IWIN2023

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 1-4, 2023
Onuma, Hokkaido

Informatics Society

Sponsored by Informatics Society

IWIN2023

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 1-4, 2023
Onuma, Hokkaido

Informatics Society

Sponsored by Informatics Society

Publication office:
Informatics Laboratory
3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan
Publisher:
Tadanori Mizuno, President of Informatics
ISBN:
978-4-902523-50-8
Printed in Japan

Table of Contents

Session 5: Multimedia and Communication ······ 171

(Chair: Yuichi Tokunaga) (13:20 - 15:20, Sep. 3)

Message from the General Chairs

It is our great pleasure to welcome all of you to Mori-machi, Hokkaido, Japan for the 17th International Workshop on Informatics (IWIN2023). This workshop has been held annually by the Informatics Society. Since 2007, the workshops were held in Naples in Italy, Wien in Austria, Hawaii in the USA, Edinburgh in Scotland, Venice in Italy, Chamonix in France, Stockholm in Sweden, Prague in Czech Republic, Amsterdam in Netherlands, Riga in Latvia, Zagreb in Croatia, Salzburg in Austria, Hamburg in Germany, Wakayama in Japan (virtually), Fukui in Japan (virtually), and Wakayama in Japan respectively.

In IWIN2023, 27papers were accepted after peer reviewing by the program committee. Based on the papers, six technical sessions were organized in a single-track format, which highlighted the latest research results in the areas such as Systems and Applications, Intelligent Transportation Systems, Network and Security, Sensing and Analysis, and Multimedia and Communication. IWIN2023 will also welcome two keynote speakers: Dr. Hidetoshi Mishima, Executive Fellow, Corporate Research and Development Group, Mitsubishi Electric Corporation and Professor Emeritus. Dr. Hitoshi Aida, The University of Tokyo. We greatly appreciate their participation in our workshop.

We would like to thank all the participants and contributors who made the workshop possible. It is indeed an honor to work with a large group of professionals around the world to make the workshop a great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN2023.

September 2023

Yoshia Saito
Akihiro Hayashi

Organizing Committee

General Co-Chairs

Akihiro Hayashi (Shizuoka Institute of Science and Technology, Japan)
Yoshia Saito (Iwate Prefectural University, Japan)

Steering Committee

Hitoshi Aida (The University of Tokyo, Japan)
Toru Hasegawa (Osaka University, Japan)
Teruo Higashino (Kyoto Tachibana University, Japan)
Tadanori Mizuno (Aichi Institute of Technology, Japan)
Jun Munemori (The Open University of Japan, Japan)
Yuko Murayama (Tsuda University, Japan)
Ken-ichi Okada (Keio University, Japan)
Norio Shiratori (Chuo University / Tohoku University, Japan)
Osamu Takahashi (Future University Hakodate, Japan)

Program Chair

Kei Hiroi (Kyoto University, Japan)

Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

Publicity Chair

Yoshitaka Nakamura (Kyoto Tachibana University, Japan)

Program Committee

Akihito Hiromori (Osaka University, Japan)
Akra Uchiyama (Osaka University, Japan)
Hideki Goromaru (Chiba Institute of Technology, Japna)
Hironobu Abe (Tokyo Denki University, Japan)
Hiroshi Mineno (Shizuoka University, Japan)
Hiroshi Sugimura
(Kanagawa Institute of Technology, Japan)
Hiroshi Yoshiura (Kyoto Tachibana University, Japan)
Kanae Matsui (Tokyo Denki University, Japan)
Katsuhiko Kaji (Aichi Institute of Technology, Japan)

Katsuhiro Naito (Aichi Institute of Technology, Japan)
Kazuyuki Iso
(Tokyo Information Design Professional University, Japan)
Kei Utsugi (Hitachi, Ltd, Japan)
Keiichi Abe (Kanagawa Institute of Technology, Japan)
Ken Ohta (Miyagi University, Japan)
Kozo Okano (Shinshu University, Japan)
Makoto Imamura (Tokai University, Japan)
Masaji Katagiri (iU, Japan)
Masaki Endo (Polytechnic University, Japan)
Masaki Nagata (Shizuoka University, Japan)
Masashi Saito
(Kanazawa Institute of Technology, Japan)
Michiko Oba (Future University Hakodate, Japan)
Mikiko Sode
(National Institute of Technology, Niihama College, Japan)
Minoru Kobayashi (Meiji University, Japan)
Naoya Chujo (Aichi Institute of Technology, Japan)
Nobuhiro Kobayashi (University of Nagasaki, Japan)
Orie Abe
(Information Communication Media Laboratory, Japan)
Ryozo Kiyohara (Kanagawa Institute of Technology, Japan)
Satoru Matsumoto (Osaka University, Japan)
Shigemi Ishida (Future University Hakodate, Japan)
Shigeyoshi Ohno (Polytechnic University, Japan)
Shinji Kitagami (Fukui University of Technology, Japan)
Takaaki Umedu (Shiga University, Japan)
Takayasu Yamaguchi (Akita Prefectural University, Japan)
Takaya Yuizono
(Japan Advanced Institute of Science and Technology)
Takuya Yoshihiro (Wakayama University, Japan)
Tetsushi Ohki (Shizuoka University, Japan)
Tetsuya Shigeyasu
(Prefectural University of Hiroshima, Japan)
Tetsuya Yokotani (Kanazawa Institute of Technology, Japan)
Tomoki Yoshihisa (Shiga University, Japan)
Tomoo Inoue (University of Tsukuba, Japan)
Tomoya Kitani (Shizuoka University, Japan)
Tomoyuki Yashiro (Chiba Institute of Technology, Japan)
Toshihiro Wakita (Kanagawa Institute of Technology, Japan)
Tsukasa Kudo
(Shizuoka Institute of Science and Technology, Japan)
Yoshia Saito (Iwate Prefectural University, Japan) Yoshiaki
Terashima (Soka University, Japan)
Yoshinobu Kawabe (Aichi Institute of Technology, Japan)
Yoshitaka Nakamura (Kyoto Tachibana University, Japan)
Yu Enokibori (Nagoya University, Japan)
Yuki Koizumi (Osaka University, Japan)
Yukimasa Nagai (Mitsubishi Electric Corporation, Japan)
Yuichi Bannai (Kanagawa Institute of Technology, Japan)
Yuichi Tokunaga
(Kanazawa Institute of Technology, Japan)
Yusuke Gotoh (Okayama University, Japan)
Yusuke Ichikawa (Saitama University, Japan)
Yusuke Takatori (Kanagawa Institute of Technology, Japan)

Session 1: Systems and Applications 1

(Chair: Katsuhiko Kaji)

A Study on a Power Interchange Method for Realizing Net-Zero Energy for Multiple Small Communities

Masashi SAITO*, Tomoki NOMURA** and Yuichi Tokunaga*

*Faculty of Graduate School of Engineering, Kanazawa Institute of Technology, Japan
** Graduate School of Engineering, Kanazawa Institute of Technology, Japan
{msaito, y.tokunaga}@neptune.kanazawa-it.ac.jp,
c6101950@planet.kanazawa-it.ac.jp

Abstract - We have developed a simulator to achieve net-zero energy through power sharing among multiple households and communities. In multi-household transactions, when the capacity of one household's electricity storage system is depleted and power supply becomes necessary, power can be purchased from other households to meet demand without having to purchase power from the grid, thus enabling power operations that do not rely on grid power. In transactions the scale of transactions is extended to multiple communities, a certain increase in the renewable energy consumption rate is confirmed even in communities with low storage battery capacity and photovoltaic power generation capacity through the inter-community exchange of electricity by scattering large-scale facility households with different electricity consumption styles. In this study, only data from the Hokuriku region is used for electricity consumption and solar power generation. By replacing these data, the feasibility of net-zero energy can be verified for all regions, showing that a wide variety of simulations in social systems is possible.

Keywords: Smart Community, Electric Power Management, Net Zero Energy House, Renewable Energy, Simulation

1 INTRODUCTION

The introduction of renewable energies in Japan has rapidly increased with the introduction of the feed-in tariff (FIT) program, which began in November 2009, requiring electric power companies to purchase electricity generated from renewable energies (solar, wind, hydro, geothermal, and biomass) for a certain period at a price set by the government. For residential PV (residential photovoltaic power generation) of less than 10 kW, surplus power after self-consumption is eligible for purchase for 10 years. The system will gradually expire from November 2019. According to the Agency for Natural Resources and Energy, the total number of residential PV systems that will expire under the FIT will be 730,000 (2.82 million kW) in 2020 and 1.65 million (6.7 million kW) in 2023[1].

The price of electricity sold during the FIT period was 48 yen/kWh in FY 2009. After the end of the FIT, the price will be about 8 yen/kWh, and compared to the purchase price of about 20 yen/kWh, it is considered more economical to consume the electricity generated on site. For this reason, attention is focused on self-consumption houses that generate electricity during the daytime when the sun is rising, store the surplus in storage batteries without selling it to the grid, and discharge the stored electricity at night.

A house that focuses on self-consumption is similar to a net zero energy house (henceforth ZEH). ZEH is a house that aims for zero primary energy consumption with the three characteristics of self-consumption, thermal insulation, and energy saving[2]. The use of air conditioners is reduced by improving the insulation performance of the house, and the entire house is made energy efficient by high-efficiency equipment and HEMS (Home Energy Management System). Since the electricity consumption is low, a combination of PV power generation and storage batteries is expected to be effective.

Storage batteries play an important role in realizing ZEH. The price of storage batteries is on a downward trend, but the price decline will slow down when the cost of raw materials for lithium-ion batteries is taken into account. Instead of installing storage batteries for home use, storage batteries installed in EVs could be used as storage batteries for home use. If EVs are the vehicles used on a daily basis, this is V2H (Vehicle to Home) which links EVs and homes[3]. The storage batteries of EVs that have reached the end of their service life could be used as a substitute for household storage batteries. In terms of used EVs, EVs that went on sale around 2010 have reached the end of their service life and are beginning to appear on the market.

Furthermore, the unstable global situation is causing electricity prices to skyrocket. According to the Agency for Natural Resources and Energy, the Japanese electricity market, which was at about 8 yen/kWh until September 2021, recorded about 26.2 yen/kWh in March 2022[4]. Against this background, a shift to distributed energy systems using renewable energy sources that do not depend on existing centralized energy systems is being promoted[5]. The operation of distributed energy systems is expected to realize "net zero energy".

Smart grids and microgrids are effective power transmission systems for building distributed energy systems. A smart grid is a power transmission system that enables control of the electricity flow not only on the supply side but also on the demand side, and this system makes it possible for consumers to buy and sell electricity from the supply side or between consumers. A microgrid is an electric power network in which a smart grid is deployed within a certain region and energy is produced and consumed locally within the community without relying on external supplies. Microgrid operations that do not rely on electricity sales from suppliers can be expected to secure electricity and heat in the event of a disaster and to reduce electricity costs.

In addition, the need to transmit electricity over long distances in rural areas, including mountainous regions of Japan, has raised concerns about the rising cost of infrastructure maintenance. By operating a microgrid and building an independent power infrastructure zone using renewable energy generation within the surrounding area, it is possible to operate power without relying on an external power supply even when the area is remote from urban areas, eliminating the need to manage the infrastructure of a centralized energy system. As a result, infrastructure maintenance costs can be significantly reduced [6].

In this study, we have first developed a simulator that assumes the installation of a PV power generation and storage system in a house for the realization of ZEH and evaluate the economic feasibility by changes in the amount and cost of the system installation. Next, assuming a small-scale community in a mountainous area, we demonstrate the feasibility of independent power operation and economical power management only in the community and its surrounding area when a microgrid is actually constructed, and discuss the evaluation results using the simulator.

2 RELATED WORK

Ishikawa and Matsuo [7] conducted a simulation using storage battery capacity, connection and charge/discharge control methods, and water heater boil-up time as parameters to evaluate the primary energy savings and economic efficiency of installing storage batteries in PV equipment in post-FIT homes. The best results were obtained by shifting the boiling time of the heat-pump water heater without installing the storage battery system, which resulted in an annual profit of 0.37 million yen and a reduction in primary energy consumption of 14 MJ (3.9 kWh). The paper also argues that if 3 kWh of storage batteries are installed at a unit price of 8 yen/kWh, one of the following conditions must be met: the storage battery installation cost must be less than 150,000 yen, the durability period must be extended to 28 years, or the electricity price must rise to 186% of the current price.

Yabe et al. [8] evaluated the economic efficiency of storage batteries by determining the changes in the self-consumption rate and the annual cost before and after the introduction of storage batteries for each consumer based on electricity demand data and actual solar power generation data for residences and businesses with solar power generation throughout Japan. The study concluded that a 5kWh storage battery at 60,000 yen/kWh can be installed in a house, and that the investment can be recovered in about 15 years. The payback period is about 11 years when electricity rates with late-night starting times are selected and varies depending on the demand characteristics and other factors. For commercial customers, the payback period is about 8 years when a 60,000 yen/kWh storage battery with the PV power generation output for 30 minutes is installed.

Goto et al. [9] proposed a framework and a solution method for the problem of determining the supply-demand operation plan of a microgrid by calculating the operating reserve based on the forecast error trends of power demand and PV power output. The validity of the proposed framework is verified through numerical simulations. The demand-supply operation planning problem is formulated as a demand-supply operation planning problem for a general microgrid. In addition, they defined a demand-supply operation planning problem that considers uncertainty by using the probability distribution of the net demand forecast. In the simulation, the operational cost for the forecasted value of net demand resulted in worse solution results compared to the conventional method. On the other hand, the solution succeeded in reducing the expected value of operational cost by approximately 20% compared to the conventional method.

Komiyama et al. [10] evaluated the optimal combination of power generation facilities, considering the introduction of renewable energy and inter-regional power transmission. Wind power generation facilities in Japan are concentrated in regions such as Hokkaido and Tohoku, and there is a need to develop inter-regional power transmission lines. Therefore, an optimal power generation model was developed as a linear programming model that considers the transmission of power using inter-regional transmission lines. As a result, it was shown that wind power can be supplied to the Tokyo metropolitan area, where solar power is the main source of renewable energy supply, by expanding the regional power grid. It was also shown that the introduction of more PV power generation facilities in the Kyushu and Shikoku regions, which have long hours of sunlight, should be promoted.

3 DEVELOPMENT OF ZEH ECONOMIC EVALUATION SIMULATOR

3.1 ZEH Model

This section discusses a house in which a PV power generation system and an energy storage system are installed. Electricity is supplied to the house from the amount of electricity generated by the PV power generation system. If consumption exceeds the amount of electricity generated, the storage batteries are discharged to make up for the power shortfall. If the amount of electricity generated and discharged from the storage batteries is still insufficient, the remaining shortage is met by purchasing electricity from the grid. Surplus power from solar power generation is charged into storage batteries, and surplus power that cannot be charged due to full battery is sold to the grid. These power transfers are controlled through HEMS.

We will also consider the use of electricity consumption forecasting and PV power generation forecasting. Smart meters measure residential electricity consumption, and PV equipment sensors measure power generation and solar radiation. The power consumption forecast, PV power generation forecast, and remaining storage battery capacity can be used to predict future excesses and deficiencies in power, enabling power control to minimize power costs.

The assumed ZEH is shown in Figure 1. In this study, the power transfer in Figure 1 is first simulated.

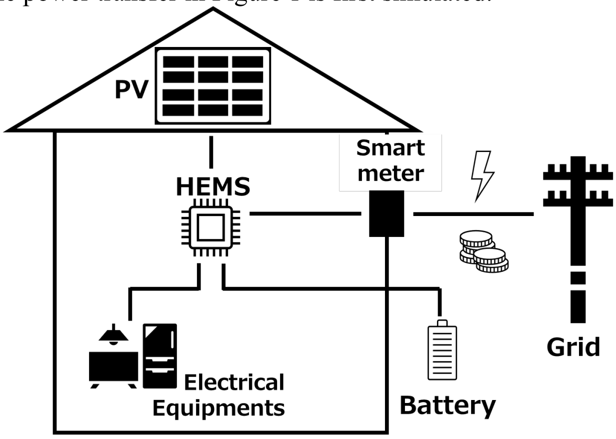


Figure 1: ZEH Model

3.2 Overview of ZEH Economic Evaluation Simulator

The data used in the simulation, such as the amount of electricity consumption and solar power generation, are recorded in CSV and DB formats. When parameters such as the amount of installed PV and energy storage systems and installation costs are entered into the simulator and executed, the data is read and the power transfer simulation is started. Power transfer calculations are performed in every 30-minute period over a year. After the simulation is complete, a graph is plotted showing the power transfer and profit/loss. The resulting data can be downloaded in CSV format.

The function and configuration of the simulator developed in this study are shown in Figure 2.

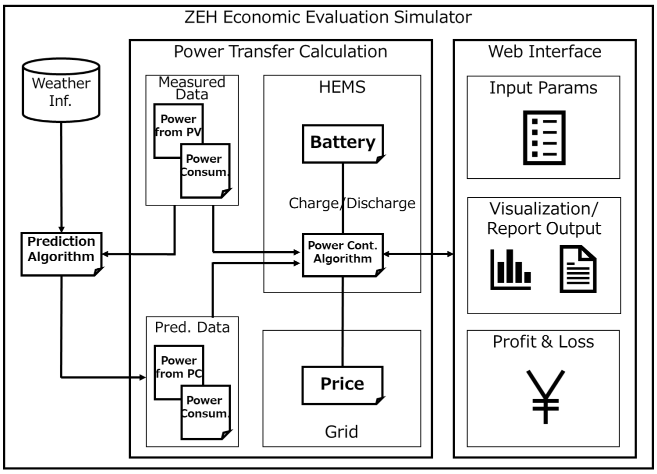


Figure 2: ZEH Economic Evaluation Simulator System Structure

3.3 Solar power generation data

An energy management experiment system for distributed energy is being conducted at the Hakusanroku Campus of the Kanazawa Institute of Technology, and energy-related data is being collected as part of the experiment [11]. Measured solar-generated electricity in 2020 is used for the power transfer simulation. Since the data is measured at 2-second intervals, we calculated the amount of electricity by using the average value of electricity generated every 30 minutes.

3.4 Electricity consumption data

The electricity consumption of eight detached houses in the Hokuriku region (Niigata Prefecture) is used to create electricity consumption data from the energy consumption data published by the Architectural Institute of Japan [12]. The electricity consumption of the eight houses was averaged and aggregated by time and classified into three categories. Using the consumption of eight houses as the basic data used in the simulation may lead to a deterioration in accuracy. In the future, if it becomes possible to obtain or measure more data, the accuracy of the simulation can be expected.

Electricity consumption and housing assignments are shown in Table 1.

Table 1: Three distinguished power consumption models

Model	House ID
Night/Midnight: Concentrated use at night or late at night when the user is away during the day, late-night hot water supply	1,2,6
Evening: Always at home and often used in the evening	4,5
Morning/Evening: Always at home and often used in the morning and evening	3,7,8

The classified data are averaged and aggregated by week, day of the week, and time, and scaled with the maximum value set to 1, and used in the simulation. The simulation reproduces electricity consumption according to households by adjusting the size according to a parameter called the average monthly electricity consumption rate. The electricity charges follow the metered B electricity rates of Hokuriku Electric Power Company.

3.5 Power transfer simulation

Figure 3 shows an overview of the simulation algorithm. First, a simulator using real data is developed, and later a simulation that takes the electricity supply-demand forecast into account and a simulation in which storage batteries are replaced by EVs are implemented.

4 SETTING UP A ZEH SIMULATION

4.1 Electricity consumption

Each power consumption model is used in the simulation with the power consumption sized as an average monthly power consumption charge of 10,000 yen. For the adjusted power consumption, Figure 4 shows the monthly power consumption and Figure 5 shows the average daily power consumption.

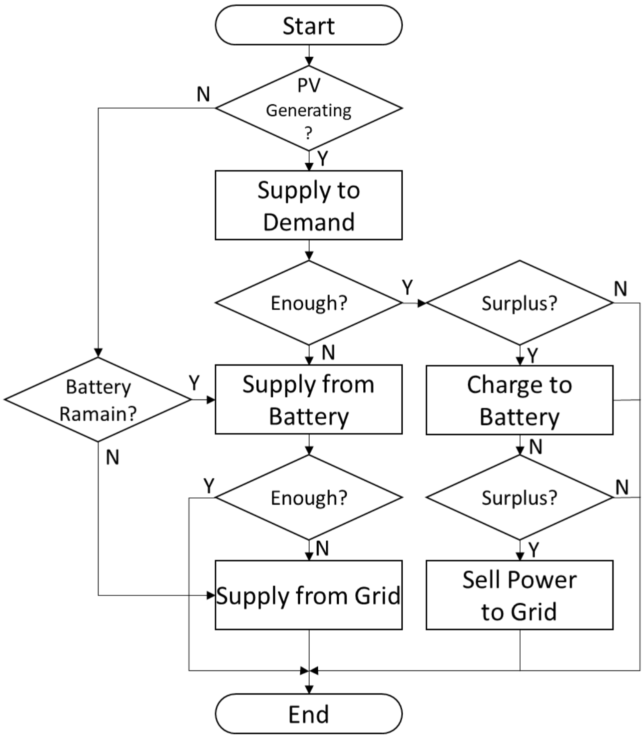


Figure 3: ZEH Economic Evaluation Simulator Algorithm Outline

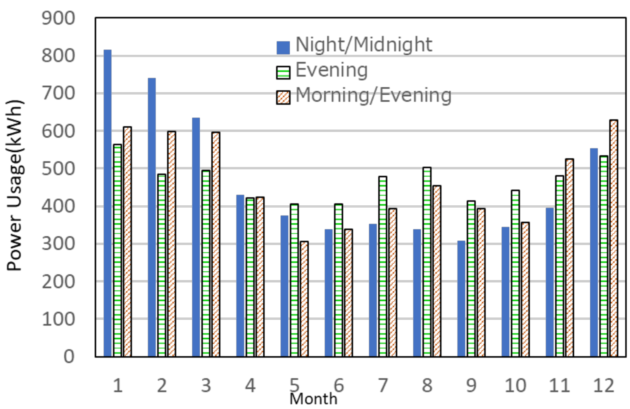


Figure 4: Monthly total power consumption amount

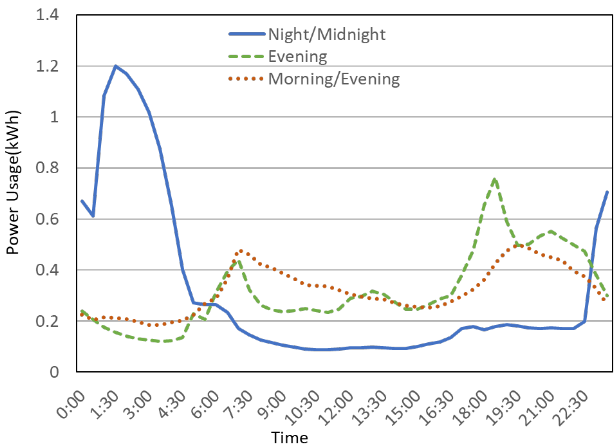


Figure 5: Average daily power consumption

Annual electricity consumption was 5621 kWh. The annual electricity consumption per household in Hokuriku is 6333 kWh, which is about 11% lower than the annual consumption per household in the same area. In all cases, electricity consumption is high in winter and low in spring and fall.

