests are concentrated on the central network, which can cause a large load. To solve this problem, edge computing, which provides services to users by directly frosting data from IoT nodes and networks, has recently attracted attention, and research has continued to efficiently provide intelligent services to users in this edge computing environment [3].

IoT devices must operate the entire system through data distributed processing and calculation, not centrally controlled through servers. Due to the nature of the edge computing system structure, real-time information detected simultaneously on each device can be efficiently processed in the entire system [4]. To solve the problem of low processing efficiency, we propose data analysis model learning techniques via Apache Spark among edge clusters. The results of performance measurement studies on based on temperature/humidity sensors are shown in comparisons with existing LSTMs by treating reliability on historical data differently. We build HDFS on the edge cluster and learn the LSTM algorithm model in parallel on the Spark platform. We also incorporate periodic data processing and model re-learning scheduling through Apache Airflow [5].

2 Related Work

2.1 *Apache Spark*

Apache Spark is a scalar-based open-source computing platform developed to enable parallel processing of large datasets. Apache Spark supports data analytics at high performance and speed for a variety of formats (static or streaming) because it performs operations in memory [6]. Spark's execution architecture consists of driver and executor. The driver is responsible for the DAG segmentation of the user code, divides it into different steps, and then submits it to the executor to calculate the task schedule for each step to run parallel tasks of the same step. The calculation process is shown in Fig. 1.

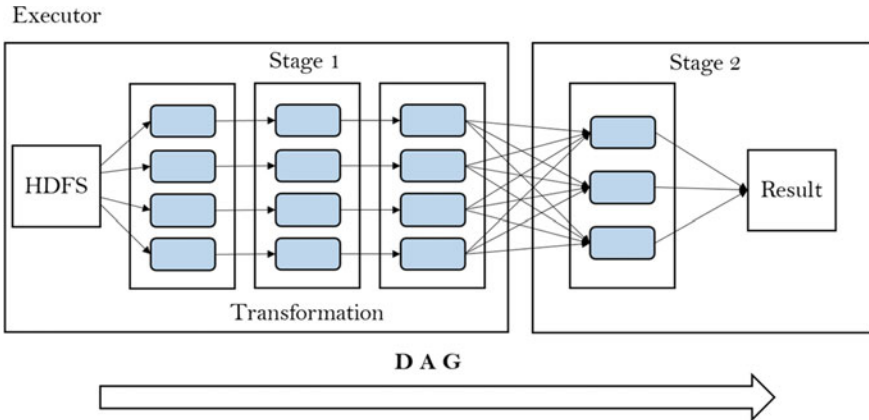


Fig. 1 Spark parallel computing flowchart

2.2 LSTM with Apache Spark

The data is processed by the workflow (Pipeline) [7] that generates the learning model, and the pipeline processing flow is shown in Fig. 2.

Pipeline models are created using learning data through machine learning algorithms in pipeline. Pipeline and pipeline model ensure that feature processing flows of training data and test data are consistent. The pipeline structure includes one or

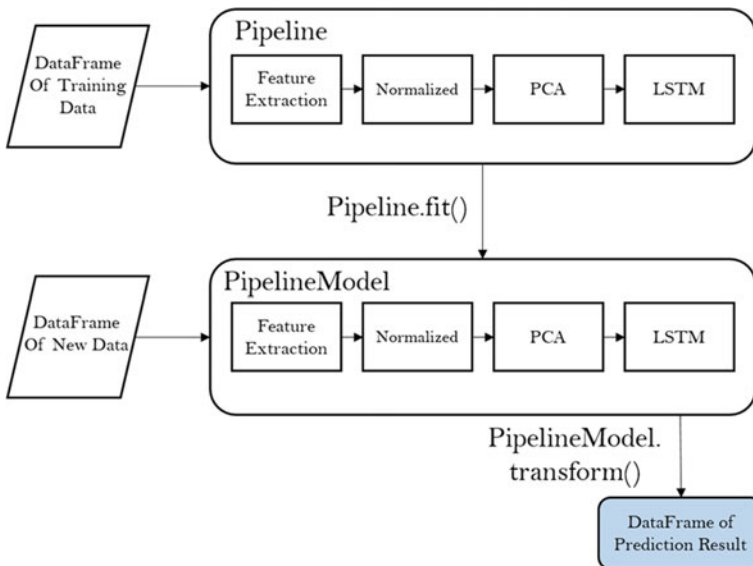


Fig. 2 Training model flow on the pipeline

more stage modules for data processing, each of which performs data processing tasks including feature extraction, normalization processing, and feature dimension reduction. By constructing and generating pipelines in a specific logical order, pipeline machine learning is made more efficient and faster than single-step independent modeling [8].

2.3 Apache Airflow

Airflow is a workflow scheduling and monitoring open-source platform developed by Apache and consists of DAG, which means the flow of data written by the developer in code, and task, which means the unit of work. The operating principles of Apache Airflow are shown in Fig. 3.

The administrator uses the airflow.cfg file to set how often to repeatedly check the port information to use the web server, the DAG script location information to be referenced by the worker node, and the scheduler Dag script file. The scheduler reads the DAG script written by the user in Python code and stores the workflow information of each task in the metadata base, and executes the executor predefined by the user in the airflow.cfg file. The executor finds a task that needs to be executed at this point in time, updates the status information of the task, and executes the task on the worker node [9].