4.2 Setting up PV power generation and energy storage systems

We simulate combinations of PV outputs of 5, 10, 15, and 20 kW and storage capacities of 0, 5, 10, and 40 kWh, respectively. The storage capacity of 40 kWh is based on the case where an EV is used as a storage battery. Degradation of the power generation and storage system is not considered in this paper. Table 2 shows the settings of the installed system.

Table 2: System parameters

Parameter	Value
Power conversion efficiency	95%
Maximum charge/discharge efficiency	95%
Maximum charge and discharge power	3kW

4.3 System pricing assumptions

The installation cost of the PV system is assumed to be 336,000 yen/kW, the main unit cost of the stationary storage battery 140,000 yen/kWh, and the main unit cost of the V2H system 493,000 yen [13]. The cost of the EV itself is not included. The construction cost of the energy storage system is assumed to be 336,000 yen per installation. Table 3 shows the unit installation costs of stationary storage batteries, PV power generation, and energy storage systems, and Table 4 shows the system installation costs in this paper.

Table 3: System costs per unit

System	Cost
solar power (generation)	336,000 yen/kW
Stationary storage battery	140,000 yen/kWh
V2H(EV)	493,000 yen/unit
Energy storage system installation cost	336,000 yen per case

Table 4: System installed costs in this paper

PV (generation)		Battery system	
Spec.	Price(M Yen)	Spec.	Price(M Yen)
5kW	1. 68	0 kWh	0
10kW	3.36	5 kWh	1.036
15 kW	5.04	10kWh	1.736
20kW	6.72	40kWh (EV)	0.829

The cost of using an EV for the energy storage system is 493,000 yen for V2H plus 336,000 yen for construction, regardless of the capacity.

Since electricity consumption tends to be low during the day and increases from evening to night, we adapted Hokuriku Electric Power Company's Kutsurogi-night 12[14] as the electricity rate plan for our simulation. Kutsurogi Night 12 is a rate plan with low electricity prices during the nighttime (20:00~8:00). This report considers only the basic charge (1,650 yen/month) and the electricity volume charge, and does not include the renewable energy surcharge and the unit price of fuel adjustment. The unit price of electricity sold is assumed to be 8 yen.

5 ZEH ECONOMIC EVALUATION SIMULATION RESULTS

5.1 Profits and losses on electricity sales

The difference between the simulated profit/loss on electricity sales and the conventional electricity rate (120,000 yen) was calculated and used as the profit/loss for one year. Basic charges and system installation costs are not included.

Figure 6 shows the one-year power trading profits and losses for the evening model.

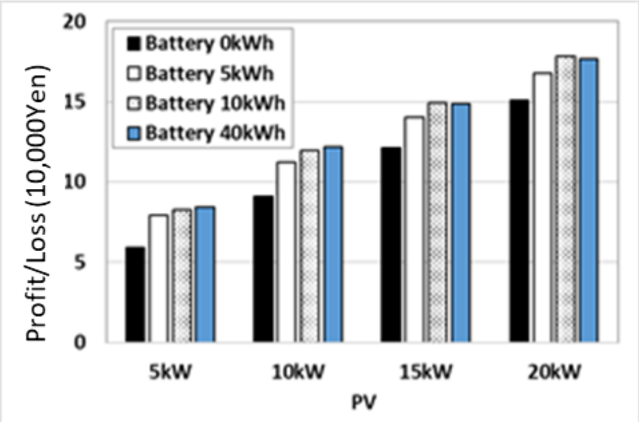


Figure 6: Profits and losses from electric power trading (Evening model)

The maximum profit was 178,000 yen in the model with 20 kW of PV power generation, 10 kWh of storage capacity, and consumption in the evening. Although profit increases with increasing PV output, it cannot be said that profit increases with increasing storage capacity.

5.2 15-year profit/loss including initial cost

A minimum loss of 26,000 yen was recorded for the concentrated nighttime model with only 5 kW of PV power generation, and a maximum loss of 6,365,000 yen was recorded for the concentrated morning and evening model with 20 kW of PV power generation and 10 kWh of storage batteries. No profit was obtained, including the initial cost.

The 15-year gains and losses for the nighttime concentration model are shown in Figure 7.

Under the conditions of this paper, it would be more economical not to install the system. If the unit cost of installing a PV power generation system is 330,000 yen, the Night/Midnight model with 5 kW of PV power generation can reduce electricity rates by 0.4 million yen from the conventional electricity rate plus the basic rate.

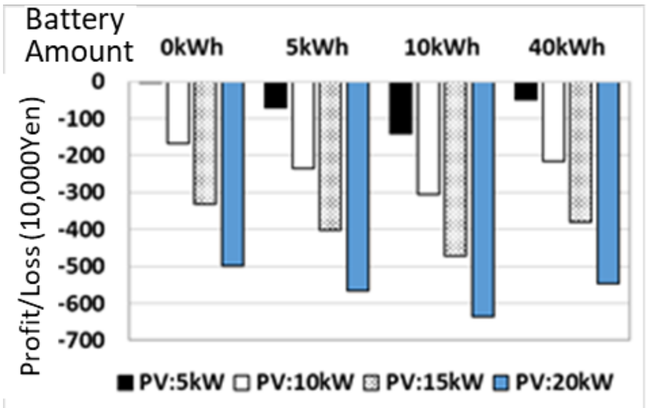


Figure 7: 15-year profits and losses (Night/Midnight model)

6 ELECTRICITY DISTRIBUTION TO MULTIPLE HOUSEHOLDS

6.1 Multi-Households Model

In this section, we assume a situation such as a community with several households equipped with PV power generation equipment and energy storage systems in the neighborhood by using electricity exchange among them to identify how electricity sharing resolve the electricity shortage. The multi-households model can also be used to simulate diverse and realistic electricity transactions by changing the electricity consumption patterns and the scale of the PV power generation and storage systems for each household. Figure 8 shows the electricity exchange model for multi-households' electricity exchange transactions.

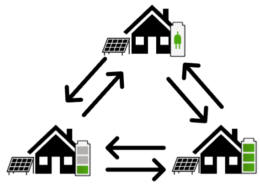


Figure 8: Multi-Households Electricity Sharing Model

6.2 Overview of Multi-Households Electricity Sharing Simulator

We simulate the flexible distribution of electricity among multiple households by running multiple simulators simultaneously and having each household disclose its electricity status and trade electricity through communication. In addition, we consider the power sharing network among multiple households as a single community and evaluate large-scale power sharing among communities with different power consumption patterns and power generation patterns.

In addition to the functions of the ZEH economic evaluation simulator, the following functions in Table 5 are implemented to simulate electricity sharing among multiple households. Details are explained in Section 6.4.

Table 5: Additional Function

implementation function	
Household-to-household communication capability	Reproduction of power trading
Algorithm for determining the amount of electricity during power sharing	Process for determining content at the time of transaction

6.3 Electricity calculation method for electricity Sharing

When conducting electricity sharing among multiple households, the amount of electricity that each household wishes to trade is determined, and simulations are conducted in a form that more closely resembles realistic transactions. As we explain in Section 3.4, the energy consumption used in the simulation is classified into three patterns. These power patterns were mixed in the community, and the simulator is operated. In addition, a minimum power configuration that assumes facilities that do not consume power but only generate, store, and supply power is also included.

6.4 Multi-Households Electricity Sharing Algorithm

Figure 9 shows the algorithm outline of the multi-households' electricity sharing algorithm. The left process shows the requester and the right one is the responder.

Households whose remaining battery capacity has been reduced below a certain level due to electricity consumption are approached by other households to sell their electricity. The households with sufficient remaining battery capacity present their tradable electricity to the electricity-deficient households. Prospective buyers adopt the optimal power trading details based on the offers from the responding households and charge their batteries until a certain capacity is met or the trading partner ceases to exist. If the battery cannot be recharged, electricity is purchased from the power company to meet the remaining demand.

7 SIMULATION OF ELECTRICITY DISTRIBUTION AMONG MULTI-HOUSEHOLDS

7.1 A Model Case for Simulation

We set up a household model case for one-to-one electricity sharing among households in a community and conducted simulations according to this setting.

Table 6 shows the electricity consumption plans and household equipment configurations used in the simulations. The reason why we choose Morning/Evening and Night/Midnight households is that their average daily power consumption is complementary.

Table 6: A Model Case Parameters

Case 1	Household-1	Household-2
Electricity Consumption Plan	Morning/Evening	Night/Midnight
Storage battery capacity (kWh)	5	5
PV power generation equipment (kW)	5	5

In this case, we simulate the storage battery capacity and PV generation capacity settings that could realistically be installed by an average household at this moment.

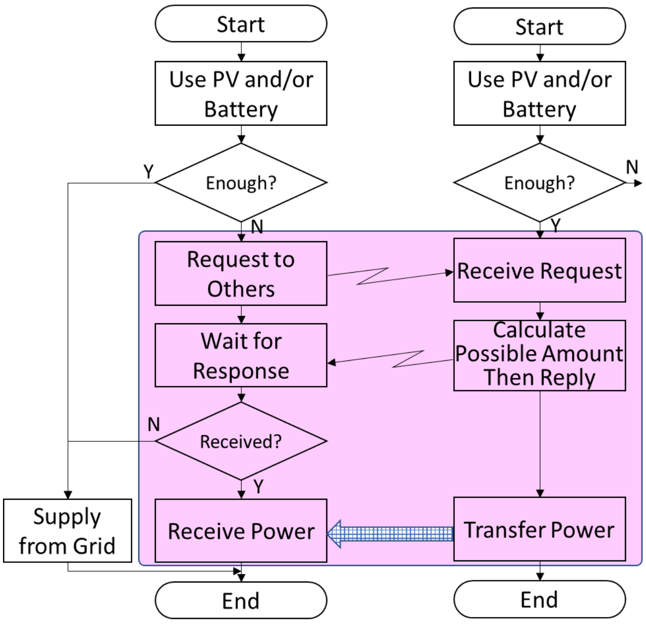


Figure 9: Multi-Households Power Sharing Simulator Algorithm Outline

7.2 Multi-households' simulation results and evaluation

The simulator is operated from July 6 to 13, when the supply to other households by PV power generation could be verified, to confirm that sharing is achieved. Figure 10 shows the electricity demand-supply ratio for Household-1.

On days when PV power generation is high during the day, electricity is supplied from the storage battery in Household 2, indicating that peak-time electricity demand could be met through sharing. However, on many days, the grid power supply is still relied on.

Table 7 shows the amount of electricity purchased and the consumption rate of PV power generation compared to the case without electricity sharing.

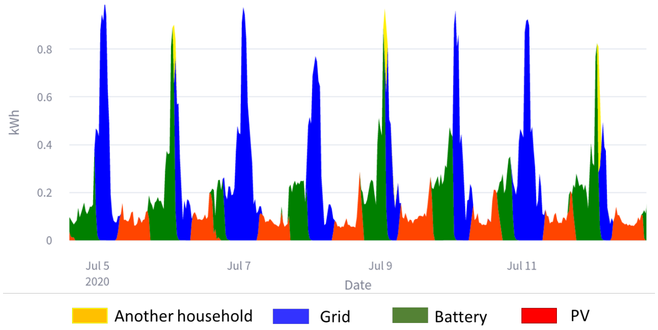


Figure 10: Household-1 Summer Electricity Supply Ratio

Table 7: Electricity Purchased and PV power Consumption Rate for Household 1

	One household	Two households Sharing
PV power consumption rate (%)	54.8	52.0
purchased from the grid amount (kWh)	46.3	50.0

The results show that the consumption rate of PV power consumption decrease, and the amount of electricity purchased from the grid increases because of the electricity sharing. These results indicate that even small-scale facilities can consume a portion of the demand through electricity sharing during the season when PV power generation can be expected.

8 SIMULATION OF MULTI-COMMUNITY ELECTRIC POWER SHARING

8.1 Overview of Electricity Transactions between Communities

We assume that there are communities with multiple households in the vicinity to trade electricity. Since the electricity storage pattern by PV power generation differs from community to community (Figure 11), a stable electricity supply in winter, which is difficult to achieve in a simulation of only one community, can be expected.

8.2 Smart Marginalized Communities

In this study, we define "Smart Marginalized Community" as a community in a smart grid system that achieves net-zero energy in mountainous areas and does not rely on external power supply.

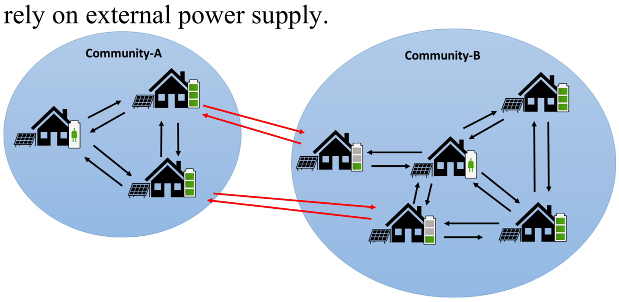


Figure 11: Model of inter-community electricity sharing

In this simulation, we are working to achieve a form of energy self-sufficiency by producing electricity and heat for their own consumption, called net-zero energy. This study aims to "achieve complete net-zero energy by means of renewable energy (mostly PV energy) and energy storage systems that do not rely on external power," and hypothesizes that it is possible to achieve these goals and realize smart marginal community. We then verify the configuration of the smallest marginal community that can satisfy the conditions and the combination of power generation and storage systems.

The optimal configuration and changes in the degree of net-zero energy realization is also examined when considering the realization of net-zero energy in regions with different demand and generation conditions.

8.3 Model case in two Communities

A model case is then set up for two neighboring communities to share electricity with each other, and simulations are conducted according to this set-up. Table 8 and 9 shows the electricity consumption plans and household facility configurations used in the simulations.

Table 8: Community-1 Parameters

Community-1			
Case 3	Household-1	Household-2	Equipment-1
Electricity Consumption Plan	Night/Midnight	Morning / Evening	a little
Battery capacity (kWh)	5	5	15
Solar power generation (kW)	5	5	15

Table 9: Community-2 Parameters

Community-2			
Case 3	Household-3	Household-4	Equipment-2
Electricity Consumption Plan	Night/Midnight	Morning / Evening	a little
Battery capacity (kWh)	5	5	15
PV generation (kW)	5	5	15

8.4 Multi-Community simulation results and evaluation

The simulator is operated for one year. Figure 12 and Figure 13 show the percentage of electricity demand supplied by Household-1 and the amount of electricity supplied by Facility 1 during the period from July 1 to July 31.

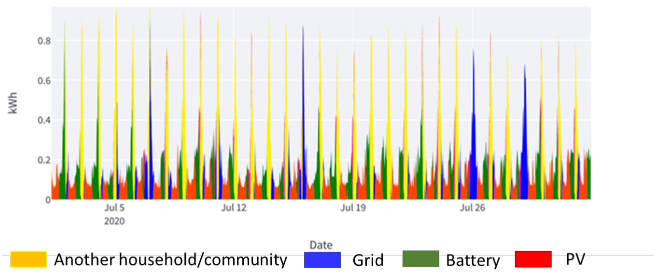


Figure 12: Household1 Summer Electricity Supply Ratio

The results in Figure 13 show that household 1 receives a large amount of electricity from Facility 1 and receives electricity from Facility 2. Table 10 shows the amount of electricity purchased and the consumption rate of renewable energy by Household 1 compared to the case without electricity supply.

The consumption rate of renewable energy is significantly increased, and the amount of electricity purchased from the grid is also reduced through the sharing power supply system. In addition to the intra-community transactions, the power acquired through transactions with power generation facilities in other communities is considered to have enabled more stable operation that is not dependent on grid power.

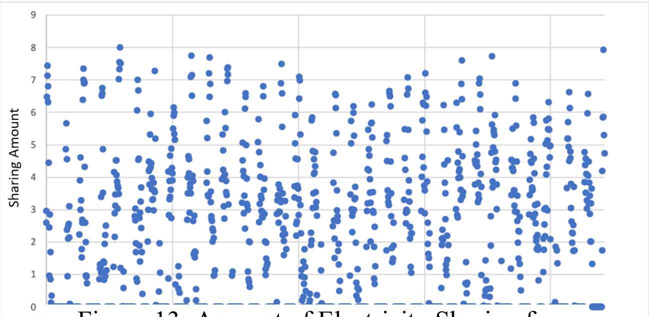


Figure 13: Amount of Electricity Sharing for Equipment-1

Table 10: Electricity Purchased and Renewable Energy Consumption Rate for Household 1

	One household	Inter-community Sharing
PV power consumption rate (%)	37.3	54.0
purchased from the grid amount (kWh)	3523	2484

9 CONCLUSION AND FUTURE WORKS

In this study, we have developed a simulator to evaluate the economic efficiency of ZEH. The simulator is useful for each household to consider the installation of equipment. The simulator can be operated through a web interface, so it can be operated by many people.

If economics is the only consideration when installing PV power generation and storage systems, it is best to refrain from installing PV power generation and storage systems.

In a one-on-one simulation assuming intra-community transactions, the amount of electricity purchased decrease, and the PV power rate increase when the storage batteries and capacity of one household are increased, while the demand during peak hours is only partially compensated by the sharing of the other household when the size of the storage batteries and PV generation facilities of both households are small.

In the case of transactions in which the scale of transactions is expanded to multi-communities, by scattering large facility households with different electricity consumption styles, the surplus electricity of large households was passed on to small facility households, and as a result, a high renewable energy rate is confirmed.

However, in this study we only simulate very simple use cases, so that the results are limited. We will set up some more use cases and evaluate and analyze towards ZEH.

Another challenge for the future is to change the electricity transaction price, which is fixed within and between communities, using an algorithm.

REFERENCES

[1] Agency for Natural Resources and Energy, "Actions toward the end of the FIT for residential photovoltaic power generation equipment," (2019).

[2] Agency for Natural Resources and Energy, Information Disclosure on ZEH (Net Zero Energy House) - Energy-Saving Housing, https://www.enecho.meti.go.jp/ category/ saving_and_new/saving/general/housing/ index03.html, (accessed July 26, 2022).

[3] New Energy and Industrial Technology Development Organization, "NEDO Renewable Energy Technology White Paper, 2nd Edition," (2014).

[4] Agency for Natural Resources and Energy, "Recent Trends in Electricity Supply and Demand and Wholesale Electricity Market," https://www.meti.go.jp/ shingikai/enecho/denryoku_gas/denryoku_gas/pdf/046 _03_02.pdf, (accessed July 15, 2022).

[5] Agency for Natural Resources and Energy, "Actions Toward the Formation of Next-Generation Distributed Energy Systems," https://www.meti.go.jp/shingikai/ enecho/denryoku_gas/denryoku_gas/pdf/046_04_02.pd f, (accessed July 15, 2022).

[6] Agency for Natural Resources and Energy, "A Guide for Establishing Regional Microgrids," https://www.meti.go.jp/shingikai/energy_environment/ energy_resource/pdf/015_s01_00.pdf, (accessed July 20, 2022).

[7] Ren Ishikawa and Hironobu Matsuo, "Evaluation of Energy and Economy of Battery System Installation in residential PV System after FIT," Journal of Japan Solar Energy Society, Vol.46, No.5, pp.63-70 (2020)

[8] Kuniaki Yabe and Yasuhiro Hayashi, "Economic Evaluation of Increased In-House Consumption of Photovoltaic Power Output by Storage Batteries," IEEJ Transactions B (Journal of the Electricity and Energy Division), Vol. 139, No. 5, pp. 363-371 (2019).

[9] Ryota Goto, Hirotaka Takano, and Hiroshi Asano, " Scheduling Method of Microgrids' Operation Considering Uncertainty in Values of Forecasted Net Load," Transactions of the Japan Society of Energy and Resources, Vol. 42, No. 1, pp. 11-17 (2021).

[10] Ryoichi Komiyama, Saeko Shibata, and Yasumasa Fujii, "Analysis on Optimal Power Supply Structure Considering Output Fluctuation of Photovoltaic and Wind Power Generation and Interregional Electricity Flexibility," IEEJ Transactions on Electric Power Industry B Vol. 133, No. 3, pp. 263-270 (2013).

[11] Eiji Fukuda et al, "Construction of Energy IoT Data Collection System (2)", Information Processing 82nd National Convention, (2020).

[12] Committee for Research and Study on Energy Consumption in Houses, Architectural Institute of Japan, Database of Energy Consumption in Houses, http://tkkankyo.eng.niigata-u.ac.jp/HP/ HP/database/index.htm, (accessed July 1, 2021).

[13] Nichicon Corporation, Product Specifications/EV Power Station/Grid-connected V2H System, https://www.nichicon.co.jp/products/v2h/about/specs/, (accessed Jan. 10, 2022).

[14] Hokuriku Electric Power Company, Hokuriku Electric Power Company Customers at Home (unit price of electricity), https://www.rikuden.co.jp/ ryokin/minsei.html, (accessed Dec. 3, 2021).

Development of a Disaster Safety Confirmation Collection System using Q-ANPI Service linked with Web Interface

MISTUKI SANO [*], TAKUMA SATO [*], MASAO ISSHIKI, and KEIICHI ABE [*],
[*] Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
abelab.sano.0914@gmail.com, abe@he.kanagawa-it.ac.jp

Abstract -In the future, when a wide-area disaster occurs in Japan, there is a possibility of the loss of electric power and communication infrastructure. In such a case, Q-ANPI, a service for collecting information on the safety of disaster victims using the Quasi-Zenith Satellite MICHIBIKI, which can cover the entire country, would be very effective. However, Q-ANPI currently has the problem that if the communication infrastructure is cut off due to a large-scale disaster, the safety information cannot be registered on the Q-ANPI shelter management PC from a smartphone owned by a disaster victim if the dedicated application is not installed on the smartphone in advance. To solve this problem, this study proposes a method for inputting the safety information of disaster victims using a Web browser already installed on their smartphones. In addition, a prototype of the proposed system was evaluated from three different perspectives: a simulated evacuation drill, a questionnaire survey, and power measurements.

Keywords: QZSS-MICHIBIKI, Victims information, Q-ANPI service, RMS.

1 INTRODUCTION

In the future, when a wide-area disaster such as the Great East Japan Earthquake occurs, there is a possibility of the loss of electric power and communication infrastructure. In fact, when the Great East Japan Earthquake occurred, ICT (Information and Communication Technology) could not be fully utilized, making it impossible to promptly disseminate information on disaster victims, etc. [1]. In the Kumamoto earthquake, it is reported that the number of victims in evacuation centers was managed at night, and the name list itself did not function because of the heavy comings and goings in the evacuation centers [2,3]. To solve these problems, we have developed a Refuge Management System (hereinafter referred to as RMS) [4]. This RMS [4] utilizes ICT to collect information on disaster victims in each shelter and to manage the long-term health of each victim, and can transmit the collected victim information via simple radio communication, such as amateur radio, as well as other means. The information collected by this RMS can also be stored in a USB flash memory and take it out manually. However, assuming a scenario in which a wide-area disaster such as a major NANKAI Trough earthquake occurs in Japan in the future, ”Q-ANPI”[5], a service for collecting safety information on disaster victims using the Quasi-Zenith Satellite "MICHIBIKI", which can cover the entire area of Japan, is considered to be very effective. There are several services provided by this Quasi-Zenith Satellite "MICHIBIKI", and Q-ANPI is one of them . This study focuses on Q-ANPI. Currently, Q-ANPI[5] provides two input interfaces for collecting disaster victim information. One is to download a smartphone application for registering safety information of disaster victims. The other is to register information directly from the Q-ANPI refuge management PC using a keyboard. However, the ”Q-ANPI" currently has the problem that safety information cannot be registered with the refuge management PC for ”Q-ANPI" from a smartphone owned by a disaster victim unless the dedicated application is installed on the smartphone in advance, in the event that the smartphone communication infrastructure is cut off due to a large-scale disaster. For this reason, people will also be allowed to register their safety information directly from the Q-ANPI refuge management PC (Personal Computer) using a keyboard. In the midst of a large-scale disaster, a great deal of congestion will probably occur if a large number of people register directly on the Q-ANPI refuge management PC. Therefore, the authors believe that the current Q-ANPI system cannot fully utilize the advantages of ICT. This paper describes the proposed technology to solve this problem. In addition , we will also propose a method for operating refuge management PCs with low power consumption even in the loss of power infrastructure disruption. We also describe the development of the prototype system proposed in this paper and the evaluation of the effectiveness of the system from various perspectives, including an actual simulated evacuation drill using the prototype system and the results of a demonstration questionnaire evaluation. In Chapter 2 ,we describe the position of this study with respect to related studies, and Chapter 3 gives an overview of the proposed system and the system flow. Chapter 4 evaluates the proposed system, and Chapter 5 presents the conclusion.

2 RELATED WORKS

A previous study of disaster victim information dissemination in response to a wide-area disaster is the RMS development [4]. This RMS utilizes ICT to collect information on disaster victims in each refuge and to manage the long-term health of each victim and can transmit the collected disaster victim information. The RMS can transmit information by simple radio communication and amateur radio, in addition to storing disaster victim information on USB flash memory devices and manually taking it outside. However, assuming a wide-area disaster such as the Nankai Trough earthquake that is expected in the future, simple radios, etc., would not be able to communicate over a wide area because their communication range is too short. For this reason, the authors believe that it would be effective to utilize a satellite communication system using the Quasi-Zenith Satellite “MICHIBIKI”, which can cover the whole area of Japan, as promoted by the National Space Policy Secretariat, Cabinet Office, Government of Japan. Therefore, this study focuses on the Q-ANPI service. In a previous study on the Q-ANPI service, a system[6] is proposed that can collect safety information on disaster victims more widely than the conventional Q-ANPI system by combining Q-ANPI and ad hoc communication using smartphones. This proposed system[6] is very useful for solving the problem of not being able to connect to a shelter management PC when the connection to the Wi-Fi access point is not known, and for collecting the safety information of victims outside of the refuge. but this system requires the dedicated application to be downloaded in advance.

On the other hand, the current Q-ANPI[5] provides two input interfaces for collecting information on disaster victims. One is to download a dedicated smartphone application and register the safety information of disaster victims through the application, and the other is to input and register the safety information directly from the Q-ANPI refuge management PC using a keyboard. However, in the event that the smartphone communication infrastructure is cut off due to a large-scale disaster, the safety information cannot be registered with the refuge management PC for ”Q-ANPI" from a smartphone owned by a disaster victim unless the dedicated application is installed on the smartphone in advance. For this reason, the current system additionally allows people to register their safety information directly from the refuge management PC using a keyboard. Since it is assumed that one such PC will be placed in each refuge, a great deal of congestion will probably occur if there are a large number of people registering directly on the refuge management PC. Therefore, the authors believe that the current system cannot fully utilize the advantages of using a smartphone.

Therefore, To solve this problem, the authors propose a method for inputting the safety information of disaster victims using a Web browser already installed on smartphones.

3 PROPOSED SYSTEM

In this study, we decided to use Q-ANPI, which uses the Quasi-Zenith Satellite “MICHIBIKI” and can cover the entire area of Japan, as a solution to the problem of the loss of communication infrastructure in the event of a large-scale disaster. However, as described in Chapter 3, it is necessary to consider a scenario in which the communication infrastructure is cut off in the event of a large-scale disaster, and the dedicated application for inputting safety confirmation for Q-ANPI cannot be downloaded, in which case the advantages of Q-ANPI cannot be fully utilized. To solve this problem, this study proposes a method for inputting the safety information of disaster victims using a Web browser already installed on smartphones. In this study, we also considered that the power infrastructure will be cut off at the same time during a large-scale disaster, and aimed to realize a practical safety information collection system that is robust against power infrastructure cutoffs by linking Q-ANPI with a system built using an embedded microcomputer board that can operate with low power consumption, instead of building a Web server and DB (Data Base) in the refuge management PCs.

3.1 Overview of proposed system

Figure 1 shows an overview of the Q-ANPI safety information collection system linked with the web interface proposed by this research. The web interface proposed by this research uses a small, low-power-consumption microcontroller (Raspberry PI 3B+; hereinafter referred to as RPI) with a web server and DB (Data Base). The RPI is used as an intermediary between smartphones owned by disaster victims and the refuge management PC. When a smartphone owned by a disaster victim connects to the web interface from its web browser, a page for inputting the victim's safety information is displayed. When the safety information is entered, it is automatically registered in the DB in the web interface. This collects the safety information of multiple victims in a refuge. The system then sends the information to the refuge management PC in a batch. When the person in charge of the refuge management presses the data transmission button on the application for the refuge management PC, the data are transmitted to the server of the control station via the Quasi-Zenith Satellite “MICHIBIKI”. The system enables relatives of disaster victims to confirm the safety of the victims by accessing the server of the control station via the Internet. By introducing the Web interface proposed in this research, the application for the refuge management PC used in the current Q-ANPI can be installed without any modification. In addition, anyone who has a smartphone with a Web browser installed can input victim information without having to download the dedicated application for inputting safety information. To reduce the power consumption of the proposed Web interface, we used RPI that can operate with lower power consumption than PC. The power supply used for the entire system is a 240Wh battery equipped with an inverter for AC power output as well as a commercially available photovoltaic panel, assuming that the power infrastructure is cut off.

3.2 How to display the safety information input page from a smartphone

As with Q-ANPI, the system in this study forms a LAN (Local Area Network) with commercially available Wi-Fi wireless routers. In addition to receiving safety information from smartphones and the Web interface, the refuge management PC serves as a gateway to the Q-ANPI terminal (920 MHz). For the connection to the web interface from a smartphone, the SSID of the Wi-Fi wireless router (in this case, "QANPI_179") was made open to allow connection to the network without a password. A printout of the URL of the safety information input page converted into a QR code was prepared in advance. By holding the QR code over the camera of a smartphone, anyone can easily display the safety information input page on the Web interface.

3.3 Prototype system of web interface

Figure 2 shows an overview of the prototype of the web interface proposed in this paper. The web server was built on RPI using nginx, and MariaDB was used for the DB. The reason for using nginx for the web server is to achieve high processing speeds that allow multiple inputs of victim information at the same time. The application for the victim safety information input interface shown in Figure 2 was developed in PHP and HTML languages. The "disaster victim information transmission application," which transmits the safety information of disaster victims registered in the DB server to the refuge management PC, was developed in python 3.7, and is an application that transmits all safety information registered in the DB. Since the application for the refuge management PC automatically discards duplicate information, we decided to use a simple application that sends all DB data at once. Figure 3 shows the appearance of the prototype developed this time. This is a prototype with a web server and DB (Data Base) built on RPI.

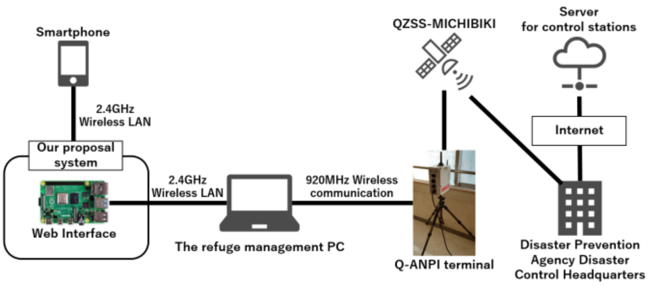


Figure 1: Overview of Proposed System.

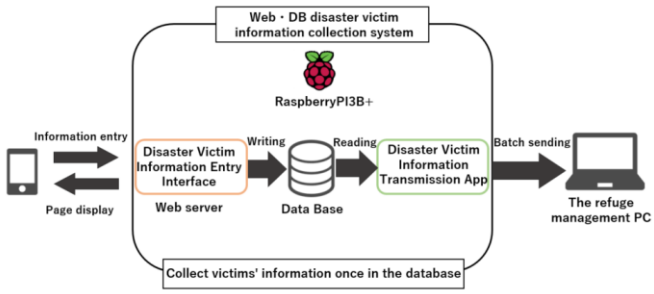


Figure 2: Overview of Web Interface.

Figure 3: Prototype of Web Interface.

3.4 How to operate the power supply of the proposed system

Assuming that power and communication infrastructures are cut off in a large-scale disaster, power supply is important in a real refuge. To operate the proposed system, it is considered necessary to use one solar panel and two or more battery units. One battery unit is used for the refuge management PC, web interface, wireless router, Q-ANPI terminal, etc. The other battery unit is charged with electricity generated by the solar panel, and the batteries should be exchanged alternately. The Q-ANPI terminal and the refuge management PC consume a lot of power, so when they are not transmitting data through the Quasi-Zenith Satellite “MICHIBIKI”, the Q-ANPI terminal can be turned off and the refuge management PC can be put in sleep mode for long-term operation.

4 DEMONSTRATION EVALUATION

4.1 Demonstration evaluation of the system

We evaluated the effectiveness of the prototype of the system proposed in this study from three perspectives: a demonstration evaluation based on a simulated evacuation drill, a questionnaire evaluation based on a demonstration, and a measurement evaluation of the prototype's power consumption. In the evaluation through a simulated evacuation drill, we conducted a questionnaire to determine whether the proposed system is practically usable, and compared it with the conventional Q-ANPI system to determine how easy it is to input safety information. In addition, a questionnaire survey was conducted by exhibiting a prototype demo to investigate the ease of use of the proposed system and find points to be improved. We also evaluated the power consumption of the prototype system using the Web interface proposed in this study.

4.2 Simulated evacuation drill

A prototype of the proposed system was utilized and evaluated through a simulated evacuation drill. The simulated evacuation drill was conducted on August 5, 2022, on the premises of the Kanagawa Institute of Technology. Figure 4 shows the evacuation route of the simulated drill. Table 1 shows the schedule of the actual evacuation drill and the actual action time required. Before the simulated evacuation drill, one Q-ANPI terminal, one refuge management PC, one web interface system, one solar panel, and one battery with 240Wh capacity were placed in the former gymnasium of the university, which served as the evacuation site. It was assumed that a major earthquake occurred at 14:00. All participants in the evacuation drill initially hid under desks, and after the earthquake subsided, they descended the stairs from the classroom (E602) on the 6th floor of Building C2 to the first floor, and then followed the evacuation route shown in Figure 4 to the entrance of the university's soccer field. After taking roll call at the evacuation assembly, all the evacuees were moved to the old

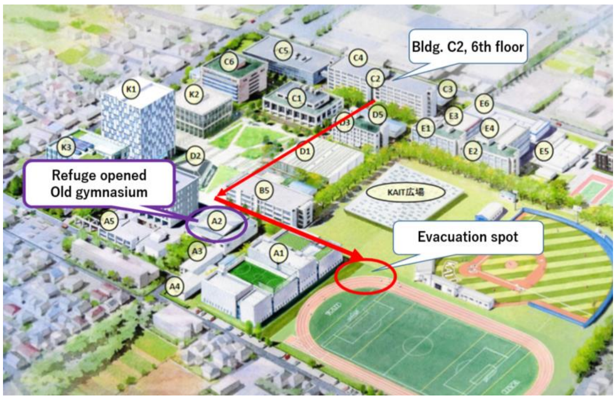


Figure 4: Evacuation route for simulated evacuation drill.

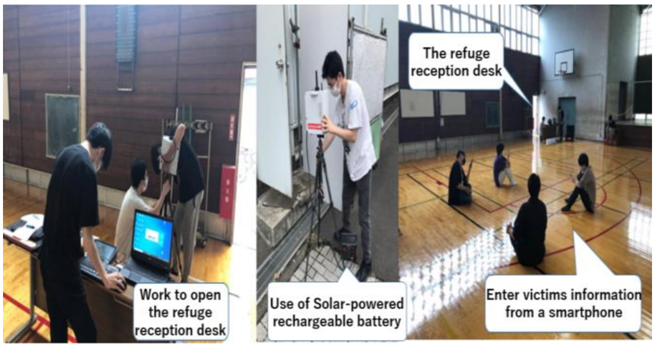


Figure 5: Overview of simulated evacuation drill.

Table 1: Schedule for the Demonstration Experiment.

Scheduled time	Action	Actual times required	Remarks
14:00	Earthquake-Hide under the desk	14:00	
14:15	Evacuation to a refuge	14:06	Evacuation by stairs from the 6th floor of Building C2
14:20	Gather at evacuation site	14:09	
14:25	Move to the refuge (old gymnasium)	14:09	
14:40	Establishment of the refuge		
	Completion of evacuation center opening	14:17	
14:55	Enter victim information (Install dedicated app)	14:17Start⇒14:28Finish (11minutes)	Mistakes in downloading the dedicated app and connecting to Wi-Fi required time to complete the input.
	Enter victim information by the proposed system	14:34Start⇒14:37Finish (3minutes)	Fastest input time
15:00	Fill out the questionnaire	14:35Start⇒14:37Finish	
15:15	The withdrawal	14:41	

gymnasium, which was the evacuation site, where three evacuation supervisors set up a proposed system and opened the evacuation desk (Figure 5). Table 1 records the time from immediately after the earthquake until the equipment was removed. The time required from the occurrence of a major earthquake to the gathering at the evacuation site was 9 minutes, which indicates that the evacuation was completed fairly quickly. The time required to open the evacuation site was about 8 minutes. In the schedule shown in Table 1, attention should be paid to the "process of inputting disaster victim information.” The proposed system required only about 3 minutes, while the conventional Q-ANPI system required about 11 minutes. The proposed system took only about a quarter of the time of the conventional system. Factors that caused the conventional Q-ANPI system to take such a long time include the time required to download the dedicated application and the time required to establish a Wi-Fi connection. The proposed system uses a QR code for easy access to the safety information input page, which is a major factor. Therefore, the proposed system is superior in terms of the time required for on-site use in actual disasters.

4.3 Questionnaire evaluation after simulated evacuation

In the demonstration evaluation through the simulated evacuation drill, a questionnaire survey was conducted after the simulated evacuation drill, since personal safety information was collected by two systems, i.e., the conventional Q-ANPI and the proposed system. The actual number of participants in the simulated evacuation drill was 9, and 6 males (67% response rate) responded to the questionnaire. Figure 6 shows the results of the questionnaire. To Question 1, "In the event of a large-scale disaster, would you like to actually use the proposed system used this time, which uses a browser for data input, in an evacuation center?", three respondents answered "Yes, I would like to use it," and three respondents answered "Yes, I would like to use it somewhat.” This resulted in a high average score of 3.5 on the 4-point scale in which the highest score of 4 was given to "would like to use" and the lowest score of 1 was given to "would not like to use". Next, to Question 2, "Which system was easier to use, the conventional system or the proposed system?”, 83% of the respondents answered that the proposed system was easier to use. Next, in response to Question 3, "In Question 2, what was easy to use about the conventional system?”, the answer was "Because there is no need to search in a browser.” In Question 4, "In Question 2, what was easy to use in the proposed system?”, five respondents answered "The fact that it can be used without downloading a dedicated application." To Question 5, "Is there any other information on disaster victims that you think is necessary besides the existing information (phone number, name, date of birth, gender, admission status, availability, address, injuries, assistance needed, disabilities, pregnant women)?”, One respondent answered, "Information such as pre-existing medical conditions, family doctor, and medications used on a regular basis would be helpful.” Next, to Question 6, "What do you think needs to be improved in the future regarding the proposed system? As shown in Figure 6, the most common response to this question was "Type of evacuee information" (five responses), followed by "Ease of text input" (two responses).

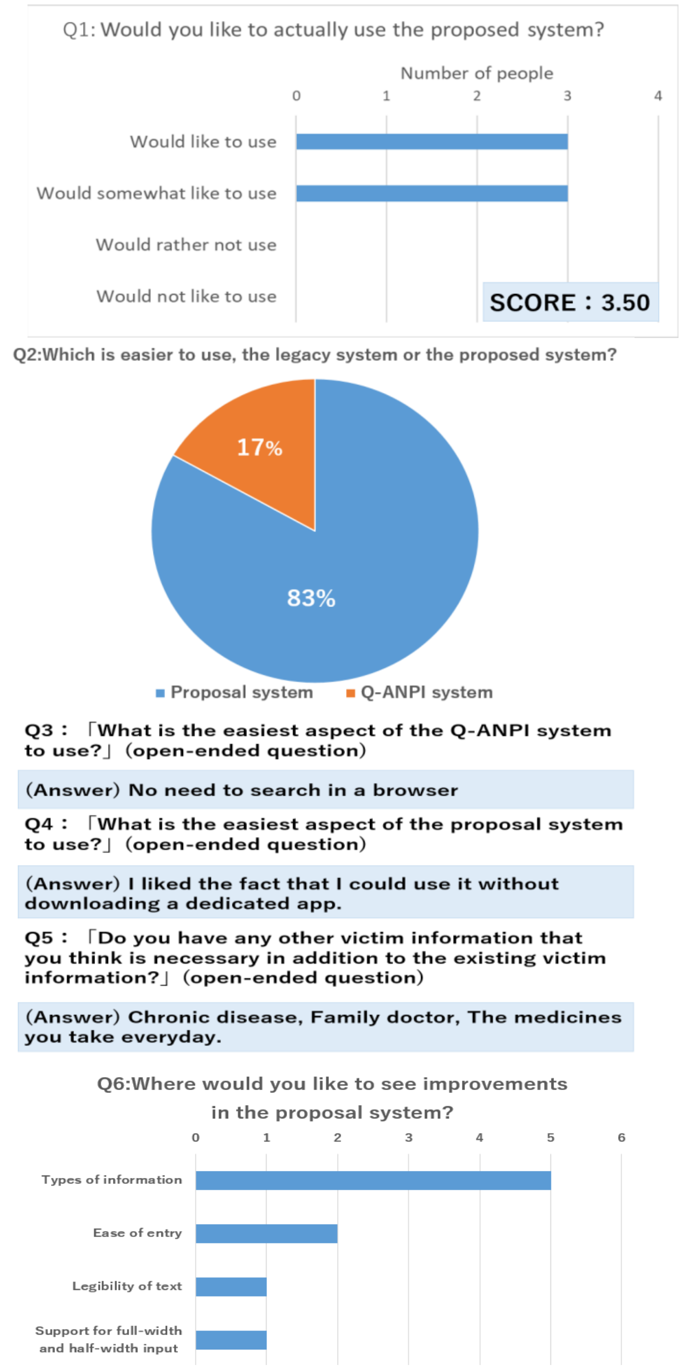


Figure 6: Questionnaire results for simulated evacuation drill.

4.4 Measurement and evaluation of prototype power consumption

Table 2 summarizes the results of power consumption measurements for each device used in the proposed prototype system. It can be seen that the Q-ANPI terminal consumes the largest amount of power (15W) among the devices used in the prototype system. The second highest is the refuge management PC, which has a maximum power consumption of 12 W. The third highest is the RPI, and the fourth highest is the wireless router. The RPI alone consumes approximately 3 W, which is a significantly low power consumption. Considering the use of a commercially available 240Wh battery, it is estimated that the RPI alone can operate for about 3.3 days. A more power-saving operation could be achieved by using the RPI as a stand-alone device without the LCD display screen, with the terminal software for the RPI installed on the refuge management PC so that the RPI's "disaster victim information transmission application" could be run remotely. It is also considered a good idea to turn on the power of the Q-ANPI terminal and the refuge center management PC only when communicating via the Quasi-Zenith Satellite “MICHIBIKI”, and turn them off and in sleep mode for the rest of the time. Assuming such an operation scenario, only the RPI unit (3W) and the wireless router (2.5W) would be used in constant operation, so the 240Wh battery would last approximately 1.8 days. If the two battery units are operated alternately, long-term operation is possible even when the power infrastructure is cut off.

Table 2: Power consumption of each device.

	Mode	Model/Manufacturer	Power consumption(W)
Q-ANPI terminal	–	–	15
The refuge management PC (Windows10)	Working	Versa Pro VX-5(NEC)	12(max)
	Sleeping		0.42(max)
Wireless router	–	WMR-433W2-BK (BUFFALO)	2.5(max)
Web interface(RPI)	–	Raspberry PI3B+	3

5 CONCLUSION

This paper addresses issues related to Q-ANPI, a safety confirmation collection service for disaster victims using the Quasi-Zenith Satellite "MICHIBIKI". It describes how to solve the problem that when the cell phone communication infrastructure is cut off, safety information cannot be registered from a smartphone owned by a disaster victim to the refuge management PC used by Q-ANPI unless a dedicated application is installed on the smartphone in advance. We proposed a method to solve this problem by using a web interface that enables the input of safety information of disaster victims using a web browser already installed in their smartphones. This web interface collects the safety information of multiple victims and sends it to the refuge management PC at any given time, which then sends the data to the control station server via the Quasi-Zenith Satellite “MICHIBIKI”. The system is a general interface that allows safety information to be input using a Web browser already installed on a smartphone. The evaluation of the prototype through simulated training showed that the time required to input safety information was shorter with the proposed system than with the conventional Q-ANPI. In a questionnaire survey conducted after the simulation training, many respondents answered that they would prefer to use the proposed system rather than the conventional Q-ANPI system, and many of them answered that the proposed system was “easier to use.” However, some respondents pointed out the following points for improvement of the proposed system: "Ease of text input" and "Types of victim information.” In the event of an actual large-scale disaster, it can be assumed that far more people will access the system than in this simulated evacuation drills. For this reason, we plan to conduct load tests with a larger number of people than in this simulated evacuation drills. The prototype of the proposed web interface was developed using RPI. Power consumption was measured, and it was found from this evaluation that the power consumption was low among the devices used. Therefore, for the sake of power conservation, the RPI can preferably be used alone, in which case the problem of the screen display can be solved by remotely controlling the RPI through terminal software installed on a refuge management PC.