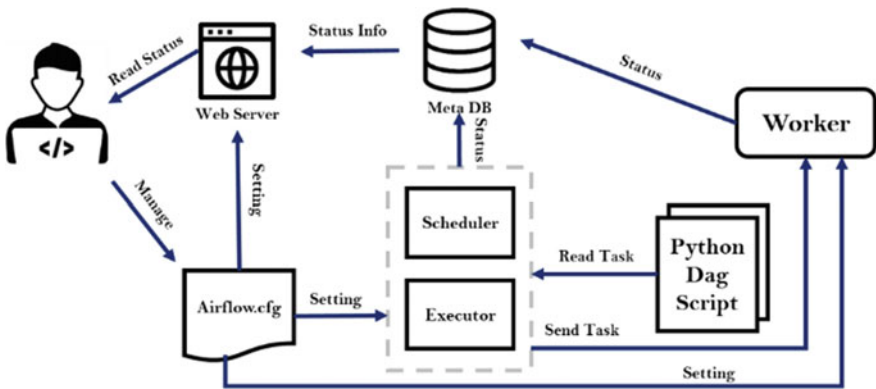


Fig. 3 Apache Airflow workflow architecture

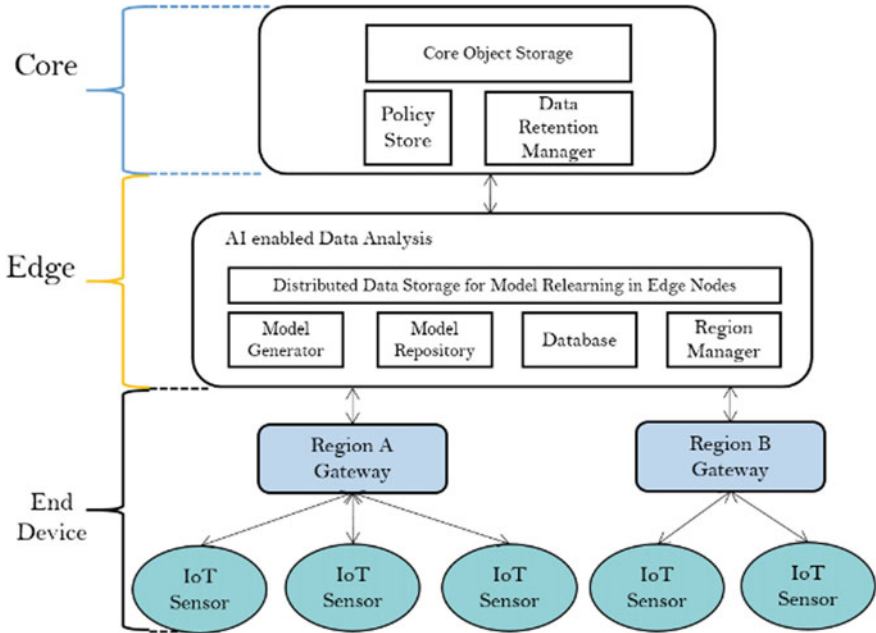


Fig. 4 Overall system configuration

3 Proposed System

3.1 System Architecture

As shown in Fig. 4, the overall system configuration diagram is divided into core, edge, and end device layers. The end device periodically transmits sensor data, which is stored in database at edge cluster. Data stored in database is periodically distributed to storage for model learning by Apache Airflow, and then LSTM model learning is triggered. Thereafter, the trained data is transferred to the core HDFS. There is a REST API that manages the sensor region and a separate dashboard is in operation. Core cloud has object storage that manages all sensor data and manages meta-information that manages learning cycles and data retention.

3.2 Data Reliability Calculate

Figure 5 shows an illustration of the LSTM model created in Apache Spark producing a predicted value. Input data is temperature/humidity sensor data that enters in real time, and it is likely to undergo rapid changes due to the nature of the data. Based on this, more accurate prediction values are derived by assigning higher reliability

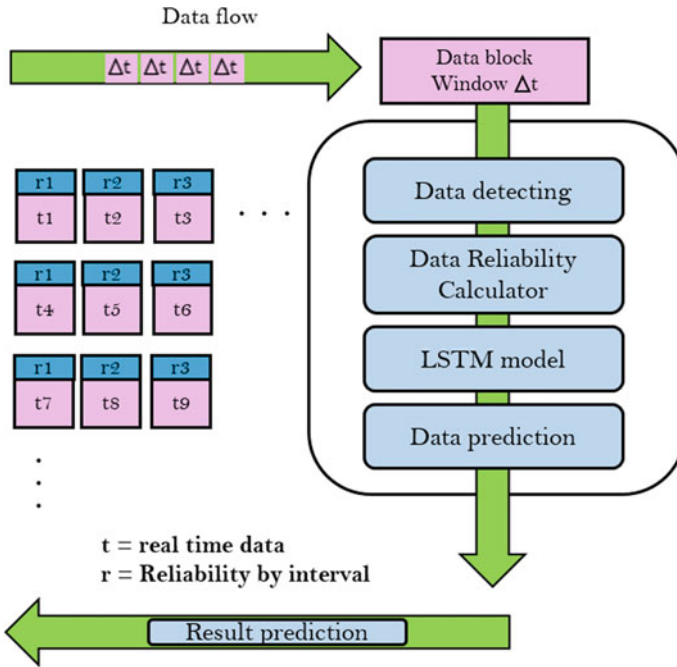


Fig. 5 Real-time LSTM processing

to incoming data. This calculation method is shown by formula (1)

$$\text{Reliability of change over time} = \sum_{n=1}^k (\Delta x_n * a_n) \left(\sum a_n = 1 \right) \quad (1)$$

Δx_n = data variation by interval, a_n = reliability of variation by interval.

4 Experiment Evaluation

In this experiment, a total of four Raspberry Pi are used, and the hardware specifications are shown in Table 1.

In the first experiment, we measure the time which takes to finish model learning from the point of time triggered by Airflow in two scenarios. We observed the time which takes to finish learning sensor data in core cloud and edge cluster, respectively (Table 2).

Table 3 is the absolute value averaged the difference between the actual data and the data predicted by the LSTM model and results of normalized scaling. The measurement results show that the average error increases over time for both models.