In the future, if the proposed system and Q-ANPI's previous research [6] are all successfully linked, it will be possible to develop a highly realistic system that can smoothly collect safety information of disaster victims at actual disaster sites where people are in confusion.

ACKNOWLEDGEMENTS

We would like to express our sincere appreciation to the National Space Policy Secretariat, Cabinet Office, Government of Japan, and the Space Systems Division of NEC Corporation for their guidance and cooperation in providing the Q-ANPI service equipment.

REFERENCES

[1] Reconstruction Agency, Cabinet Office (in charge of disaster prevention), and Fire and Disaster Management Agency: "Number of deaths related to earthquake disaster in the Great East Japan Earthquake",<http://www.reconstruction.go.jp/topics/main-cat2/sub-cat26/20171226_kanrenshi.pdf>, (Referred:2023-06).

[2] Ai Urano, Executive Director, Rescue Stock Yard (RSY): "Actual conditions and issues of shelter management in the Kumamoto earthquake - comparison with shelter management in past earthquakes."<https://www.isad.or.jp/pdf/information_provision/information_provision/h28/H28_dai3bu1.pdf> , (Referred:2023-06).

[3] Ministry of Land, Infrastructure, Transport and Tourism, "Issues and Discussion Points in Light of the Kumamoto Earthquake.",<https://www.mlit. go.jp/co-mmon/001135911.pdf>, (Referred:2023-06).

[4] K. Akasaka, T. Namba, M. Isshiki, K. Abe, "Proposal on Victims Information Management System", International Journal of Informatics Society (IJIS), Vol.12, No.1, pp.17-27, Aug. (2020).

[5] Cabinet Office, Government of Japan, "Q-ANPI Satellite Safety Confirmation Service.", <https://qzss.go.jp/overview/services/sv09_qanpi.html>,(Referred:2023-06).

[6] N. Ohuchi, M. Nishiura, S. Aikawa, A. Yumura, T. Hamada, "Disaster Safety Confirmation Collection System Using QZSS Michibiki and Smartphones", The Institute of Electronics, Information and Communication Engineers, Technical Report, Vol.119, 266, pp.27-32, Nov. (2019).

River Water Level Prediction Using Radar Rainfall with Deep Learning

Futo Ueda†, Hiroto Tanouchi‡, Nobuyuki Egusa‡, Takuya Yoshihiro‡

†Graduate School of Systems Engineering, Wakayama University, Japan
‡Faculty of Systems Engineering, Wakayama University, Japan
{s246034, tanouchi, egusa, tac}@wakayama-u.ac.jp

Abstract - Accurate prediction of river water levels during floods is important to help residents make appropriate evacuation decisions and is the subject of much research. Traditional river water-level predictions are based on physical models. Recently, several studies reported that deep-learning-based models achieve more accurate water-level predictions. Existing deep learning-based water level prediction methods predict downstream water levels based solely on rainfall and water levels at upstream observatories. However, in many cases, there is not a sufficient number of upstream observatories to make accurate water-level predictions.

In this study, we propose a river level prediction model based on deep-learning models, CNN and LSTM, with radar rainfall data. We predict the river levels up to several hours after the current time with the proposed model that uses the historical water levels at the prediction place and the radar rainfall data, without using the data of upstream observatories. The evaluation results showed that the proposed model using the historical water level at the prediction place and the radar rainfall in the basin can predict water levels with a small error, even if it does not use the data of the upstream observatories.

Keywords: river water level prediction, deep learning, radar rainfall, convolutional neural network(CNN), long short-term memory(LSTM)

1 INTRODUCTION

Because Japan has a lot of mountains, there are many rapid-flowing rivers. Since river water-level in Japan rises rapidly in a short period of time, it is important to evacuate quickly in case of heavy rainfall. In particular, elderly people, who may take a long time to evacuate, need to evacuate earlier. In addition, the recent climate change leads to frequent and short-time heavy rainfall. Because the short-time heavy rainfall rapidly rises river water levels, the time to flood became short and there is a higher risk of extensive flooding damage.

For this reason, river water-level prediction has been actively studied as a disaster countermeasure. Accurate prediction of river water-levels during heavy rainfall is important for residents to make appropriate evacuation decisions. Traditional river water-level predictions are based on physical models that represent the rainfall-runoff process [1][2]. Recently, several studies reported that deep-learning-based models achieve more accurate water-level predictions [3]-[10]. These studies typically learn models from the rainfall and the water-level measurements at upstream observatories to predict downstream water levels.

However, rainfall measurements at observatories don't include all rainfall information in the river basin. Using all rainfall information in the river basin will improve accuracy to predict river water levels. In addition, in many cases, there is not a sufficient number of upstream observatories to make accurate water-level predictions. Many small rivers in Japan do not even have a single observatory upstream. Typhoon Lionrock in 2016 caused flooding in Komoto River in Iwate Prefecture where there is only one water level observatory, resulting in numerous human casualties [11]. Thus, highly accurate river water-level prediction even in small rivers is required. Even when there is no upstream observatory, highly accurate river water-level prediction enables early evacuation even in small rivers.

As other rainfall data sources, radar rainfall observed by Japan Meteorological Agency [12] is available in Japan. In this data, rainfall strength values are observed at 1 km mesh whole of Japan based on radar reflections. Unlike rainfall at observatories, radar rainfall contains rainfall information for the entire watershed because it has been observed at 1km mesh. So even when there is no rainfall observatories at the river watershed, we can obtain all rainfall information in the river basin by using radar rainfall. Water-level prediction method using radar rainfall with deep-learning-based models has been proposed [10], but it also requires a number of rainfall and water-level observatories upstream, and no method without using upstream observatories has been proposed in the literature.

In this study, without using rainfall and water-level observatories upstream of a prediction place, we predict river water levels up to several hours after the current time using only historical water-level at the prediction place and the radar rainfall in the basin. We achieved highly accurate water-level prediction without using upstream observatories, which shows that water-level prediction is possible even for small rivers with only one water-level observatory.

This paper is organized as follows. In Section 2, we present related work. In Section 3, we describe the proposed method. In Section 4, we evaluate the proposed method by using actual measurements of the river. Finally, in Section 5, we summarize this study.

2 RELATED WORK

In Japan, Hitokoto et al. proposed a model for river water level prediction using deep learning, which showed that it can predict the river water level more accurately than the conventional prediction methods that are based on physical models [8]. This prediction model is built with MLP (Multi Layer

Perceptron), and predicts the downstream water level value by using the rainfall and the water-level data at observatories upstream of the prediction place. Yamada et al. showed that LSTM (Long Short-Term Memory), which is a recurrent neural network, is capable of predicting river water-level more accurately than the conventional MLP models [9], meaning that LSTM can learn chronological characteristics of the water-level and the rainfall observation values. However, because these methods use only the rainfall and the water level of the observatory upstream of the prediction place, they do not include all rainfall information in the river basin.

Araki et al. proposed a model based on CNN (Convolutional Neural Network), which uses radar rainfall data in addition to the rainfall and the water level observation values upstream of the prediction place [10]. Each channel of the CNN layer in this model is each time point of radar rainfall measurements. Each of the rainfall and the water level in each observatory forms one independent channel, and adjusts the matrix size by filling the time-series values into the matrix to make the same-size matrix as the radar rainfall mesh. These data are input to the CNN to predict river water level. However, in this method, since each channel contains the time-series rainfall or the water level values, it cannot learn chronological characteristics directly. Furthermore, these deep-learning-based models depend on a number of observatories upstream of the prediction place.

In a similar study using the radar rainfall data, Kimura et al. proposed a model for predicting the water level in sewer culverts [13]. In this model, radar rainfall information is input into CNN, and the water level at close observatories of the prediction place are combined and input into fully connected layers. The output of the fully connected layer is the prediction value. This study shows that using radar rainfall can predict more accurately than without using radar rainfall. However, this study predicts the water level in sewer culverts, so it differs from river water level prediction.

3 PROPOSED METHOD

3.1 Overview of the Proposed Prediction Method

In this study, we propose a deep-learning-based model for river water prediction using radar rainfall data. Radar rainfall data are input to convolution and pooling layers, compressing rainfall information. The compressed rainfall information, and measurements (rainfall and water level values) of the observatories upstream and the prediction place, are input to the LSTM layer. Each time step of the LSTM layer is an hourly interval because the interval of observatories' measurements is one hour. Thus, we combine radar rainfall data, which is 10-minutes interval, into hourly time series, and input them into CNN and each LSTM step. Figure 1 shows a diagram of the proposed river water-level prediction model. In Fig. 1, the left side is the input, and the process proceeds to the right direction, to output at the right side. Each row represents the process for a single time-point of the time series, and an one-hour dataset is processed in the convolution and pooling layers and is input to one step of LSTM. The each step, the learning results are compiled through LSTM, and a water-level prediction value is finally output through the fully connected layers.

3.2 Convolution Layer

The first layer of the prediction model is the convolution layer. In the convolution layer, we convolute the radar rainfall data, and compress its information without losing the areal rainfall information of the radar rainfall data. We use ReLU as an activation function.

Convolution and pooling layers are built for each LSTM step. Rainfall and water-level observation values are at hourly intervals, whereas the radar rainfall data is at 10-minute intervals. Thus, one step of LSTM corresponds to a set of one hour of radar rainfall data including 6 time-points of the 10-minute-interval data, where each channel of the convolution and pooling layer is each time point of the radar rainfall data. Because 10-minute intervals of data are for 1 hour, there are 6 channels. As we will discuss later, a target watershed is divided into 40×30 areas with 1km mesh. Therefore, the input of radar rainfall data into the convolution layer at each LSTM step is $40 \times 30 \times 6$ tensors.

3.3 Pooling Layer

The second layer of the prediction model is the pooling layer. In the pooling layer, the rainfall information compressed at the convolution layer processes pooling, and summarizes information without losing the areal rainfall information of radar rainfall data, and reduces size. Those output values are input to LSTM as compressed information of the radar rainfall. In the proposed method, the pooling process uses max pooling. Because max pooling can extract the maximum values for each divided area, we expect that the model can learn the local rainfall intensity of radar rainfall.

3.4 LSTM Layer

The final layer of the prediction model is the LSTM layer and the MLP layer. In the LSTM layer, rainfall information, which is compressed by the convolution and pooling layer, and the observatories' observation values are input to the LSTM layer, and learn information on each step in chronological order. In the last step of the LSTM layer, outputs of the LSTM layer are input to a fully connected MLP layer, which outputs the water-level prediction value. As we have already noted, each time step of the LSTM layer is an hourly interval, and radar rainfall information that is compressed for one hour. Also, the rainfall and water-level observation values are measured for one hour. Those hourly data are input to each LSTM step. Furthermore, the number of LSTM steps is determined by the number of historical points used in prediction. Specifically, in this study, because we use 11 hours historical data to learn models, the LSTM layer is built with 11 time steps.

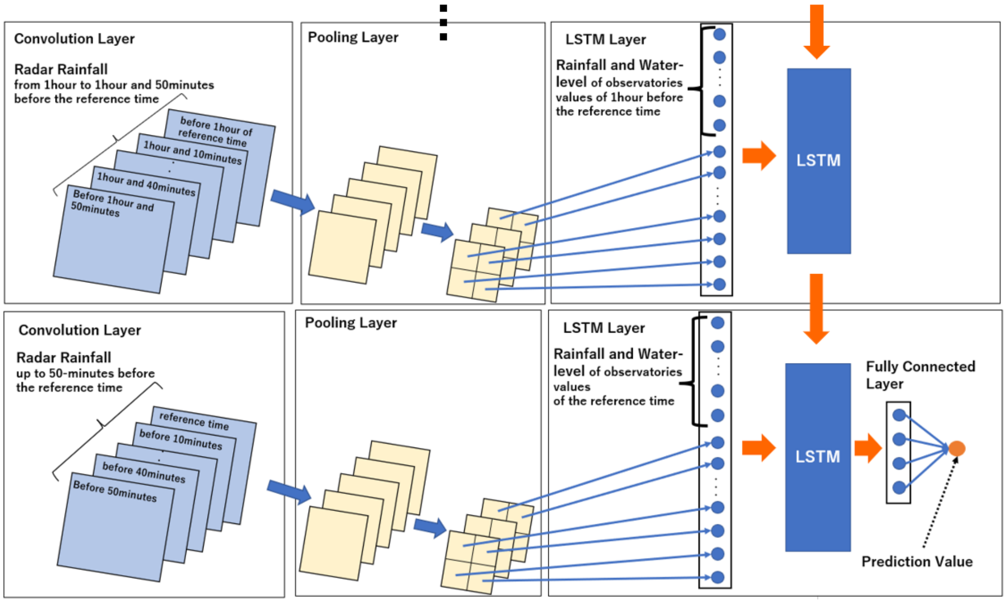


Figure 1: Proposed Prediction Model

4 EVALUATION

4.1 Evaluation Method

In this study, we predict the water level using river proposed model, and show that it is possible to predict it with high accuracy even when we do not use rainfall and water-level observatories upstream of a prediction place. Specifically, we compare the following five prediction models.

A. Prediction model with MLP, which uses rainfall and water level data obtained by the observatories located at upstream and the prediction place.

B. Prediction model with LSTM, which uses rainfall and water level data obtained by the observatories located at upstream and the prediction place.

C. Prediction model with CNN and MLP, which uses radar rainfall data in addition to rainfall and water level data upstream and at the prediction place.

D. Prediction model with CNN and LSTM, which uses radar rainfall data in addition to rainfall and water level data upstream and at the prediction place.

E. Prediction model with CNN and LSTM, which uses radar rainfall data and water level data at the prediction place (without data from upstream observatories).

We train the prediction model using values up to 11 hours before the reference time for prediction, and predict 1 to 12 hours after the reference time. We use Mean Squared Error (MSE) to evaluate the accuracy of our model.

4.1.1 Study Area

In this study, the study area is the Hiwatashi point of the Oyodo River, which is a first-class river in Miyazaki prefecture, Japan. There are 5 water level observatories and 12 rainfall observatories. The locations of these observatories are shown in Fig. 2. In this area, there are no flood control facilities upstream of the Hiwatashi point, which is the prediction place. There are a lot of water level and rainfall observatories upstream.

4.1.2 The Dataset

We acquire water level and rainfall data, which are hourly data, from the website of the hydrology and water-quality database [15]. We acquire radar rainfall data, which are 10-minutes-interval, from Japan Meteorological Business Support Center [16]. Radar rainfall data are observed at 1km mesh whole of Japan, and we use as input data 40 km long and 30 km wide regions shown in Fig. 2.

The prediction place is the Hiwatashi water level observatory, and the target period is the 13 periods from 2006 to 2021 when the Hiwatashi water level observation value is higher than the dangerous water level 1 (5.4m) [17][18] defined at each river observatory by the government. (The actual number of such periods is 14 periods, but we exclude one period because radar rainfall has missing values.) We define that a such period is 120 hours long, from 72 hours before to 48 hours after the time of peak water level of Hiwatasi observation value during the target periods. Among them, we select three evaluation periods for water level prediction (July 3, 2010, September 20, 2011, and July 3, 2019) when the Hiwatashi water level observation value is higher than the dangerous water level 3 (8.3m) [17][18] in the 13 periods. We evaluate the prediction accuracy of the models A-E by predicting one of the 13 periods with the other 12 periods used as the training data, which is applied for evaluating the three test periods.

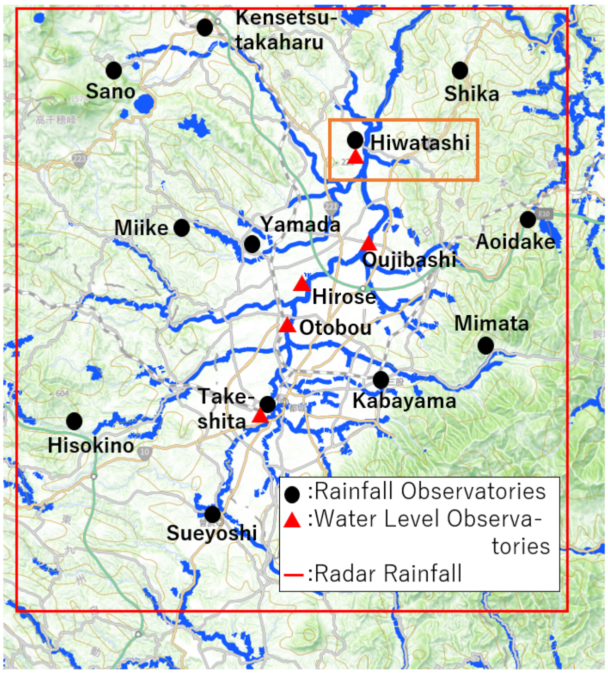


Figure 2: Study Area
(Created Using GIS Map of Geospatial Information Authority of Japan [14])

Table 1: Structure and Parameters of the Model

Classification	Parameter
Optimizer	Adam (0.0001)
Epoch	Early Stopping (50 times)
Error function	MSE (Mean Squared Error)
Batch size	90
Language	Python
Library	PyTorch

4.2 Details of the Prediction Model

4.2.1 Normalization of Training Data

If training data contain bias, they impose a negative effect on learning. Therefore, in this proposed method, we normalize and convert to the range of $[0, 1]$ the training data before learning the model. Normalization is made as shown in the following Eq. (1).

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \tag{1}$$

where x'_i is the normalized value of the i-th observation value, x_i is the raw observation value, $\min(x)$ is the minimum value, $\max(x)$ is the maximum value in the training data. This normalization is applied to all the input data, i.e., water level and rainfall data of the observatories, and the radar rainfall data.

4.2.2 Structure and Parameters of the Model

We show the structure and parameters of our learning model in Table 1. We use Adam [19], which is a well-known dynamic

Table 2: The Number of Hidden Units for MLP

	Hidden Units
No Radar Rainfall (Model A)	30-20-10
	50-30-10
	100-50-20
	150-100-50
	500-300-100
	800-400-200
	1000-500-100
	2000-1000-500
Using Radar Rainfall (Model C)	2000-1000-500
	3000-1000-500
	5000-3000-1000
	7000-4000-1000
	9000-5000-1000
	10000-6000-2000

Table 3: The Number of Hidden Units for LSTM

	Hidden Units
No Radar Rainfall (Model B)	6
	10
	14
	18
	50
	100
	150
	200
	500
Using Radar Rainfall (Model D,E)	200
	300
	500
	1000
	2000
	3000

Table 4: Parameter of CNN

Parameter		Value
Convolution Layer	Kernel size	3×3
	Number of Filter for MLP	128
	Number of Filter for LSTM	6
	Stride	2
	Activation Function	ReLU
Pooling Layer	Classification	Maxpooling
	Window size	2×2

learning parameter adjuster. To prevent overlearning, we determine the number of epochs by using Early Stopping [20], and stop learning if accuracy does not improve 50 times. We use stochastic gradient descent (SGD) to learn the network, and use backpropagation to calculate the gradient of the error function.

Table 5: Parameter of the Best Accuracy

Prediction Model	Hidden Units	Number of Layers	Dropout
A. MLP with upstream observatories	50-30-10	-	0.1
B. LSTM with upstream observatories	100	2	0.3
C. CNN+MLP with upstream observatories and radar rainfall	9000-5000-1000	-	0.1
D. CNN+LSTM with upstream observatories and radar rainfall	500	1	-
E. CNN+LSTM with water level at prediction place and radar rainfall	1000	1	-

4.2.3 Parameter Evaluation

In this study, we tried a variation of the parameter value for each of the five prediction models A-E, and compare the prediction accuracy to determine the optimal parameter values. For the MLP model (model A and C), we use ReLU as an activation function, and build a five-layer network consisting of an input layer, three hidden layers, and an output layer. Table 2 shows the number of hidden units in the prediction model of using MLP. Each number in the triplet expresses the number of neurons of an input, intermediate, and output layer, respectively. The dropout parameter is set as 0.1 or 0.3. We predict water levels with all combinations of the number of hidden units and the dropout, and evaluate prediction accuracy.

For the LSTM model (model B, D, and E), we use the number of hidden units shown in Table 3. The number of layers is 1, 2, or 4, and in the case of 2 or 4 layers, a dropout is set as 0.3. We predict water levels with all combinations of the number of hidden units and the number of layers of the LSTM, and evaluate prediction accuracy.

We show the CNN parameter for the prediction model with radar rainfall data in Table 4. For the prediction model with CNN and MLP (model C), each channel of the CNN is 10-minute intervals of radar rainfall data. In prediction model C, because 10-minute intervals of data are input for 11 hours, there are 66 channels. Therefore, the input of radar rainfall data is a $40 \times 30 \times 66$ tensor.

4.3 Results

4.3.1 Parameter Results

We show the parameter values with the best accuracy for each of the prediction models A-E in Table 5. For each prediction model, we predict water levels by using these parameter values to evaluate the prediction accuracy in the next section.

4.3.2 Accuracy for the Three Prediction Periods

We show the prediction accuracy of each prediction model from 1 to 12 hours after the reference time for the three prediction periods in Fig. 3. The vertical axis is MSE, and the horizontal axis is the prediction time. Figure 3 shows that the prediction accuracy in all five cases is not significantly different. In particular, the MSE of the prediction model E, which does not use the measurements of the upstream observatories, does not increase MSE significantly compared to the MSE of A-D.

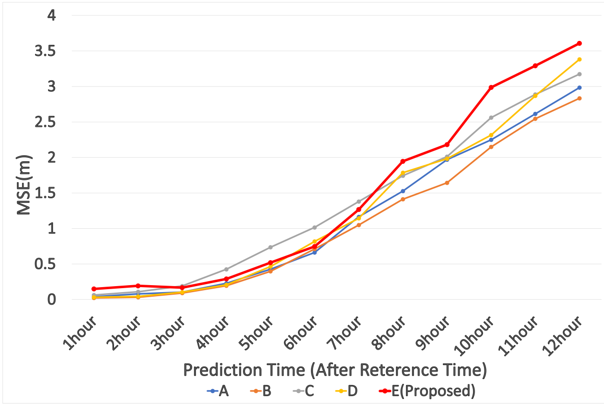


Figure 3: MSE of Each Prediction Model at Three Prediction Periods

5 CONCLUSION

In this study, we propose a river level prediction model based on deep-learning models, CNN and LSTM, with radar rainfall data. We predict the river levels up to several hours after the current time with the proposed model that uses the historical water levels at the prediction place and the radar rainfall data, without using the data of upstream observatories.