Table 1 Edge node hardware specifications

Resources	Specifications
Device model	Raspberry Pi 4 model B
OS	Raspbian GNU/Linux 10 (buster)
CPU	1.5 GHz quad-core 32-bit ARM Cortex-A72
RAM	4 GB
Local storage	128 GB
Spark version	3.3.0

Table 2 After task trigger, the time it takes to complete a task for each dataset size

(epoch = 200)	Units (seconds)			
	500 Mb	1 Gb	2 Gb	5 Gb
Core cloud	2687	6055	16,810	42,934
Edge cluster	2470	5122	13,627	34,192

Table 3 Mean error for datasets for period

Min = 0, max = 1	1 days	7 days	14 days	21 days
LSTM	0.159	0.231	0.251	0.269
Relational LSTM	0.108	0.155	0.167	0.172

However, errors in relational LSTM is lower than the traditional LSTM. This seems the proposed scheme is more suitable for data that change rapidly.

5 Conclusion

In this paper, IoT sensor data is collected, converted/loaded into analysis data, and the time it takes to complete the learning of the predictive model according to the location is measured. We also show the results of increasing the prediction accuracy for sensor data by varying the reliability of data over time. As experimental results show, as the size of the dataset increases, the time it takes to complete the learning increases exponentially. This shows that the entire system needs to be operated through data distributed processing and calculation, not centrally controlled, and that it is necessary to efficiently provide intelligent services to users in an edge computing environment. In addition, the prediction accuracy tends to decrease over time. However, we show that the rate of increase in error by varying the reliability of each data period is lower than that of the existing LSTM model.

In the future, a correlation analysis model between each input data will be created, and the LSTM model will be further developed based on the corresponding coefficient. Through this study, it is expected that it can develop in a positive direction

to increase manufacturing efficiency by early prediction of abnormal conditions in plant factories where numerous IoT sensors are used or by various applications in manufacturing processes that are heavily affected by the external environment.

Acknowledgements This work was supported by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00818, Machine Learning Based Low Bandwidth Image Communication Edge Computing System for Proactive Anomaly Detection on Smart Plant Environment).

References

1. Fatima Z, Tanveer MH, Waseemullah, Zardari S, Naz LF, Khadim H, Ahmed N, Tahir M (2022) Production plant and warehouse automation with IoT and Industry 5.0. *Appl Sci* 12:2053. <https://doi.org/10.3390/app12042053>
2. Nancy AA, Ravindran D, Raj Vincent PD, Srinivasan K, Gutierrez Reina D (2022) IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics* 11:2292. <https://doi.org/10.3390/electronics11152292>
3. Naveen S, Kounte MR (2019) Key technologies and challenges in IoT edge computing. In: Proceedings of the third international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC 2019). IEEE Xplore Part Number: CFP19OSV-ART. ISBN: 978-1-7281-4365-1
4. Huang Y, Fan J, Yan Z, Li S, Wang Y (2022) A gas concentration prediction method driven by a spark streaming framework. *Energies* 15:5335. <https://doi.org/10.3390/en15155335>
5. Lv Z, Qiao L (2021) AI-enabled IoT-edge data analytics for connected living. *ACM Trans Internet Technol* 21(4):Article 104
6. Sheng S, Chen P, Chen Z, Wu L, Yao Y (2021) Deep reinforcement learning-based task scheduling in IoT edge computing. *Sensors* 21:1666. <https://doi.org/10.3390/s21051666>
7. Belov V, Tatarintsev A, Nikulchev E (2021) Choosing a data storage format in the Apache Hadoop system based on experimental evaluation using Apache Spark. *Symmetry* 13:195. <https://doi.org/10.3390/sym13020195>
8. Yangyang Z, Wei N, Meinan W (2021) Research on parallel LSTM algorithm based on spark. In: 2021 IEEE/ACIS 20th international fall conference on computer and information science (ICIS Fall). IEEE. <https://doi.org/10.1109/ICISFALL51598.2021.9627382>
9. Anveshritaa S, Lavanya K (2020) Real-time vehicle traffic analysis using long short term memory networks in Apache spark. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE). IEEE. <https://doi.org/10.1109/ic-ETITE47903.2020.97>
10. Apache Airflow. <https://airflow.apache.org/>

MetaOps: Metaverse Operations for Large-Scale Metaverse Services on Distributed Cloud



Teh-Jen Sun, Seung-Woo Jeong, Hong-Ju Jeong, Choong Seon Hong, Seong-Bae Park, and Eui-Nam Huh

Abstract Recently, the contactless lifestyle has shifted to a new paradigm due to social issues including diseases and natural disasters caused by COVID-19. In this study, the core technology ‘MetaOps’ is derived from ‘Metaverse’ and ‘Operation’, which employed to data processing unit (DPU) in a core–edge distributed cloud environment that enables real-time mutual interaction between human-to-avatar in the Metaverse space. It provides an efficient real-time communication for huge data from Metaverse services. The performance was observed for simultaneous access with hundreds of user interactions and shows practical possibility of large-scale Metaverse services.

Keywords Metaverse · Distributed cloud · Machine learning

T.-J. Sun

Department of Artificial Intelligence, Kyung Hee University, Yongin-si 17104, South Korea
e-mail: dlwmznzll@khu.ac.kr

S.-W. Jeong · H.-J. Jeong · C. S. Hong · S.-B. Park · E.-N. Huh (✉)

Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea
e-mail: johnhuh@khu.ac.kr

S.-W. Jeong

e-mail: jyng2227@khu.ac.kr

H.-J. Jeong

e-mail: sub06038@khu.ac.kr

C. S. Hong

e-mail: cshong@khu.ac.kr

S.-B. Park

e-mail: sbpark71@khu.ac.kr

1 Introduction

As the 4th industrial and online service market expands, Metaverse services that provide various and interactive contents in virtual spaces based on the real world are expanding [1]. Moreover, as the contactless lifestyle has shifted to a new paradigm due to social issues such as COVID-19 and natural disasters, Metaverse is required in various industries including games, finance, entertainment, and education [2]. Therefore, it is necessary to prepare operation platform supporting a large-scale Metaverse service in practical domains.