The prediction accuracy of the proposed model is evaluated by comparing it with the models that use the upstream water level and rainfall measurements. The results showed that the proposed model using the historical water level at the prediction place and the radar rainfall in the basin can predict water levels with a small error, even if it does not use the data of the upstream observatories. This means that water level prediction is possible even for rivers with only one water-level observatory, which are often the case of many small rivers in Japan.

ACKNOWLEDGEMENT

This work was supported by JST, PRESTO Grant Number JPMJPR1939, Japan.

REFERENCES

[1] Y. Tachikawa, T. Sayama, K. Takara, H. Matsuura, T. Yamazaki, A. Yamaji, and Y. Michihiro, “Development of a Real-Time Runoff Forecasting System Using a Physically-Based Distributed Hydrologic Model and its Application to the Yodo River Basin, ” Natural Disaster Science J. JSNDS, Vol. 26-2, pp. 189-201, 2007 (in Japanese).

[2] Y. Nakamura, T. Koike, S. Abe, K. Nakamura, T. Sayama, and K. Ikeuchi, “Development of Real-Time Flood Prediction Utilizing the RRI Model with a Particle Filter, ” Journal of Japan Society of Civil Engineers, SER. B1 Hydraulic Engineering, Vol. 74, No. 5, I_1381-I_1386, 2018 (in Japanese).
[3] S-S. Baek, J. Pyo, and A. Chun, “Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach,” Water, 12, 3399, 2020.
[4] P. Li, J. Zhang, and P. Krebs, “Prediction of Flow Based on a CNN-LSTM Combined Deep Learning Approach,” Water, 14, 993, 2021.
[5] X. Li, W. Xu, M. Ren, Y. Jiang, and G. Fu, “Hybrid CNN-LSTM models for river flow prediction,” Water Supply, 22(5), 4902, 2022.
[6] B. Alizadeh, A.G. Bafti, H. Kamangir, Y. Zhang, D.B. Wright, and K.J. Franz, “A Novel Attention-based LSTM Cell Post-processor Coupled with Bayesian Optimization for Streamflow Prediction,” Journal of Hydrology, 601, 126526, 2021.
[7] C. Chen, J. Jiang, Z. Liao, Y. Zhou, H. Wang, and Q. Pei, “A Short-term Flood Prediction Based on Spatial Deep Learning Network: A Case Study for Xi County, China,” Journal of Hydrology, 607, 127535, 2022.
[8] M. Hirokoto, M. Sakuraba, and Y. Sei, “Development of the Real-Time River Stage Prediction Method Using Deep Learning, ” Journal of Japan Society of Civil Engineers, SER. B1 Hydraulic Engineering, Vol. 72, No. 4, pp. I_187- I_192, 2016 (in Japanese).
[9] K. Yamada, Y. Kobayashi, M. Nakatsugawa, and J. kishigami, “A Case Study of Flood Water Level Prediction in the Tokoro River in 2016 Using Recurrent Neural Networks, ” Journal of Japan Society of Civil Engineers, SER. B1 Hydraulic Engineering, Vol. 74, No. 5, pp. I_1369- I_1374, 2018 (in Japanese).
[10] T. Araki, K. Hakoishi, M. Hitokoto, T. Shimamoto, and K. Fusamae, “River Water Level Prediction by Convolutional Neural Network Using Radar Rainfall, ” Advances in river engineering, Vol. 25, pp. 297-302, 2019 (in Japanese).
[11] Ministry of Land, Infrastructure, Transport and Tourism: River Information Policies <http: //www.river.or.jp/kougi2020_5_rev.pdf>(accessed 2023-06-05).
[12] Japan Meteorological Agency Weather Radar <https: //www.jma.go.jp/jma/kishou/know/radar/kaisetsu.html >(accessed 2023-06-05).
[13] Y. Kimura, “Prediction of Water Level in Sewer Pipe Using Convolutional Neural Network, ” Journal of the Society of Instrument and Control Engineers, Vol. 58, No. 6, pp. 449- pp. 453, 2019 (in Japanese).
[14] Ministry of Land, Infrastructure, Transport and Tourism: Geospatial Information Authority of Japan, <https:// www.gsi.go.jp/top.html >(accessed 2023-06-05).
[15] Ministry of Land, Infrastructure, Transport and Tourism: the hydrology and water-quality database, <http:// www1.river.go.jp/ >(accessed 2023-06-05).
[16] Japan Meteorological Business Support Center, <http: //www.jmbsc.or.jp/jp/ >(accessed 2023-06-05).
[17] Ministry of Land, Infrastructure, Transport and Tourism: List of Water Levels Related to Flood Prevention for Directly Controlled Rivers, <https://www.mlit.go.jp/ river/toukei_chousa/kasen_db/pdf/2021/12-1-8.pdf >(accessed 2023-06-05).
[18] Ministry of Land, Infrastructure, Transport and Tourism: Water and Disaster Management Bureau Implementation Guidelines for Revision of Disaster Prevention Information System for Floods, <https://www.mlit.go.jp/river/shishin_guideline/ gijutsu/saigai/tisiki/disaster_info-system/ >(accessed 2023-06-05).
[19] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization, ” arXiv: 1412. 6980, 2014.
[20] L. Prechelt, “Early Stopping - But When?, ” Springer Berlin Heidelberg, Berilin, Heidelberg, pp. 53- pp. 67, 2012.

Developing an Information System to Enhance Intellectual Productivity in Teleworking Environments

KANAE MATSUI†* KATSUMA TAKAGI†**

*School of System Design and Technology, Tokyo Denki University, Japan
**Graduate School of Informatics, Tokyo Denki University, Japan
{matsui@mail, 21amj12@ms }.dendai.ac.jp

Abstract –The COVID-19 pandemic highlighted the necessity for efficient telework settings, as intellectual productivity is closely linked with workplace satisfaction. Addressing the lack of productivity systems in home environments, we created an IoT tool that uses temperature, humidity, and CO2 measurements to provide relevant feedback. A two-month trial demonstrated its effectiveness in promoting behavioral changes, such as using humidifiers, ensuring its operational reliability.

Keywords: IoT application, Healthcare, Data collection

1 INTRODUCTION

The COVID-19 outbreak led to a significant shift from office to homework, with a survey showing 80% of Japanese firms implementing telework (Figure 1). However, over 80% of both employees and employers reported a productivity drop in teleworking, due to factors like poor communication, inadequate equipment, and home distractions. Environmental factors, such as temperature, humidity, and CO2 levels, though initially underestimated, are proven to be crucial for productivity [1]. In the U.S., improving these aspects could result in economic gains of $40-$200 billion annually (Table 1), prompting companies to integrate BEMS and smart offices for better office productivity.

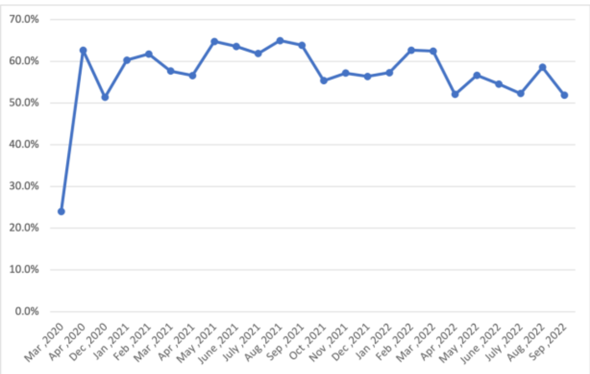


Fig 1. Telework implementation rate at companies in Tokyo (30 or employees) [2]

Table 1. Economic effect of improving air quality in the United States

Source of productivity gain	Potential US annual savings
Reduced respiratory illness	$6-$14 billion/year
Reduced allergy and asthma	$2-$4 billion/year
Reduced sick building syndrome symptoms	$10-$30 billion/year
Improved worker performance from changes in thermal environment and lighting	$20-$160 billion/year

Both workers and employers observed over an 80% decline in telecommuting productivity due to communication challenges, inadequate equipment, and home disturbances. Environmental factors such as temperature, humidity, and CO2 levels, crucial for productivity, were often overlooked, with their improvement in the U.S. potentially offering $40-$200 billion in economic benefits (Table 1). Consequently, many firms are now implementing systems like BEMS to improve workspace productivity.

Additionally, about 70% of telecommuters faced home air quality issues, underscoring the need for better environmental conditions and information systems. Despite known benefits, few companies have optimized home thermal conditions for productivity. Our research utilized the “Netatmo Weather Station [2]” IoT device to monitor environmental factors, leading to an IoT-based system to enhance home work environments and productivity.

2 RELATED STUDY

2.1 Telework situation in Japan

The Ministry of Internal Affairs and Communications describes telework as a “flexible approach leveraging ICT for efficient time and place usage [3].” There are three primary types of telework based on work location: telecommuting, mobile work, and satellite office work. Due to COVID-19, telework adoption has surged, with many companies making home-based telework standard practice [4]. Owing to its adaptability, telework offers numerous benefits for society, corporations, and employees. Our research specifically examines telecommuting within these telework categories.

2.2 Novelty of our study Indoor environment and intellectual productivity

Hamanaka et al. [5] developed a teaching bot designed to optimize indoor comfort. Integrated into the chat tool “Slack,” the bot offers real-time feedback on indoor temperature and humidity. It sources data from a sensor on the Raspberry Pi 3, recording temperature and humidity values every minute via Wi-Fi. Upon querying the bot, users receive recommendations for enhancing comfort, derived from real-time data and optimal comfort ranges. This bot boasts two core features: querying the database to generate and send graphical data on demand.

2.3 Novelty of our study

The majority of IoT systems for indoor optimization target office spaces, with a gap in catering to individual telecommuters or small teams. Systems like HEMS focus on residential energy use rather than telecommuting needs. Our research aims to create a health-conscious, efficient workspace by developing a recommendation system tailored for home telecommuting. This system faces challenges in universal adaptability due to the variety of household air-conditioning devices. It provides customized information to improve working conditions based on indoor parameters, allowing manual adjustments to air-conditioning. The system design prioritizes intuitiveness, with easily understandable information even during work hours. We will detail this system's structure, technical aspects, experimental methods, evaluation, and our findings in the following sections.

3 SYSTEM DESCRIPTION

3.1 Overview

Our system uses strategically placed sensors to monitor workspace conditions, generating notifications for maintaining or enhancing the work environment. Workers can act upon these alerts to foster productivity.

3.2 System configuration

The system, outlined in Fig. 2, relies on IoT devices across telecommuting setups, sending environmental data to a central server. This data, both real-time and historical, is used to create recommendations, delivered to workers as notifications. Upcoming sections detail the system's design and functionality.

3.2.1 Data acquisition part

Our system's data acquisition relies on Netatmo devices, consisting of a primary indoor module for internal environment monitoring and a secondary outdoor module. The primary device sends data directly to Netatmo's server via Wi-Fi, while the secondary device transmits data indirectly through the primary device. Netatmo's server offers unlimited storage and API access for data integration into custom systems.

Our developed platform centralizes data from multiple Netatmo devices, linked to their respective accounts, ensuring data consolidation and accessibility. Data is updated every five minutes and stored in BSON format using MongoDB, with Heroku as the application server. The development environment details are in Table 3.

Table 3. Development environment of Platform

	Server-side	Client-side
OS	MacOS	MacOS
Language	Node.js (v10.18.1)	HTML, CSS, (SCSS), JavaScript
Library	express, ejs, crypto, Socket.IO	Vue.js, Bootstrap-Vue, Core UI, chart.js, Axitos
Database	MongoDB	

3.2.2 Information generator

This component analyzes real-time and historical data to create insights for an optimal work environment. It generates:

1) Provides immediate values of indoor environmental factors with feedback for improvement.
2) Displays daily fluctuations in indoor factors using graphs, with comments to ensure optimal conditions.
3) Summarizes the previous day's and week's environment suitability, with averages and improvement suggestions.
4) Offers advice on maintaining or enhancing the indoor environment.

Our system uses building environmental hygiene control standards to define ideal workspace conditions (Table 4) [7].

Table 4. Ideal working environment

Category	Detail
Carbon dioxide concentration	Less than 1000 ppm
Temperature	1) Range: 17°C to 28°C 2) If the indoor temperature is lower than the outdoor, the difference should be significant.
Relative humidity	Range: 40% to 70%

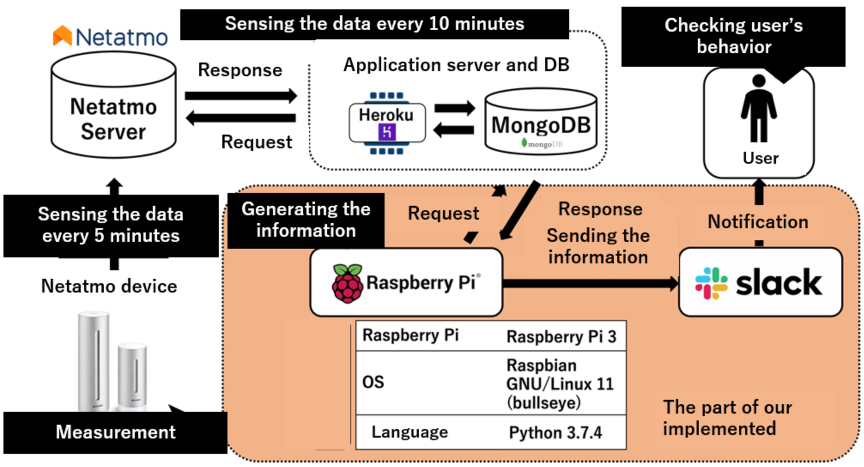


Fig 2. System overview

This presents hourly updates on the current state of the workspace environment, specifically focusing on temperature, humidity, and carbon dioxide concentration. Relevant comments are determined based on conditional branching, rooted in the previously defined criteria. Here are the tables for humidity and CO2.

Table 5. Conditional branching and comments in humidity

Condition	Comment
Less than 40%	The room is dry. Dryness increases the number of times you blink, causing discomfort and reducing work efficiency. It's recommended to humidify.
From 40% to 70%	Humidity is optimal. Let's keep this humidity.
More than 70%	Humidity in the room is high. In a humid environment, the sympathetic and parasympathetic nerves are out of balance, making it easier to feel tired. Let's dehumidify.

Information of type 2 provides daily visuals and comments on the fluctuation of temperature, humidity, and CO2 from 8:00 AM to 7:00 PM. Here are the conditions and comments for temperature, humidity, and CO2.

This displays how closely the workspace met the building environmental hygiene standards. It presents the "indoor environment adequacy rate", which is calculated using the following formula [A].

$$Indoor\ environment\ adequacy\ rate = \frac{Number\ of\ instances\ meeting\ standards\ during\ work\ hours}{Total\ observed\ instances\ during\ work\ hours} * 100 \ldots [A]$$

The system keeps telecommuters informed about their workspace environment, allowing them to make necessary adjustments for maximum comfort and productivity. Every day at 12:00, a report is generated highlighting research findings on the correlation between indoor environment and productivity, along with suggested measures to enhance the indoor ambiance in information of type 4. By grounding suggestions in scientific evidence and providing explicit action steps, the aim is to foster a heightened awareness and proactive attitude towards the workspace environment among users.

3.2.3 Information notification part

In our system, we integrated three key notification functions using Slack [6] as the communication platform (Figure 3 and 4). Each participant had a dedicated Slack channel for receiving notifications from the "Indoor Environment Notification Bot." The functionalities include:

A feature enabling periodic dispatch of the four types of data previously detailed.

1) Periodic Data Dispatch: Sends the four types of data (current status, daily trends, historical analysis, and guided articles) at set intervals, focusing on work hours for minimal distraction.
2) Threshold Alerts: Monitors environmental data every 30 minutes and alerts users when humidity and CO2 levels exceed set thresholds, offering concise evaluations and suggestions.
3) Interactive Chatbot Response: Provides real-time feedback to users' queries about their current work environment. It utilizes the Predicted Mean Vote (PMV) index to assess thermal comfort, considering factors like temperature, humidity, air velocity, and personal factors such as clothing insulation and metabolic rate. This allows tailored advice based on environmental data from Netatmo and user-specific inputs like clothing type, enhancing comfort and productivity. The system's approach to measuring and responding to thermal comfort is detailed, including predefined clothing insulation levels for accurate PMV calculations.

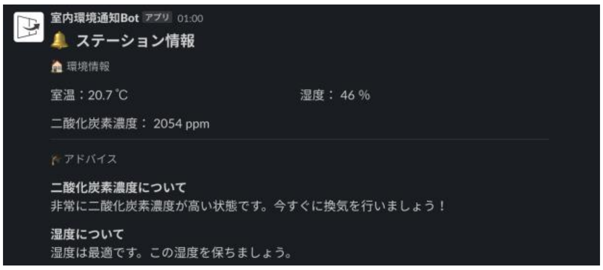


Fig 3. Actual condition of information 1

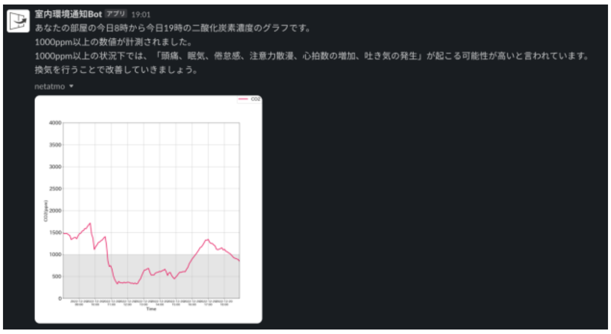


Fig 4. Actual condition of information 2 (CO2)

When a user mentions the term "working environment" in their designated channel, the bot prompts them to specify their current clothing [8 and 9]. After receiving the user's clothing input, the system incorporates the Predicted Mean Vote (PMV) value into the data from Presentation Information 1 and then notifies the user accordingly.

4 EXPERIMENT

4.1 Overview of the experiment

The experiment aimed to collect indoor environmental data for telecommuting using Netatmo, and utilize it to enhance indoor conditions and influence behavior. Effectiveness was evaluated through interviews before, during, and after the experiment, which lasted 33 days from November 10 to December 28, 2022.

4.2 Participant demographics

Three participants from the Kanto region, aged 20s to 50s and working from home, were involved. Their demographics are detailed in Table 14, including age, gender, occupation, residence, and workspace specifics.

Table 5. Participants attributes of the study

	A	B	C
Age	50's	40's	20's
Gender	male	male	Female
Occupation	Engineer	Technical sales	Office work
Residence status	Apartment	Apartment	Apartment
Workplace	Study	Living	Living
Size of workplace	18 m^2	22 m^2	14 m^2

4.3 Experimental method

4.3.1 Participants

Participants set up accounts on Netatmo and Slack and received device kits (Table 7). They installed indoor modules in their workspaces and outdoor modules in sheltered outdoor areas.

Table 7. Contents of the experimental kit

No	Details
1	Netatmo Weather Station outdoor module
2	Netatmo Weather Station indoor module
3	AC adapter for Netatmo Weather Station indoor module
4	Kit bag
5	Covering letter

4.3.2 Pre-experiment questionnaire

A survey, inspired by Daikin Industries [6], assessed participants' initial perceptions of their telecommuting environment. It inquired about housing types, environmental factors, and potential improvements.

4.3.3 Questionnaire during the experiment

Daily surveys via SurveyMonkey asked about work location, duration, and environment evaluations, focusing on days spent telecommuting.

4.3.4 Post-experiment questionnaire

A post-experiment questionnaire (Table 8) evaluated the system's effectiveness and any shifts in participants' environmental improvement awareness. It included questions on device placement, installation ease, notification frequency, and the system's impact on work environment quality.

Table 8. Questionnaire after the experiment

No.	Details
1	Where did you position the indoor module?
2	Where was the outdoor module placed?
3	Please specify the dimensions of the room where the device was set up. (Provide the measurement in square meters.)
4	Did you find the installation process for Netatmo to be intuitive?
5	Was the notification frequency of once every hour satisfactory?
6	Could you provide further clarification on your answer to question 5?
7	Was receiving a daily summary notification at 19:00 suitable for you?
8	Please elucidate your response to question 7.
9	Was the daily 8:00 am summary notification, recapping the previous day/week, convenient for you?
10	Please tell me if there is a reason for answering question 9.
11	Was the 12:00 noon notification for the column function apt for your needs?
12	Please tell me if there is a reason for answering question 11.
13	Kindly expound upon your response to question 11.
14	Which set of indoor environmental data was deemed most valuable by you?
15	In your estimation, did the system enhance the work environment's conduciveness?
16	When telecommuting, are there particular facets of your home setting that present concerns? (You may select multiple options.)
17	Looking ahead, are there specific components or resources you're considering investing in or improving for an optimized telecommuting experience?

5 EVALUATION

5.1 System evaluation

The system's effectiveness in enhancing telecommuting environments was assessed. Notifications were periodically sent to participants, including alerts for sub-optimal humidity and CO2 levels. The frequency and types of these notifications, both during and outside work hours, are detailed in Table 17. An "information provision rate" was calculated to measure the system's efficiency in delivering notifications (Table 9).

Table 9. Number of notifications

		A		B		C	
		During office hours	Out of the office hours	During office hours	Out of the office hours	During office hours	Out of the office hours
Notification function	Type of information 1	182	296	160	349	82	134
	Type of information 2	20		20		9	
	Type of information 3	13		14		5	

	Type of information 4	13		14		5	
Alert function	Humidity	42	149	0	0	0	0
	CO2	94	7	255	537	7	45
Chatbot function	-	0	0	0	0	0	0

To assess the system's operational efficacy, we compared the actual frequency of notifications to an ideal count for the days participants worked from home, deriving a metric termed the "information provision rate". The calculation for this rate, denoted by equation [B], is as follows:

$$Information\ provision\ rate = \frac{Number\ of\ actual\ notifications}{Ideal\ number\ of\ notifications} * 100 \quad \ldots \text{[B]}$$

Table 18 encapsulates the information provision rates deduced using the formula.