As the proportion of the software and service market expands [3], there is a limit to the communication speed according to high-volume data communication and physical distance. To solve this problem, a server that provides real-time services in the form of a cloud is required. However, there is a limit to the physical response to handle all these tasks on the central server, so it might be helpful to develop a core-edge distributed cloud. This technology enables users to support real-time online services faster [4].

Core-edge distributed cloud architecture can improve response speed by positioning the computing infrastructure at a point nearby the client [5]. In this study, we intend to conduct a study to the extent that interaction is possible in the Metaverse space accessed by the user at the urban communication distance.

MetaOps, the subject of this study, applies distributed cloud computing to enable real-time human interaction in the next-generation Metaverse space. The goal is to develop a real-time Metaverse operation environment for large-scale Metaverse services that is efficiently enabled by employing a core DPU node in the cloud as a global synchronizer and edge DPU node as a local collector. In addition, it supports realistic user real-time avatar generation to enhance the user experience and also analyzes the user's face to identify the location and motion characteristics of the face. Thus, various neural networks are supported to map realistic features between the user and the avatar.

Using a webcam, 3D avatars are configured based on user face information, and the avatars in the Metaverse move equally according to the pose in which the actual user moves, supporting the construction of a more realistic Metaverse service.

2 Related Work

- **Distributed Cloud Service:** Essential elements of cloud services include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service [6]. The user may use the computing resources they need at any time they want. Applications utilizing cloud computing-based systems enable high-throughput computing (HTC) [7]. And load-balancing technology can be used to distribute the load to prevent overload and reduce resource usage to maintain response time [7].

A distributed server is used in the space where the user is located to support simultaneous access of the user [8]. Each server may control a user to efficiently manage resources.

To optimize the distributed system, a method of reducing data usage was proposed [9]. The method keeps the usage of dynamic tasks occurring in distributed systems. For efficient server management, task scheduling algorithms such as least loaded first (LLF), least flow-time first (LLF), and shortest remaining processing time (SRPT) are applied.

There are two requirements for avatars in MetaOps pursued in the study. One is a realistic avatar 3D model, and the other is a natural avatar movement.

- **Realistic Avatar 3D Model:** In a recent 3D face reconstruction study, a method of generating 3D models with principal component analysis (PCA) for 3D shape and generative adversarial networks (GANs) for texture has emerged, And it render UV maps generated by 3D shape and Texture GAN and optimize them iteratively via convolutional neural networks (CNN). And an improved study is one-shot texture completion (OSTeC). This work produces realistic 3D avatar by re-rendering the 3D model from several angles and optimizing texture UV maps on the derived images.
- **Natural Avatar Movement:** Features were extracted from the user's face picture to generate an avatar expressing user characteristics. In order to detect the user facial movement in the webcam, a face mesh is generated by extracting a feature point of the face. The movement is expressed by changing the distance and angle between the feature points of face mesh. About detect blinking, it can be implemented with an eye aspect ratio (EAR).

Distributed server processing is more efficient instead of central server processing in order to proceed with real-time interaction on the Metaverse based on these various technologies. In this study, we aim for real-time user interaction through hundreds of user in one space on the next-generation Metaverse. Therefore, we propose a service that enables real-time interaction in the Metaverse space using core-edge cloud distributed processing.

3 Method

3.1 Main Framework

The proposed system of MetaOps developed in this study is responsible for distributed processing for high-volume data with core-edge cloud environment of concurrent users and real-time user interactions in Metaverse space. And faster data processing and synchronization are required at the core-edge server, respectively. The core and edge DPU nodes are developed for communication with edge DPU nodes and user devices, respectively.

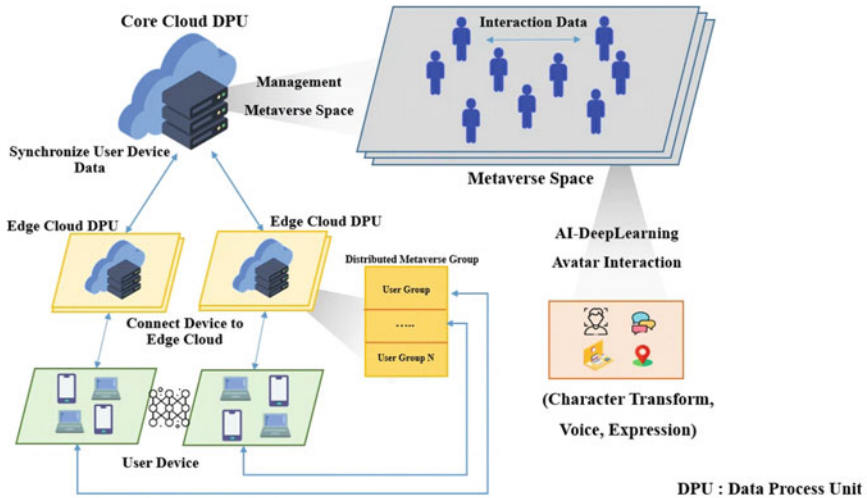


Fig. 1 Metaverse distributed cloud architecture

As shown in Fig. 1, ‘Core Cloud DPU’ (denoted to CCD) processes user interaction data by communicating in real-time the user’s information, game objects, avatars, messages, and voice data in the Metaverse space and controls data delivered by the edge DPU in a timely manner. ‘Edge Cloud DPU’ (denoted to ECD) performs distributed processing within the edge cloud by grouping users who are joined to the service.

3.2 Edge Cloud DPU Node

ECD requires simultaneous access and real-time interaction of hundreds of users in the Metaverse space with distributed cloud. In order to develop ECD, it is essential to support real-time synchronization by transmitting user data to the DPU.

Figure 2 shows a structure diagram of a process in which the DPU performs distributed processing for user data communication efficiently in ECD. First, when the edge DPU receives the user’s interaction data in services, the users are divided into N groups in ‘Group Distribution’. The groups are controlled by the thread pool, so that manages resource scheduling of groups efficiently. Second, ‘Checksum User process’ validates each user group data. Third, ‘Data Pooling’ groups validated data for ‘Packet Generation’. The grouped data is allocated to the memory buffer in real time. And, ‘Data Sharing’ sends processed data to the ‘Core Clod DPU’ to exchange user data to other edge DPUs. Finally, the ‘Core Cloud DPU’ calculates amount of the received traffic and memory usage and dynamically transmits data to the edge DPUs for providing efficient data interactions among users.

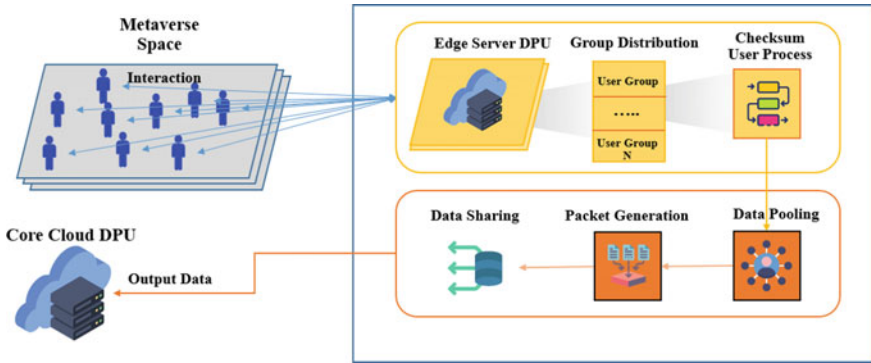


Fig. 2 Distributed Metaverse process in edge cloud

3.3 3D Metaverse Avatar

To develop a more realistic Metaverse services, we propose a method that applies real-world user information to the Metaverse space by linking the user object with the Metaverse game object, especially remote, realistic education service in Metaverse space (Fig. 3).

First, our system used CNN-based deep learning technology for ‘Detect User Face’ using webcam. Second, ‘Data Processing’ applies the information to ‘update 3D avatars’. Finally, our system expresses 3D avatars in Metaverse space by detecting the user’s movement at 60 frames per second in real-time.

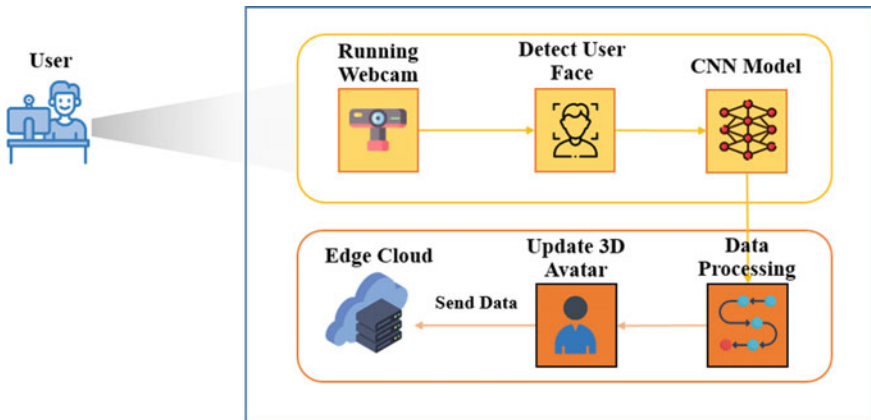


Fig. 3 Detect user face and update 3D avatar in user device

4 Evaluation Metrics and Analysis

Simulations were performed to verify the technology of MetaOps proposed in this study. For measuring MetaOps performance, server processing speed and CPU usage were selected as the main evaluation list. Each server is assigned a virtual machine (VM) to the cloud platform to establish a server communication environment.

The experimental environment for DPUs consists of 2VM with 4 GB RAM in Ubuntu. To verify the performance of the core-edge DPU, the number of simultaneous user connections, DPU node processing speed, user interaction speed, latency, and CPU usage of the server are measured.

Performance Analysis

We develop dummy clients for accurate measurement. Performance analysis is conducted by assessing simultaneous 10-50-100-200-300 user sessions in the Metaverse space.

4.1 DPU Node—CPU Performance

The performance and CPU usage were measured when processing the user data in DPU node after transmitting data to the Edge server.

The DPU node initially recorded a speed of 5 ms among 10 and 50 simultaneous connections, but the speed decreased after 200 connections, recording an average speed of 21 ms at 300 connections. The server is overloaded when more than 300 users are joined. CPU usage of the servers increased significantly when 100 users were connected to cloud server, recording 18% and 38% utilization of the CCD node and ECD node, respectively. And in case of 300 users, 32 and 70% usages are observed at each DPU node. In the CPU usage of the CCD node and ECD node, the user data is first transmitted to the ECD node when the user synchronizes in the Metaverse space. At this time, it is observed that ECD nodes for data pre-processing of 300 users' interactions performs 28 ms. After that, the noise-filtered data was transferred to CCD, and the processing in CCD becomes 21 ms, relatively low (Figs. 4 and 5).

4.2 User Interaction—Client Performance

We analyze user synchronization speed and client performance during user interactions in Metaverse space.

As shown in Fig. 6, the synchronization speed is increased in linear by 100 users and smoothly increased. User synchronization time increases in proportion to the number of user data.