Table 10. System availability

	A	B	C
Type of information 1	84%	84%	83%
Type of information 2	83%	83%	83%
Type of information 3	100%	100%	100%
Type of information 4	100%	100%	100%

5.2 Questionnaire evaluation

The post-experiment questionnaire assessed participants' adjustments to their work environments, focusing on air conditioning, humidification, and ventilation (Table 11). Despite prior management of temperature and humidity, there was room for improvement in environmental management, as revealed in Table 12. The system was found to significantly contribute to a better work environment, with participants intending to continue improving their telecommuting spaces.

Table 11. Questionnaire about notification frequency

	A	B	C
Type of information 1	4	4	4
Type of information 2	5	5	4
Type of information 3	3	5	4
Type of information 4	4	5	4

Table 12. Questionnaire about system usefulness

	A	B	C
Was the type of information 1 useful?	3	5	4
Was the type of information 2 and 3 useful?	5	5	4
Was the type of information 4 useful?	3	5	4

5.3 Improvement of work environment

5.3.1 Quantitative evaluation

We assessed how the system influenced participants' work environments using data on temperature, humidity, and carbon dioxide levels. Participant An experienced stable

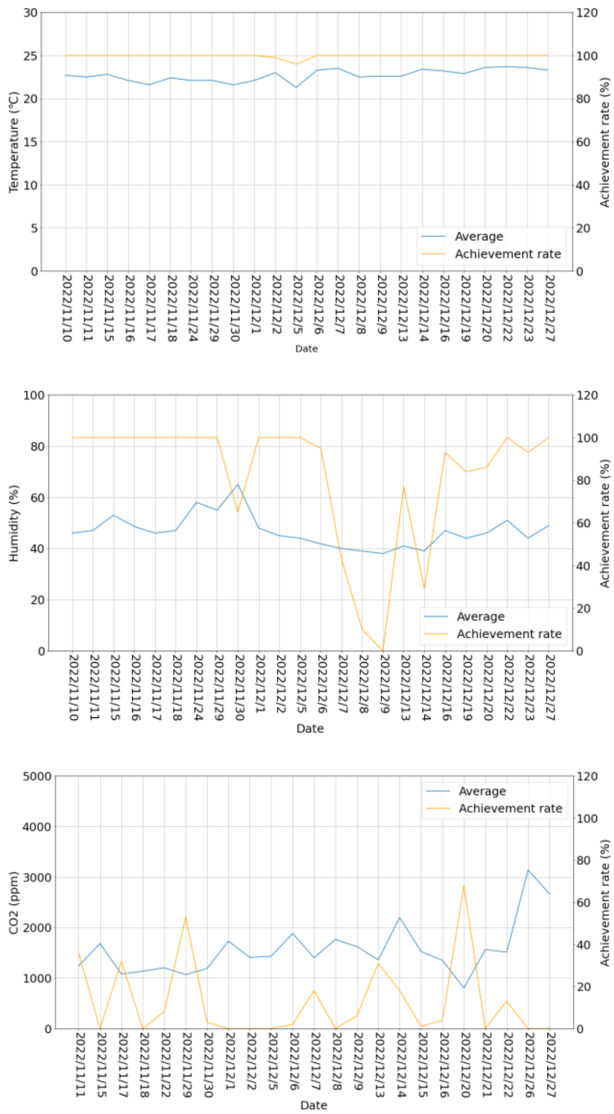


Fig 10. Environmental data of each participants

temperature but variable humidity and CO2 levels. Participant B maintained stable temperature and humidity, but CO2 levels fluctuated. Participant C had a consistently stable environment.

5.3.2 Qualitative evaluation

Feedback from the post-experiment questionnaire was used to evaluate improvements in the working environment. Participants confirmed using air conditioning, humidifiers, and ventilation, and monitoring environmental parameters. Despite prior efforts to manage the environment, there were gaps in effective utilization of data and environmental control. Table 14 shows participants' ratings of the system's impact on their work environment and their intentions for continued improvement. Scores on a 5-point Likert scale indicated a positive contribution of the system to the work environment and a strong intention to maintain these improvements.

Table 14 . Questionnaire about improvement of telecommuting environment

	A	B	C
System's contribution to a good work environment?	5	4	5
Intent to continue improving the work environment?	5	5	5

6 DISCUSSION

6.1 Information notification system

We discuss our system, which transmitted four types of data and included a chatbot feature for enhancing work environments. Despite a non-operational period, due to Heroku account issues, the system was generally efficient. However, the chatbot feature was underutilized, possibly due to participants' work constraints.

6.2 Discussion derived from questionnaire results

This segment deliberates on the insights obtained from questionnaires administered pre, during, and post-experiment.

6.2.1 Information notification system

Usability insights of the information notification system, sourced from the post-experiment questionnaire, are explored here. Feedback on the four pieces of presented information was generally favorable. Notification intervals, the shortest being hourly, were well-received due to their concise and instantly comprehendible design.

6.2.2 Improvement of work environment

Participants understood their workspace conditions better, leading to conducive work environments. For example, Participant A used a humidifier post-notification, showing a behavioral change, while Participant B did not significantly alter behavior despite high CO2 levels. Challenges such as ventilation impacts and external noises were noted, highlighting the limitations of relying solely on information notifications for behavioral change. The study suggests a need for automated environmental regulation but acknowledges its complexity.

Table 13 . Type of work and the total days during experiment

Type of work	A	B	C
Office work	0	2	5
Meeting	11	14	3
Programming	12	0	0
Write documentation	13	12	3
Thinking	5	5	0
Other	0	0	0

6.2.3 Questionnaire Considerations

The study assessed the system's utility and the shift in workspace environment awareness. Questionnaire responses indicated potential improvements in notification timings and data presentation. Participants showed an increased drive for environmental improvement, validating the experiment's impact. Daily task inquiries were found useful for data analysis and could lead to more tailored suggestions in the future.

7 CONCLUSION AND REMARKS

The system, aimed at improving the telecommuting environment, faced challenges, especially with Slack as the notification platform due to its chat-based design. Overall, the system proved useful but highlighted areas for further improvement.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP22H03572.

References

[1] Kawaguchi, G. et al., "Intellectual Productivity Evaluation in Indoor Environments: (Part 8) Impact of Improved Thermal Environment Satisfaction through Quantitative Methods on Intellectual Productivity," Journal of Air Conditioning and Sanitary Engineering of Japan, 2004.

[2] Netatmo, "Smart home weather station", URL : https://www.netatmo.com/en-gb/weather/weatherstation/indoor-module, (Accessed November15, 2023).

[3] Nishino, J. et al., "Research and Development of AI & IoT-based Intellectual Environment Control System to Realize Spaces that Enhance Intellectual Productivity (3rd Report) Verification Evaluation of the Effect of Work Efficiency Improvement by Air Conditioning Control Based on Sleepiness," Proceedings of the Air Conditioning & Sanitary Engineering Society Conference, Vol. 8, pp. 381-384, 2019.

[4] Ministry of Health, Labor and Welfare, "Guidelines for the Proper Introduction and Implementation of Telework," URL: https://www.mhlw.go.jp/content/000759469.pdf , (Accessed November15, 2023).

[5] Hamanaka, G and et al., "Implementation and Evaluation of a Teaching Bot for Improving Indoor Comfort," Research Report on Ubiquitous Computing Systems, No. 9, pp. 1-6, 2017.

[6] Slack, URL : https://slack.com/intl/ja-jp/, (Accessed August 3, 2023).

[7] Ministry of Health, Labor and Welfare, "Standards for Building Environmental Health Management," URL: https://www.mhlw.go.jp/bunya/kenkou/seikatsueisei10/, (Accessed November15, 2023).

[8] National Institute of Health and Nutrition, URL : https://www.nibiohn.go.jp/eiken/programs/2011mets.pdf, (Accessed November15, 2023).

[9] FREE HAND, "Clo Value Laboratory," URL : http://freehandjapan.com/, (Accessed November15, 2023).

Keynote Speech 1:

Dr. Hidetoshi Mishima

(Mitsubishi Electric Corporation
Corporate Research and Development
Group Executive Fellow)

(Chair: Akihiro Hayashi)

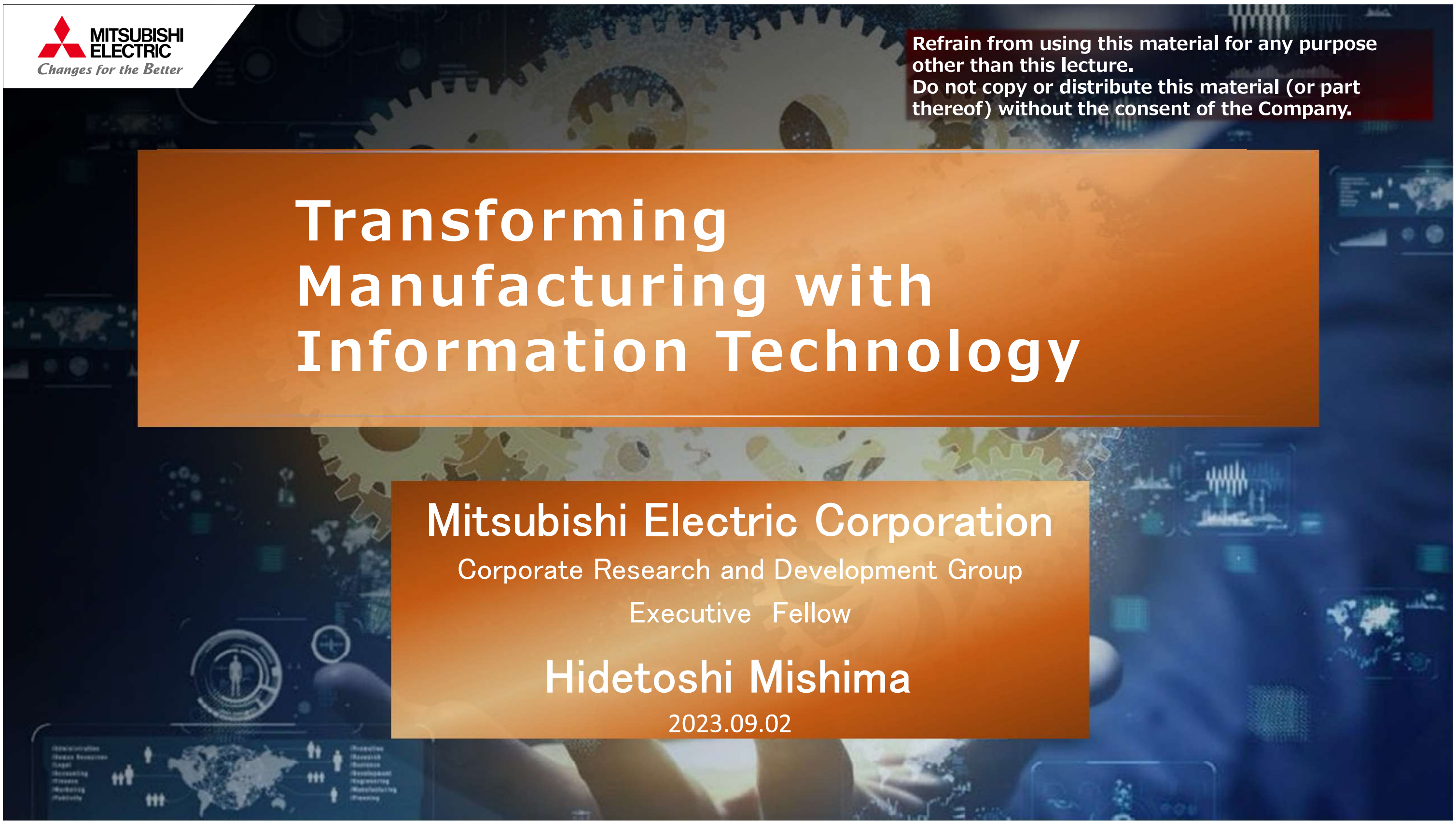


Table of contents

 2

Sustainability Management

3. Initiatives for Sustainable Growth

Respond to key societal challenges in five basic areas in order to

Key societal challenges

Carbon neutral	Circular economy	Safety/Security	Inclusion	Well-being
Decarbonize society	Build a circular society	Resilient society	Society that values diversity	Vibrant lives

Initiatives for realizing sustainability

Contribution to a sustainable society through business

- Decarbonize society with energy management and heat pumps technologies etc.
- Build a circular society through innovation in the fields of maintenance/ operation and recycling
- Build a resilient society through maintenance/management of infrastructures and facility automation etc.
- Realize vibrant lives of all people through sophisticated use of people flow data and personal data

Progress in social initiatives causes positive impacts on the Company

Realize sustainability

Apply initiatives within the Company to business

Responsibilities for a sustainable society

- Strengthen efforts for cutting greenhouse gas emissions
  - FY2051: Aim for net zero greenhouse gas emissions in the entire value chain
  - FY2031: Aim for net zero greenhouse gas emissions from factories and offices
- Respect for human rights based on international norms
- Realize workplace where diverse and versatile human capital gathers and works together
- Adherence to compliance

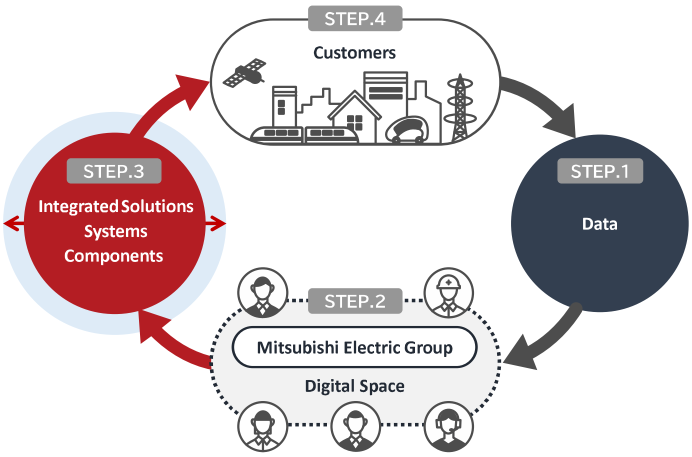
3

"Circular Digital-Engineering"

3. Initiatives for Sustainable Growth

Leverage strong intra-group connections and knowledge sharing in order to consolidate and then analyze data in the digital space to create new value and contribute to solving key social challenges

Solving social challenges through Circular Digital-Engineering



STEP.1
Consolidate diverse data
Consolidate data generated through usage of our components and systems by a wide range of customers into a digital space

STEP.2
Ascertain potential issues and needs of customers
Analyze the consolidated data via a digital space and ascertain potential issues and needs of customers

STEP.3
Create new value
Create new value through evolving components, systems and Integrated Solutions based on potential issues and needs

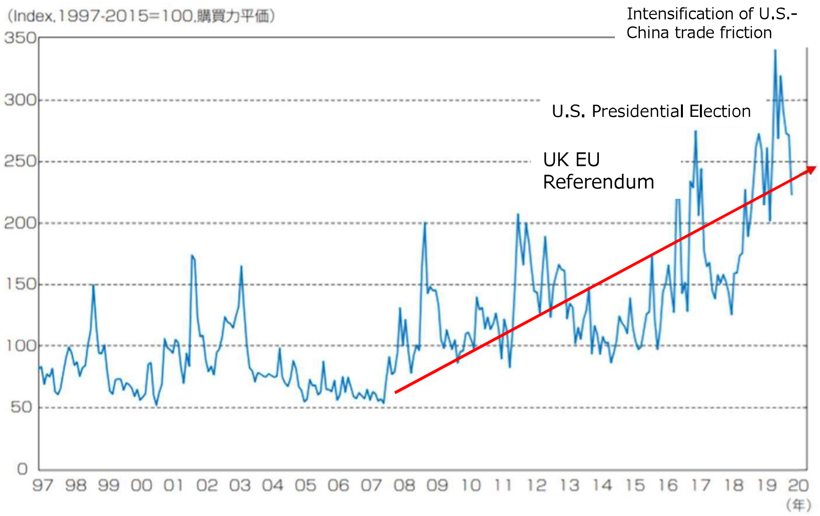
STEP.4
Return value to a wide range of customers
Return new value to a wider range of customers, and solve social challenges with customers

4

An era of uncertainty

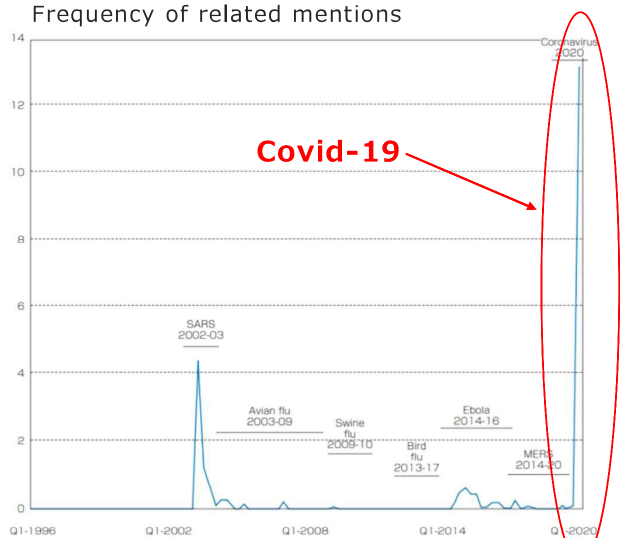
Global Policy Uncertainty Index

※Indexes the frequency of policy uncertainty terms in major newspapers



Global Pandemic Uncertainty Index

※Pandemic or contagion and global uncertainty Frequency of related mentions



備考：EIU によるカントリーレポートにおいて、｢パンデミック (pandemics)｣ または ｢伝染 (epidemics)｣ の語の近くで世界の不確実性に言及する頻度 (10万語中)
資料：IMF " Global Uncertainty Related to Coronavirus at Record High" https://blogs.imf.org/2020/04/04/global-uncertainty-related-to-coronavirus-at-record-high/

Source: Ministry of Economy, Trade and Industry 2020 Monozukuri White Paper

 6

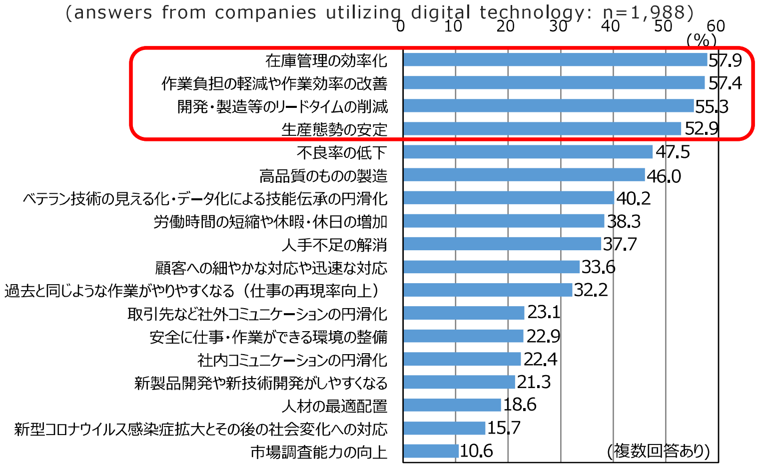
MITSUBISHI ELECTRIC
Changes for the Better

Japan Manufacturing Challenges

Equipment related to the production process
Is data being collected such as operation status?

出展：経済産業省 2020年版ものづくり白書

Aim of utilizing digital technology
(answers from companies utilizing digital technology: n=1,988)



出展：経済産業省 2021年版ものづくり白書

Not as advanced as expected

Improving operational efficiency and reducing lead time

Investing resources in new businesses and business model transformation
Uncertainty increases due to the influence of corona, and resource investment
⇒Necessity of promoting digitalization to strengthen corporate transformation capabilities

 7

MITSUBISHI ELECTRIC
Changes for the Better

Digital is a weapon for corporate transformation

Three Competencies for Dynamic Capabilities

① Sensing threats and opportunities
② Seize opportunities to rebuild and recombine resources,
Gain a competitive advantage
③ Making your competitive advantage sustainable
Transform your entire organization

Enhanced by digitalization

- Data collection and collaboration
- Prediction and prediction by AI
- 3D design and simulation
  Accelerate product development
- Variants
- Flexible process changes

From the Ministry of Economy, Trade and Industry's "2020 Monozukuri White Paper"
Note: Proposed by Professor David J. Thiess, UC Berkeley Business School

Transforming the business model will lead to a competitive advantage in manufacturing.
Digitalization is necessary to increase the capacity to do so

 8

MITSUBISHI ELECTRIC
Changes for the Better

Manufacturing site issues to be solved

Challenges on the shop floor

labor	1	Labor shortage, skill inheritance, work-life balance	productivity	2	Prompt and flexible response to demand fluctuations and troubles
Social needs	3	Carbon Neutrality, Security	predominance	4	Design capabilities through front-loading

By process Breakdown	Design/Development	manufacture	Operation/Maintenance
	■ Strengthening design capabilities ■ Remote Design	■ Reduced cycle times ■ Flexible manufacturing ■ Energy Saving	■ A factory that never stops ■ troubleshooting

Automation of human work by replacing robots and automatic machines and devices
More automation by replacing human judgment with AI

 9

MITSUBISHI ELECTRIC
Changes for the Better

Digitally optimized manufacturing lifecycle

Design **Front-loading reduces rework**
Using a 3D simulator, advance mechanical, electrical and control programsEnabling verification in a digital space. Reduced rework and adjustment time

Manufacture **Speeding up the PDCA cycle by utilizing data**
Operational monitoring, data analysis using AI, diagnosis of abnormalities in equipment and facilities
Speeding up the improvement cycle at the manufacturing site

Maintenance **Reduce costs through AI and remote use**
AI-based maintenance of equipment and facilities, remote from remote. Reduced maintenance costs through work support, etc.

e-F@ctory

Design ▶ Manufacture ▶ O & M

Feedback to Design

Design and simulation in digital space

Reduction of rework during commissioning

Utilization of IT systems
Worldwide surveillance, remote

Data delivery feedback

Leveraging Data at the Edge Layer
EDGECROSS ×Visualization, Analysis, Diagnosis

Data delivery feedback

IoT support / AI utilization of equipment and facilities

CC-Link IE TSN
Maisart

 10

e-F@ctory Technology

Connect, process, and control all data in the field

Integration of field-based technology and digital technology

FA-IT linkage technology	• Technology that easily connects production sites and IT systems • Primary processing of data in real time, Technology for shaping meaningful data	 [Edge Computing Products]
Control technology	• Robot technology required for automation • Comprehensive control of sensor devices, Technology to collect necessary data from production sites	 [PLC]  [Industrial Robots]
Industrial Networking Technology	• High-speed large amounts of data on the production floor Network technology for efficient collection • Fusion of FA and IT, mixing technologies of different networks	CC-Link IE TSN [Industrial Networks]

 11

Cooperative Strategies in e-F@ctory

Accelerating collaboration and innovation

e-F@ctory Alliance Co-creation with more than 1,000 global alliance partners and organizations

e-F@ctory Concept presentation


MELSEC Q series

②Alliance

EDGECROSS ③open platform

Collaborated with startups Acceleration Program ④open innovation

2000 2003 2005 2007 2011 2014 2017 2018 2019

CC-Link → CC-Link IE TSN

①Open Network

Maisart Mitsubishi Electric AI technology announcement

 12

MITSUBISHI ELECTRIC *Changes for the Better*

Record and analyze operating status to reduce downtime

"Record the whole data"when trouble occurs

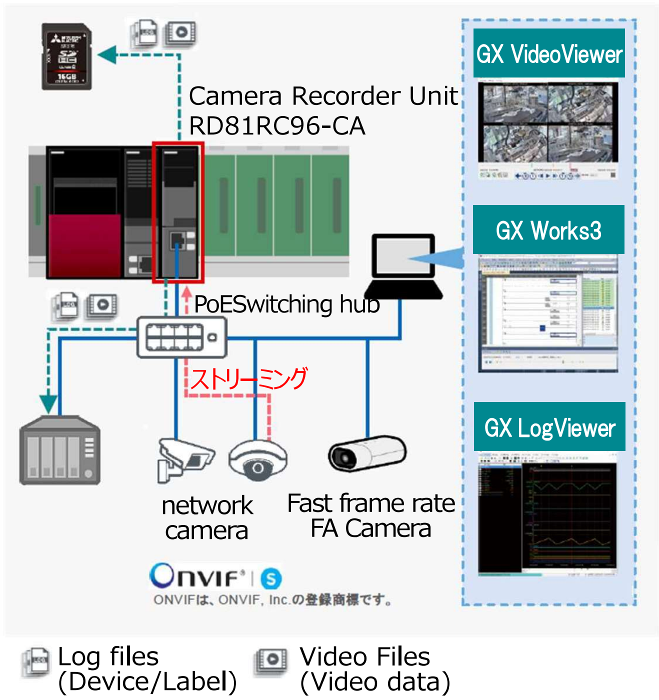
When an abnormality occurs, control data of multiple devices and devices, Record the entire workpiece status and operator operation history

"Easy analysis" with camera image and device confirmation

Control data and camera images collected by recording the whole Match and check to quickly identify the cause of anomalies

Can be connected to a wide variety of network cameras

In addition to MELCO network cameras, a wide variety of easy connection to third-party ONVIF compatible cameras



 13

MITSUBISHI ELECTRIC *Changes for the Better*

AI for manufacturing Realize data utilization without specialized knowledge

AI-powered programming-less analysis
Deep learning, multiple regression analysis, and many other AI capabilities, including
Using machine learning methods and statistical analysis methods, various
Can be used for applications. In addition, since it does not require programming,
Easily implement data analysis solutions.