Fig. 4 DPU node processing result

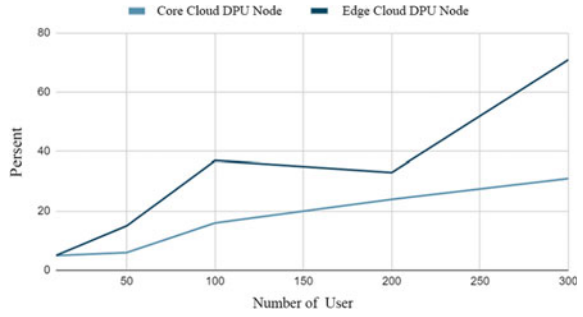


Fig. 5 Node CPU

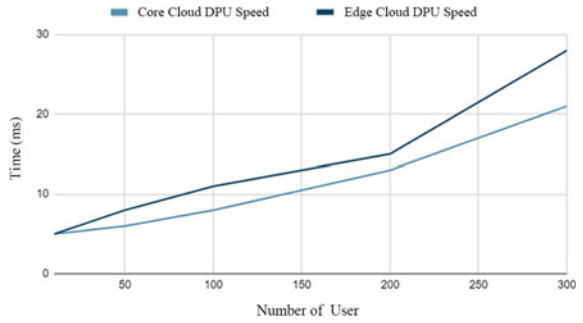
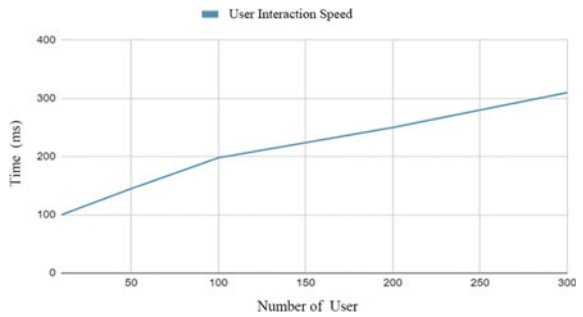


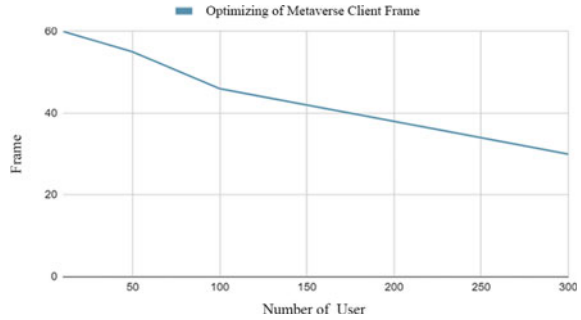
Fig. 6 User interaction synchronization



Since frame optimization of the user scene is also affected by the ECD processing speed, the number of frames decreases at about 100 users. And the average number of the video frame at 300 simultaneous user connections is observed by 30 frames (Fig. 7).

The observed performance of the proposed system by developing a core–edge DPU (CCD and ECD) shows accommodating up to 300 users successfully in the Metaverse service.

Fig. 7 Optimizing of Metaverse connection clients



5 Conclusions

In this paper, we propose a distributed processing technique that can effectively interact in the Metaverse space using core–edge cloud and DPU nodes. After processing of the user’s information in the edge cloud DPU node, the global processing is performed in the CCD node that optimizes the user synchronization time.

Through the implementation and performance of MetaOps technology that supports the services of distributed cloud-based large-scale Metaverse services, it showed an average of 200 ms of interaction between users and up to hundreds of user accessible performance. And in future work, we will implement intent base networking (IBN) and improve the performance of DPU processing speed and CPU usage. We also will study MetaOps optimization technology enabling more than 1000 users’ connections.

Acknowledgements This work was partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) [No. RS-2022-00155911, Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University)] and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1A4A1018607).

References

1. Wang Y, Lee LH, Braud T, Hui P (2022) Re-shaping post-COVID-19 teaching and learning: a blueprint of virtual-physical blended classrooms in the metaverse era. In: The 1st international workshop on social and metaverse computing and networking in conjunction with IEEE ICDCS’22. Bologna, 10–13 July, Italy
2. Kim JH, Lee BS, Choi SJ (2022) A study on metaverse construction and use cases for non-face-to-face education. *J Convergent Cult Technol* 8(1):483–497
3. Ojala A, Tyrväinen P (2007) Market entry and priority of small and medium-sized enterprises in the software industry: an empirical analysis of cultural distance, geographic distance, and market size. *J Int Mark* 15(3):123–149

4. Endo PT, de Almeida Palhares AV, Pereira NN, Goncalves GE, Sadok D, Kelner J, Mangs JE (2011) Resource allocation for distributed cloud: concepts and research challenges. *IEEE Netw* 25(4):42–46
5. Taleb T, Samdanis K, Mada B, Flinck H, Dutta S, Sabella D (2017) On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration. *IEEE Commun Surv Tutor* 19(3):1657–1681
6. Mell P, Grance T (2011) The NIST definition of cloud computing
7. Rashid ZN, Zebari SR, Sharif KH, Jacksi K (2018) Distributed cloud computing and distributed parallel computing: a review. In: 2018 international conference on advanced science and engineering (ICOASE). IEEE, pp 167–172
8. Ng B, Si A, Lau RW, Li FW (2002) A multi-server architecture for distributed virtual walk-through. In: Proceedings of the ACM symposium on virtual reality software and technology, pp 163–170
9. Semchedine F, Bouallouche-Medjkoune L, Aissani D (2011) Task assignment policies in distributed server systems: a survey. *J Netw Comput Appl* 34(4):1123–1130

A Study on the Improvement of Internet Comment Accuracy Using Deep Learning



Heeji Park, Jimin Ha, and Jungho Kang

Abstract Research on comment classification algorithms using AI is frequently conducted to solve the problem of malicious comments. However, these algorithms have a problem such that the accuracy of comment classification is low due to overfitting. Accordingly, this study is premised on Deep Learning Classification Algorithm Based on Crowdsourcing (DCAC) and the system using the DCAC algorithm to improve the problem of low malicious comment classification and low judgment accuracy due to overfitting, using crowdsourcing. Furthermore, it was confirmed by testing that the proposed method with 92% accuracy helps to increase the judgment accuracy reduced by overfitting.

Keywords Crowdsourcing · AI · Deep learning · Overfitting · Malicious comments

1 Introduction

The development of information technology not only allows to convey information quickly on the Internet but also allows freedom of opinions in real-time by writing posts or posting comments through SNS. However, there have been a series of damages including posting malicious comments, such as obscenity, disparagement, and abusive language to offend others by exploiting the anonymity on SNS [1]. In fact, it has emerged as a social problem as the damage of malicious comments has occurred such as the suicide of celebrities, and the damage of malicious comments

H. Park · J. Ha · J. Kang (✉)

Department of Software Security Convergence, Baewha Women's University, Seoul, Republic of Korea

e-mail: kangsamm@naver.com

H. Park

e-mail: pheeji@naver.com

J. Ha

e-mail: basejm07@naver.com

has spread in entertainment, sports, and politics. In addition, problems of malicious comments are occurring in various forms such as abusive language and remarks that cause racial conflict. To solve this problem, researchers are proposing various AI comment deletion models and AI-based malicious comment detection models. But AI fundamentally has overfitting which reduces accuracy when learning new data in case of overlearning certain test data [2].

In this paper, we propose ‘Deep Learning Classification Algorithm Based on Crowdsourcing (DCAC)’, which improves the problem of comment classification and judgment accuracy degradation due to overfitting using the general public’s opinion and demonstrates accuracy improvement through experiments.

2 Related Work

2.1 BERT

Bidirectional Encoder Representations from Transformers (BERT) is the bidirectional language model developed by Google for Natural Language Processing (NLP) [3]. BERT was pre-trained with a 3.3-billion-word corpus, providing convenience to use by fine-tuning the required tasks. Furthermore, BERT determines the meaning of the word after grasping the entire context through the process of transforming the sentence into a token without looking at each word in the input sentence independently. This feature reduces the probability of ambiguous interpretation of words and increases classification accuracy.

2.2 Crowdsourcing

Crowdsourcing is a combination of crowd and outsourcing, a web-based business model devised by Jeff Howe and Mark Robinson in 2006. Crowdsourcing is the process of outsourcing through public participation [4]. A prime example of crowdsourcing is Wikipedia, which aims to produce a free Internet encyclopedia. Wikipedia is the world’s largest WikiWikiWeb that allows users to participate in creating and modifying content [5].

2.3 Malicious Comments Detection Method Using AI

There is a long history of research for detecting and preventing malicious comments. Through research using SVM and sentiment analysis, Hong et al. [6] calculated the malicious index with a sentiment dictionary specialized in Internet comments

and suggested the technique to detect malicious comments by deriving a calculation formula for determining malicious comments with SVM. Van Huynh et al. [7] proposed a model that separates into three categories, 1D CNN, bidirectional LSTM, and bidirectional GRU, through research based on the Bi-GRU-LSTM-CNN classifier, and effectively detected malicious comments by supervised learning this model.

3 Proposed System

In this paper, we propose Deep Learning Classification Algorithm Based on Crowdsourcing (DCAC) and a comment classification system based on the corresponding algorithm.

3.1 DCAC-Based Comment Classification Scheme

The comment classification system in Fig. 1 consists of three stages: preprocessing step, training step and DACA step. In the preprocessing step, the collected data is divided into train data and test data for data cleaning. In the training step, Training Data is processed through embedding, self-attention, and subword tokenizer to increase the accuracy of deep learning, and BERT trained it. Finally, in the DCAC step, the AI Judgment generated by test data and trained AI and the CS Judgment obtained through crowdsourcing are cleaned by Malicious CS, and then the two delegate values are compared to determine the final value. CS means crowdsourcing, and delegate indicates the representative values of AI and CS. Final value is divided into normal comments and bad comments, and bad comments are further divided into spam, abuse, disparagement, and obscenity. The following is a detailed description of the DCAC step.

True) If the AIDelegate and CSDelegate are the same, the AIDelegate is selected as the Final Value.

False) If the AIDelegate and CSDelegate are different, the Final Value is made as follows.

- (a) When the AI Ratio and CS Ratio have the ratios of 4:6, all classification ratios are calculated using the $F(\text{AI Ratio}, \text{CS Ratio})$.
- (b) The calculated two Ratios are compared, and the Delegate is selected as the Final Value.

The DCAC in which the CS Judgement process is added may be free from the problem of overfitting compared to existing algorithms. Using the advantages of

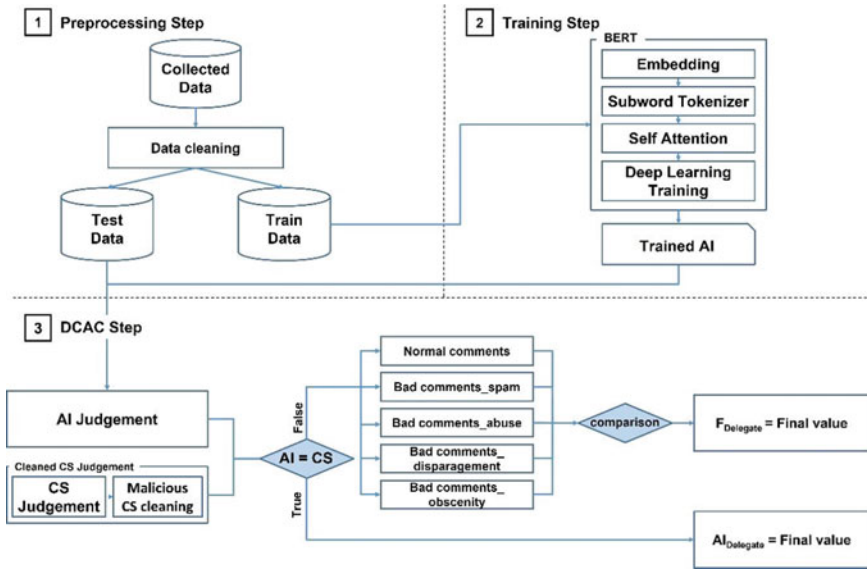


Fig. 1 DCAC-based comment classification scheme

DCAC, malicious comments can be more accurately identified than existing AI-based comment classification models, and it is possible to create a correct comment environment.