Achieve functions in one SW
Phase of analyzing production site data in the office, analysis results. In order to realize the phase of on-site diagnosis in one SW,The learning model obtained from data analysis can be applied directly to the production site It is possible.

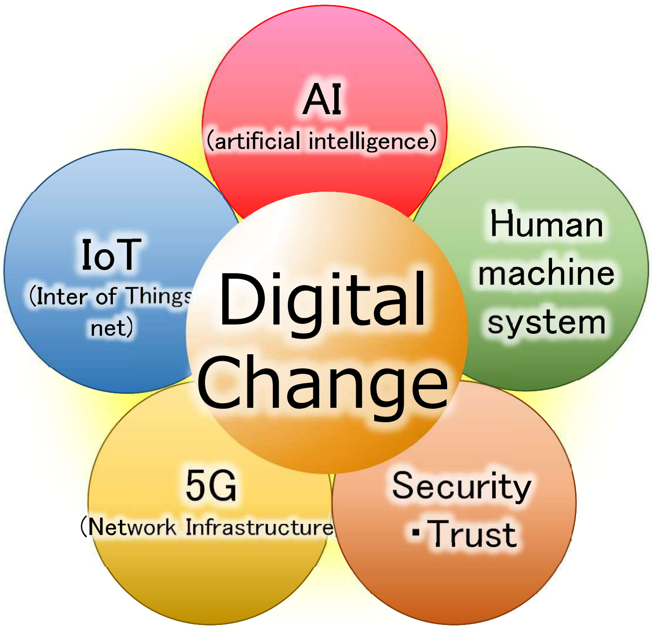
Provision of data analysis services
Data analysts who are familiar with FA will support your data analysis. We provide "Data Analysis Support Service" to analyze customer data at our company, and "Data Analysis Training" to lecture on data analysis know-how.

Data Science Tools MELSOFT MaiLab

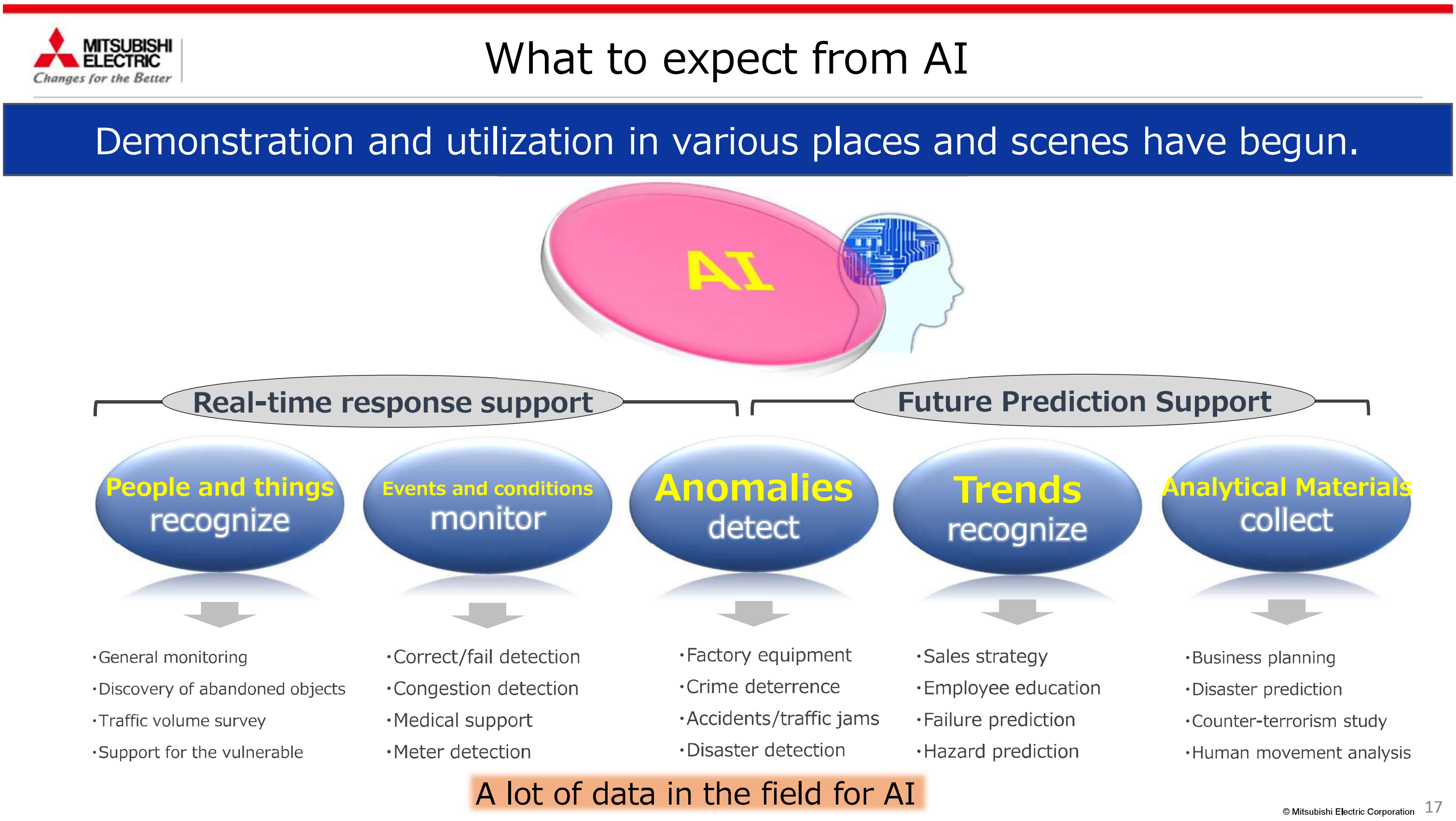
 14

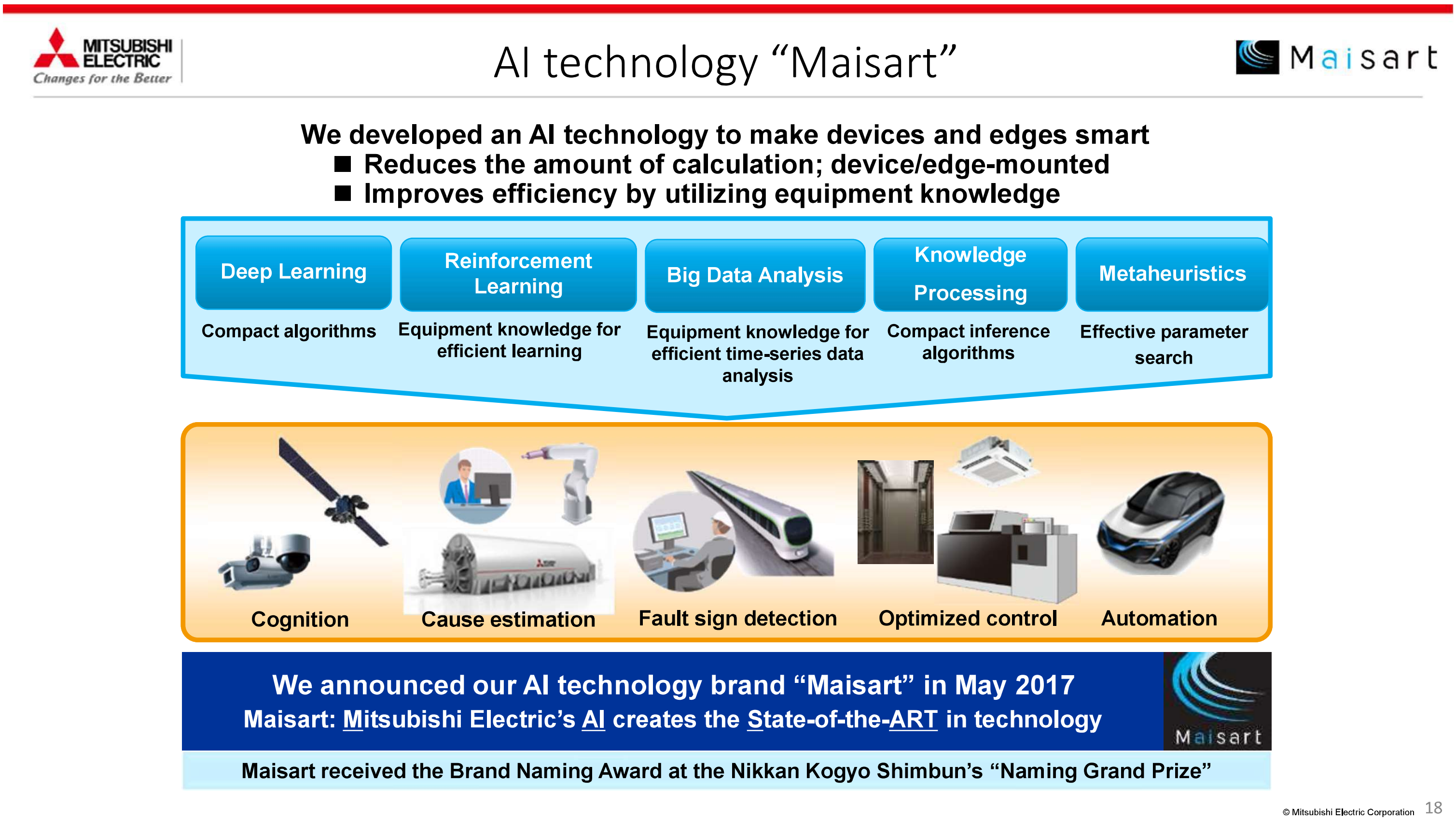
MITSUBISHI ELECTRIC
Changes for the Better

Major digital technologies required for on-site transformation



 16


What to expect from AI
Demonstration and utilization in various places and scenes have begun.
AI
Real-time response support
Future Prediction Support
People and things recognize
Events and conditions monitor
Anomalies detect
Trends recognize
Analytical Materials collect
・General monitoring
・Discovery of abandoned objects
・Traffic volume survey
・Support for the vulnerable
・Correct/fail detection
・Congestion detection
・Medical support
・Meter detection
・Factory equipment
・Crime deterrence
・Accidents/traffic jams
・Disaster detection
・Sales strategy
・Employee education
・Failure prediction
・Hazard prediction
・Business planning
・Disaster prediction
・Counter-terrorism study
・Human movement analysis
A lot of data in the field for AI
© Mitsubishi Electric Corporation 17

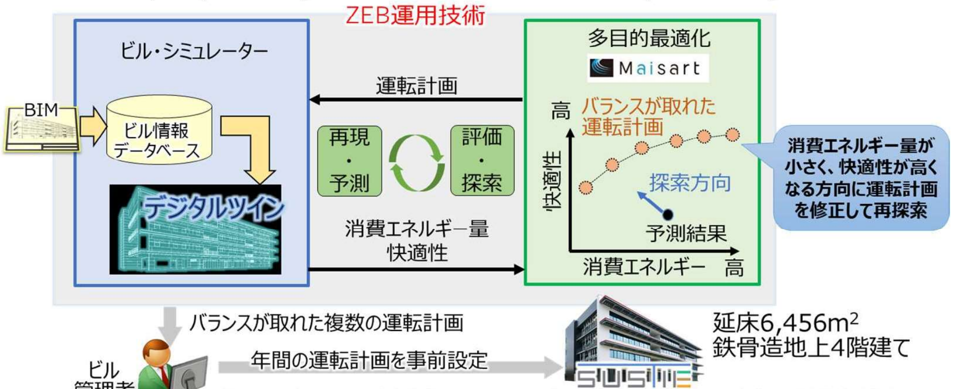

AI technology “Maisart”
Maisart
We developed an AI technology to make devices and edges smart
■ Reduces the amount of calculation; device/edge-mounted
■ Improves efficiency by utilizing equipment knowledge
Deep Learning
Compact algorithms
Reinforcement Learning
Equipment knowledge for efficient learning
Big Data Analysis
Equipment knowledge for efficient time-series data analysis
Knowledge Processing
Compact inference algorithms
Metaheuristics
Effective parameter search
Cognition
Cause estimation
Fault sign detection
Optimized control
Automation
We announced our AI technology brand “Maisart” in May 2017
Maisart: Mitsubishi Electric’s AI creates the State-of-the-ART in technology
Maisart received the Brand Naming Award at the Nikkan Kogyo Shimbun’s “Naming Grand Prize”
© Mitsubishi Electric Corporation 18

Development of AI-based ZEB operation technology

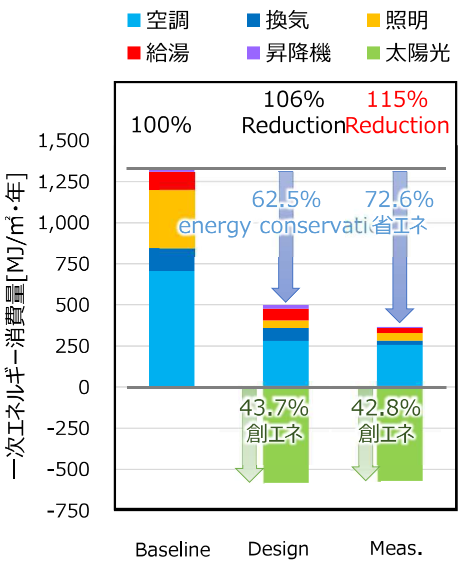
ZEB achieved in ZEB-related technology demonstration building "SUSTIE"

✓World's highest level of medium-sized office building BEI=－0.06
✓Achieved operational "ZEB"
✓Annual energy balance compared to base ▲115%

＜Features of ZEB operation technology＞
❶Building simulator using digital twin
❷Multi-purpose optimization to derive optimal operation



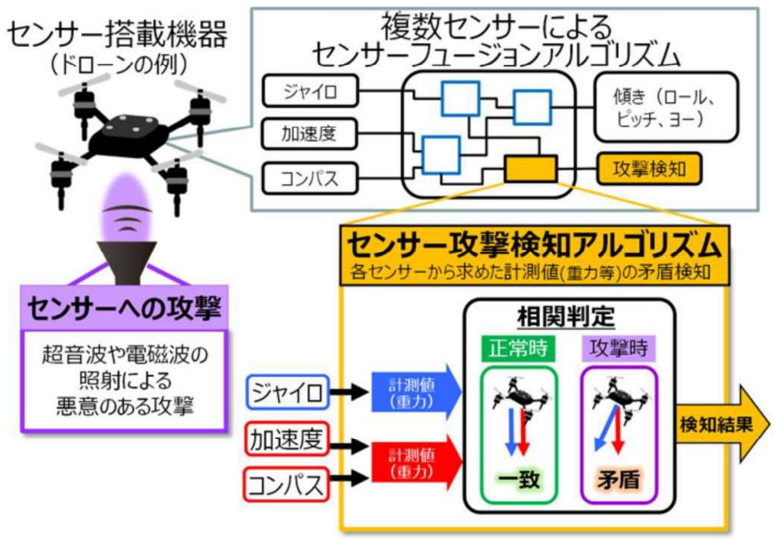
Pre-planned ZEB operation introduced in SUSTIE



 19

Detect sensor attacks with high accuracy "Sensor Security Technology"

✓ Sensor fusion algorithm implements a unique sensor attack detection algorithm
✓ "Sensor security technology" that detects discrepancies in measurement data with high accuracy



Sensor Security Technology
Application examples to drones

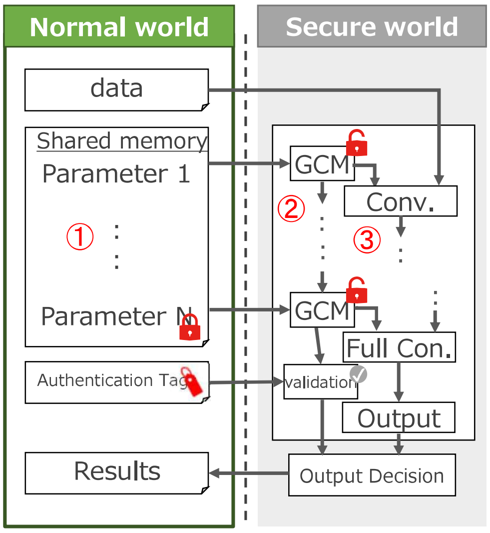
 20

MITSUBISHI ELECTRIC Changes for the Better

AI Security Technology in Embedded Systems

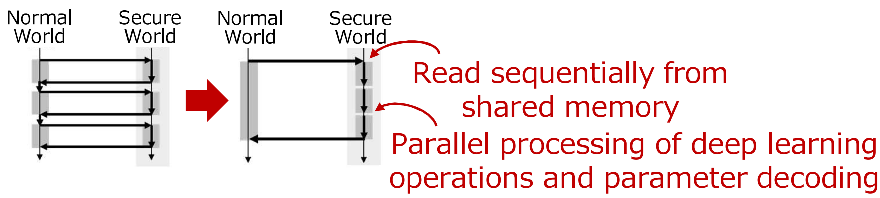
✓Future security for embedded devices enables secure execution of applications methods for utilizing TEEs (*1) are beginning to be considered.
✓Issues when using TEE are (1) insufficient memory, (2) high execution time overhead,(3) Develop a method that can solve the problem that parameter tampering is not detected

The proposal method



① Memory saving by dividing parameters into layers and units of operation and expanding memory

② Parameters are read sequentially from shared memory in split units, and the execution environment Reduced changeover, reducing processing time overhead



③ GCM(Galois/Counter Mode) makes parameter secrecy and tamper detection

(※1)Trusted Execution Environment

This research and development was supported by JST, Future Society Creation Project, and JPMJMI19B6.