4 Experiment and Related Results

4.1 Environment

The co-laboratory development browser and KoBERT among several models were selected as the development environment. As test data, comments on Twitter and YouTube were collected, and 10,000 comments were randomly selected. 8000 Training Data and 2000 Judgement Data were chosen from the selected comments. At this time, the ratio used to divide Training Data and Judgement Data is 8:2, which is mainly used to compare overfitting error rates. By analyzing 10,000 data directly, each Z_Value of the data is determined, and accuracy is measured using Z_C . Z_Value means correct value and Z_C means a set of Z_Values of all comments. For the Judgment Data used for CS Judgment, questionnaires were received from 30 individuals and crowdsourcing was conducted. As a result of checking the collected crowdsourcing, there were cases of cursory participation, such as checking by one classification. In this case, the weight of the malicious crowdsourcing value was lowered through the Malicious CS cleaning process.

4.2 Experiment

The test was conducted provided that CS Judgment was superior to AI Judgment, but the Malicious CS cleaning process was conducted considering that exceptions (i.e., Malicious crowdsourcing) could occur. And the ratio between Judgment and CS Judgment is 4:6, and the accuracy was measured using it. At this time, the classification and judgment accuracy of malicious comments were calculated as

$$\frac{\text{Number matching } Z_C}{\text{Total number of comments}} \times 100.$$

4.3 Result

Figure 2 is a graph showing the accuracy of experiments in eight cases from 0.02 to 0.20 to determine the value used in the Cleaned CS Judgment to lower the proportion of maliciously prepared CS Judgment values. It can be seen that the value to be used for Cleaned CS Judgment increases rapidly at 0.5 and then gradually increases the highest at 0.12 and then decreases gradually.

Table 1 is the result of comparing AI/CS Judgment Decision and Z_C . AI Judgment showed 79.2% accuracy with a total of 1584 comment categories matching Z_C , CS Judgment showed 86.8% accuracy with a total of 1736 comment categories matching Z_C , and Cleaned CS Judgment showed 89.95% accuracy with a total of 1799 comment categories matching Z_C .

Table 2 compares the accuracy of the comment classification system using DCAC with the accuracy of the comment classification system using other algorithms, including AI and crowdsourcing accuracy. It was confirmed that the value of the

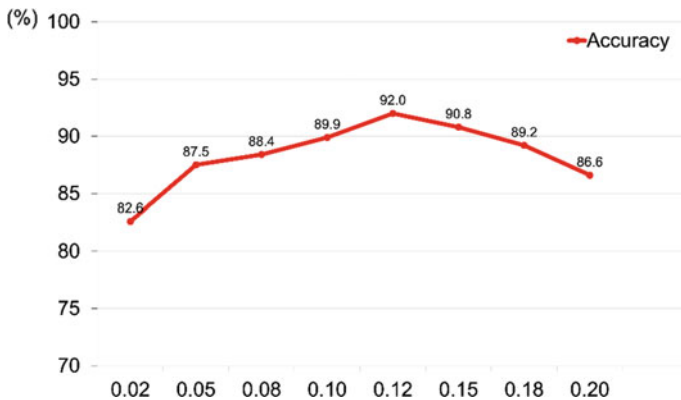


Fig. 2 Graph of accuracy comparison based on judgment ratio

Table 1 Accuracy comparison table between AI, CS, cleaned CS judgment, and Z_C

	Normal comment	Bad comment				Total	Accuracy (%)
		Spam	Abuse	Obscenity	Disparagement		
Z _C	550	365	397	300	388	2000	100
Number of AI judgment matching Z _C	389	277	305	295	318	1584	79.2
Number of CS judgment matching Z _C	424	300	385	300	327	1736	86.8
Number of cleaned CS judgement matching Z _C	452	320	385	300	342	1799	89.95

Table 2 Judgment accuracy comparison table

Algorithm	AI	SVM algorithm [6]	Bi-GRU-LSTM-CNN algorithm [7]	CS	Cleaned CS judgement	DCAC
Accuracy (%)	79.2	87.8	70.576	86.8	90.0	92.0

DCAC malicious comment classification and judgment accuracy experiment was 92%, an increase of 5.2% over CS Judgment accuracy, and a noticeable improvement in accuracy to 12.8% over AI Judgment accuracy.

Since the experiment was not conducted in the same environment, absolute comparison is difficult, but the results of experiments in other papers were compared and presented for relative comparison. Through experiments, it can be confirmed that the accuracy of DCAC’s comment classification is improved compared to the existing comment classification algorithm.

5 Conclusion and Follow-Up Research

This test proved that the comment classification system using DCAC has higher accuracy in comment classification than other algorithms. DCAC which used the people’s judgment obtained through the crowdsourcing process found additional malicious comments that were not identified due to overfitting. Although there was a limitation that malicious crowdsourcing could be less accurate than conventional

algorithms due to the nature of DCAC where humans intervene, these weaknesses were compensated by ‘Malicious Crowdsourcing Cleaning’ that compares and calculates the correct answer value (results judged by Unsupervised Learning AI) with the results of crowdsourcing. This paper presents a new direction for solving the overfitting problems and improving the accuracy of comment classification, and it is judged that it will help to form a good comment culture.

References

1. Ybarra ML, Mitchell KJ (2004) Youth engaging in online harassment: associations with caregiver–child relationships, Internet use, and personal characteristics. *J Adolesc* 27(3):319–336
2. Santos CF, Papa JP (2022) Avoiding overfitting: a survey on regularization methods for convolutional neural networks. *ACM Comput Surv (CSUR)* 54(10s):1–25
3. Devlin J et al (2018) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
4. Brabham DC (2008) Crowdsourcing as a model for problem solving: an introduction and cases. *Convergence* 14(1):75–90
5. Potthast M (2010) Crowdsourcing a Wikipedia vandalism corpus. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval
6. Hong J, Park J, Choi J (2016) A malicious-comment detection technique on the Internet using the support vector machine. *International Information Institute (Tokyo). Information* 19(6B):2383
7. Van Huynh T et al (2019) Hate speech detection on Vietnamese social media text using the Bi-GRU-LSTM-CNN model. arXiv preprint [arXiv:1911.03644](https://arxiv.org/abs/1911.03644)