 21

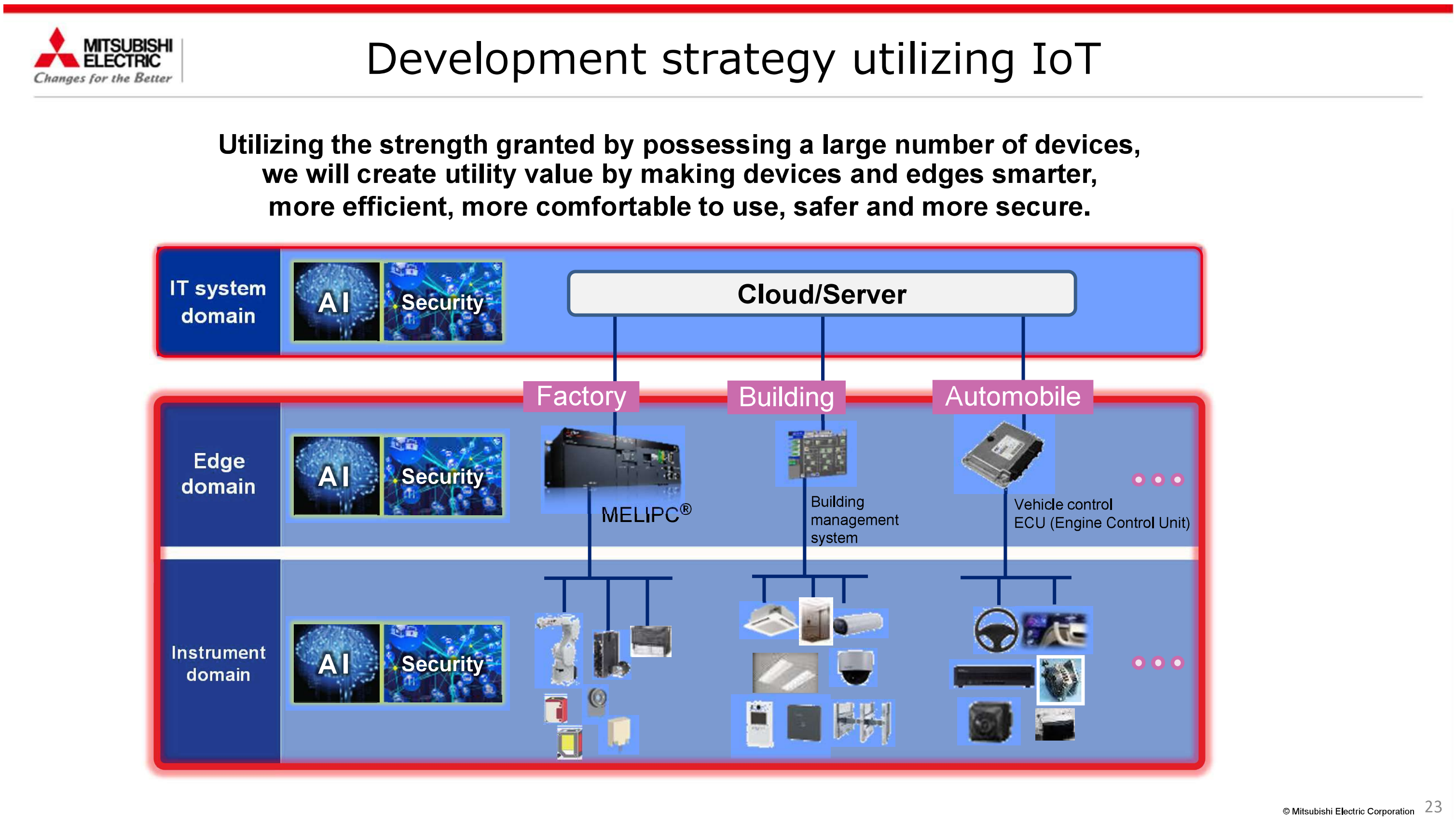
MITSUBISHI ELECTRIC Changes for the Better

Integrated IoT「ClariSense」

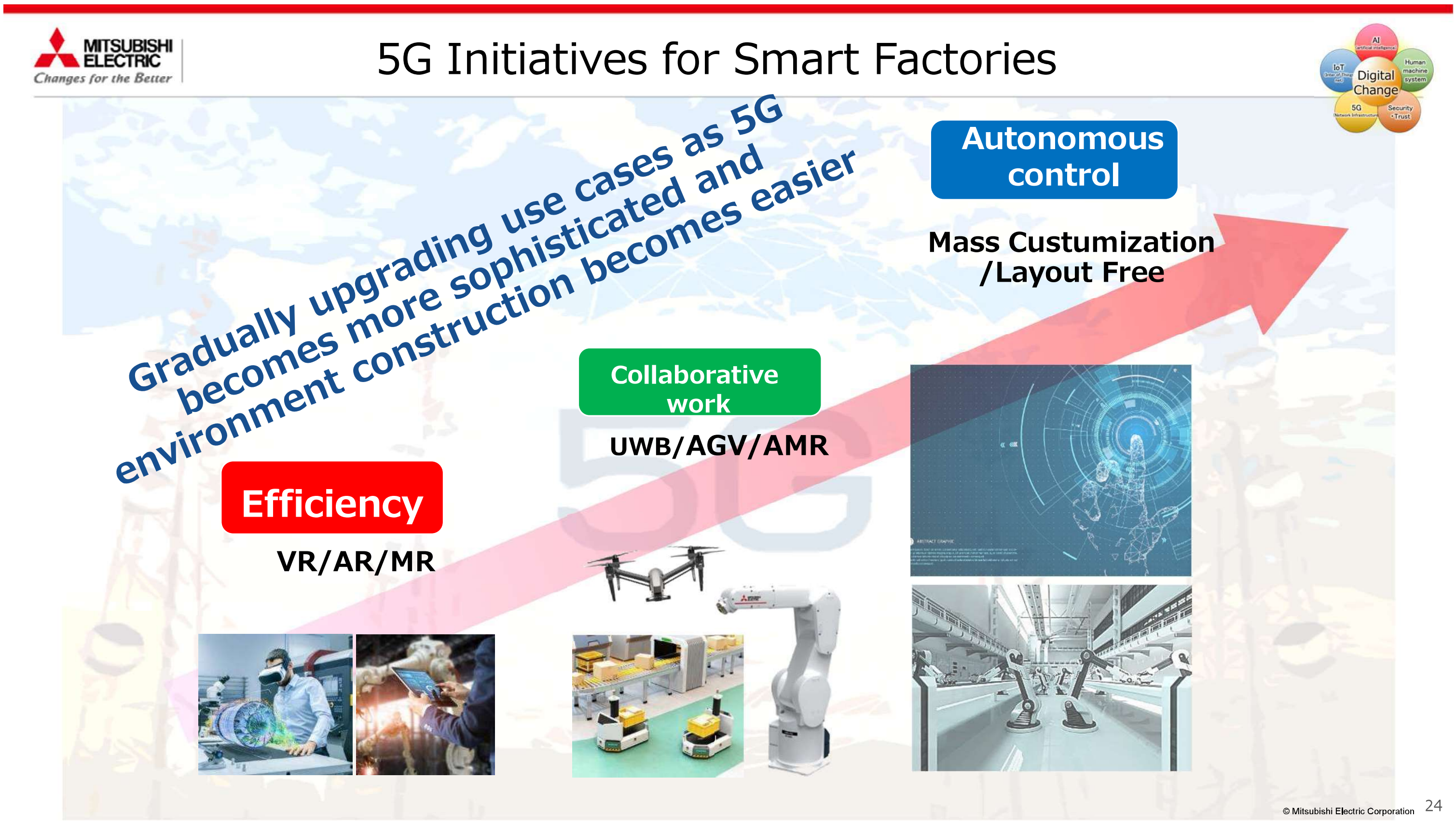
✓ Knowledge of various devices, technologies such as Maisart, security, integrated into IoT system unified design guidelines and solution libraries (*1)
ClariSense centrally organizes core technologies and enables rapid development of IoT system solutions

*1 Reuse common functions for solution Packaged in a possible form

 22



Development strategy utilizing IoT
Utilizing the strength granted by possessing a large number of devices,
we will create utility value by making devices and edges smarter,
more efficient, more comfortable to use, safer and more secure.
IT system domain
AI
Security
Cloud/Server
Factory
Building
Automobile
Edge domain
MELIPC®
Building management system
Vehicle control ECU (Engine Control Unit)
Instrument domain
© Mitsubishi Electric Corporation 23

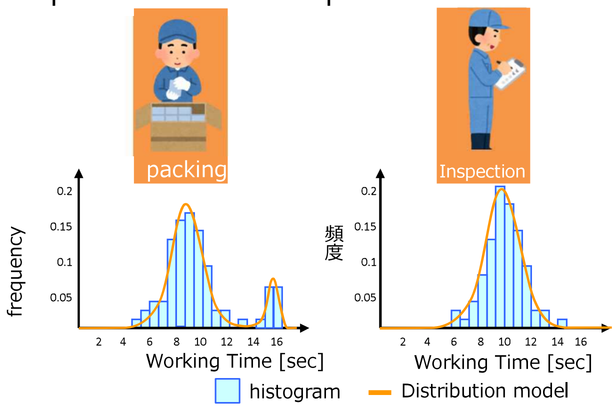


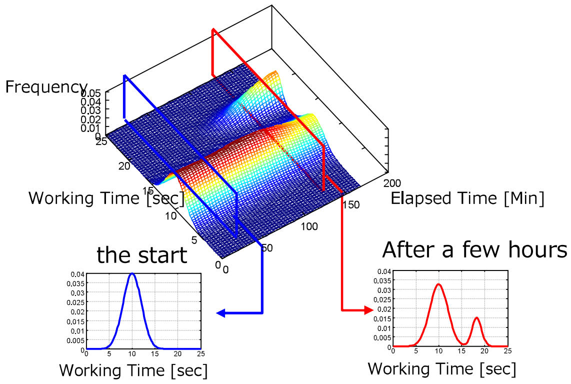
5G Initiatives for Smart Factories
Gradually upgrading use cases as 5G becomes more sophisticated and environment construction becomes easier
Autonomous control
Mass Custumization /Layout Free
Collaborative work
UWB/AGV/AMR
Efficiency
VR/AR/MR
Digital Change
© Mitsubishi Electric Corporation 24

Production forecasting

Forecasting production based on workers and work hours

Predict worker fluctuations and improve the accuracy of estimates of how much production has improved





Variations depending on what you're working on

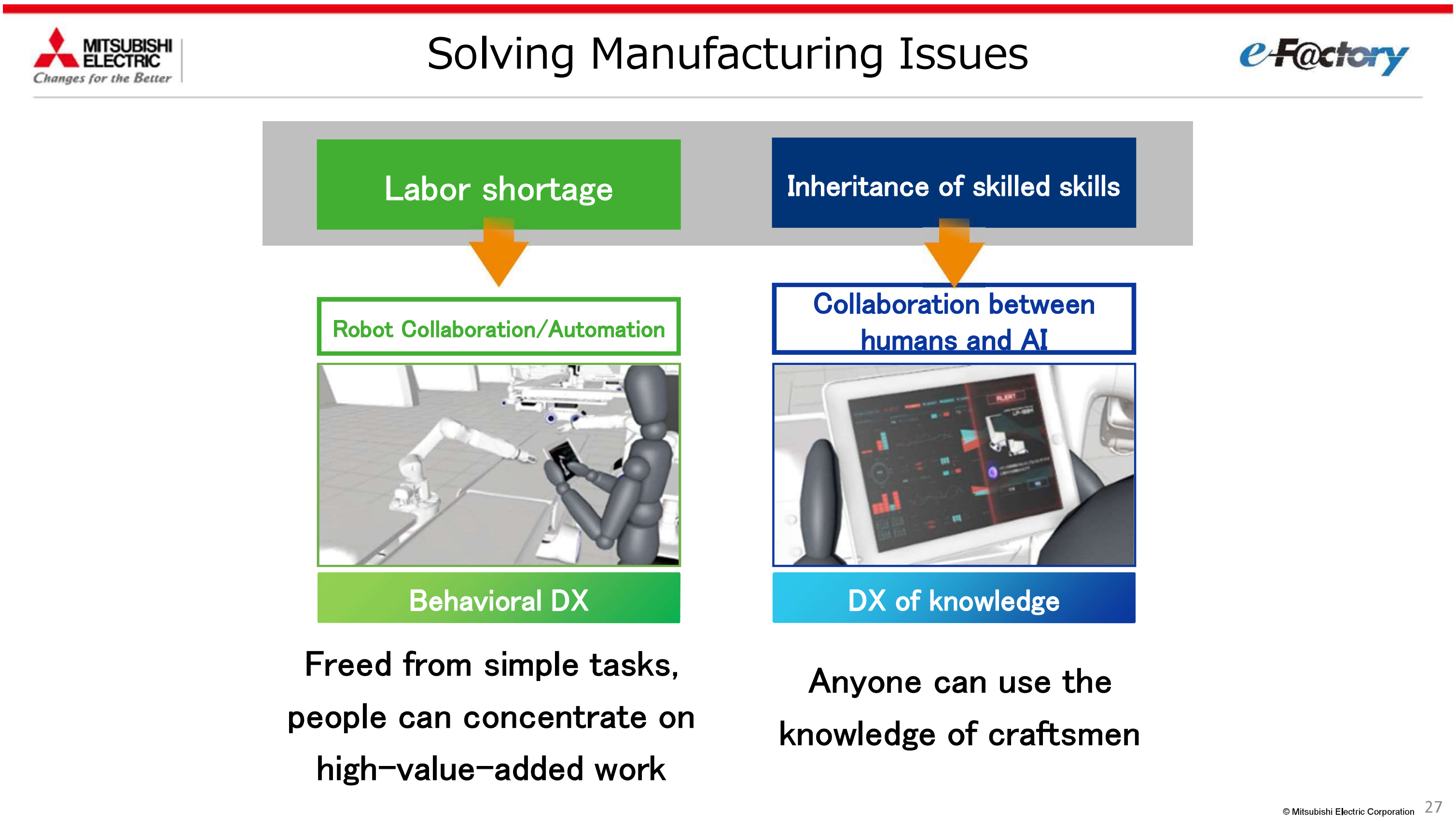
Variation depending on the time of day

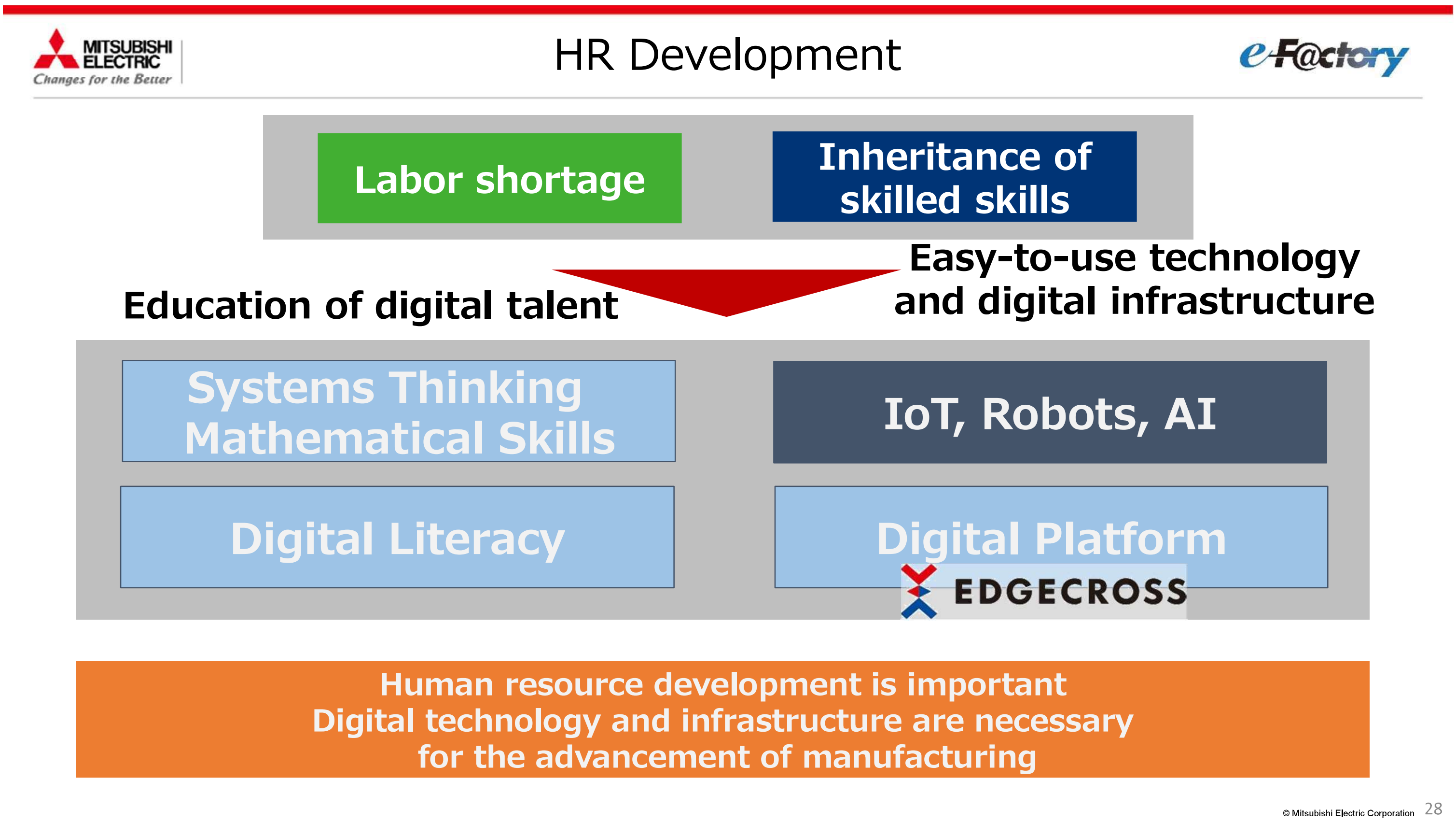
⇒ Learn variability in work time and generate probabilistic models

 25

chapter 03

AI development for the transformation of manufacturing sites


MITSUBISHI ELECTRIC
Changes for the Better
Solving Manufacturing Issues
e-F@ctory
Labor shortage
Inheritance of skilled skills
Robot Collaboration/Automation
Collaboration between humans and AI
Behavioral DX
DX of knowledge
Freed from simple tasks, people can concentrate on high-value-added work
Anyone can use the knowledge of craftsmen
© Mitsubishi Electric Corporation
27

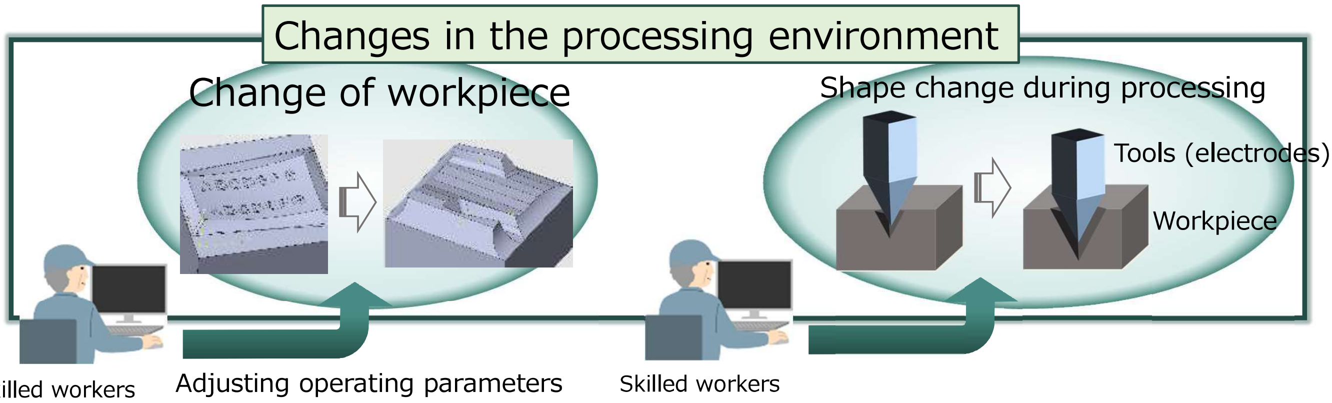

MITSUBISHI ELECTRIC
Changes for the Better
HR Development
e-F@ctory
Labor shortage
Inheritance of skilled skills
Education of digital talent
Easy-to-use technology and digital infrastructure
Systems Thinking Mathematical Skills
IoT, Robots, AI
Digital Literacy
Digital Platform
EDGECROSS
Human resource development is important
Digital technology and infrastructure are necessary
for the advancement of manufacturing
© Mitsubishi Electric Corporation
28

MITSUBISHI ELECTRIC Changes for the Better

Challenges on the shop floor

Maisart

variable-type and variable-volume production is required



Issues Productivity in coordination work decreases due to shortage of skilled workers

Strategy: Adapt to changes in the processing environment in real time and improve productivity even in variable-type and variable-volume production

 29

MITSUBISHI ELECTRIC Changes for the Better

Development Target

The machining environment constantly changes during machining
Difficult to keep the operating parameters optimal at all times

AI control technology that can adapt to changes in the machining environment in real time

High reliability

Real-time AI control requires high reliability・・・

Index the reliability of inference results and appropriately control FA equipment

Fast inference

FA equipment requires high-cycle control・・・

Development of AI capable of high-speed inference in parallel with FA equipment control

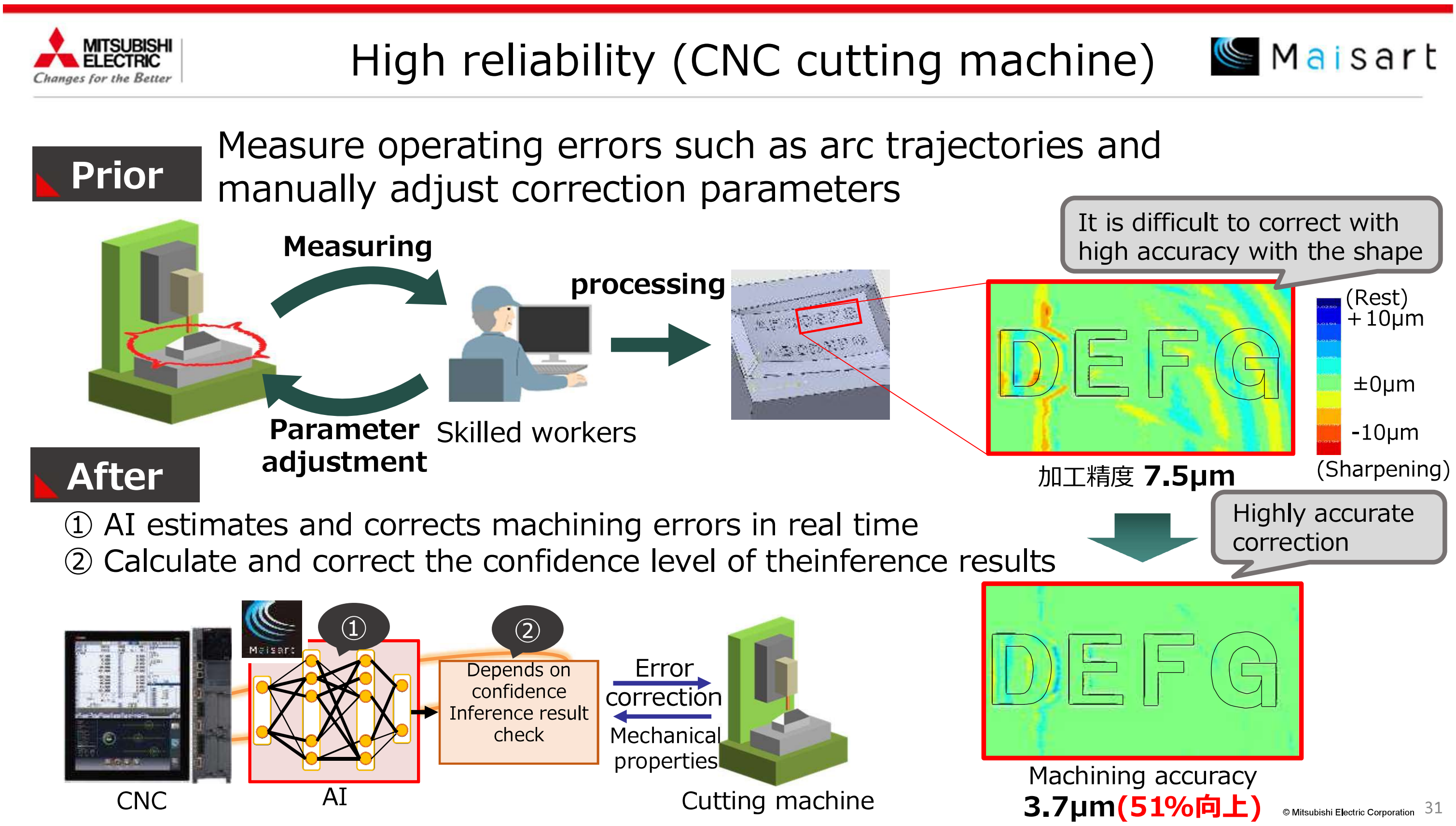
Environmental adaptation

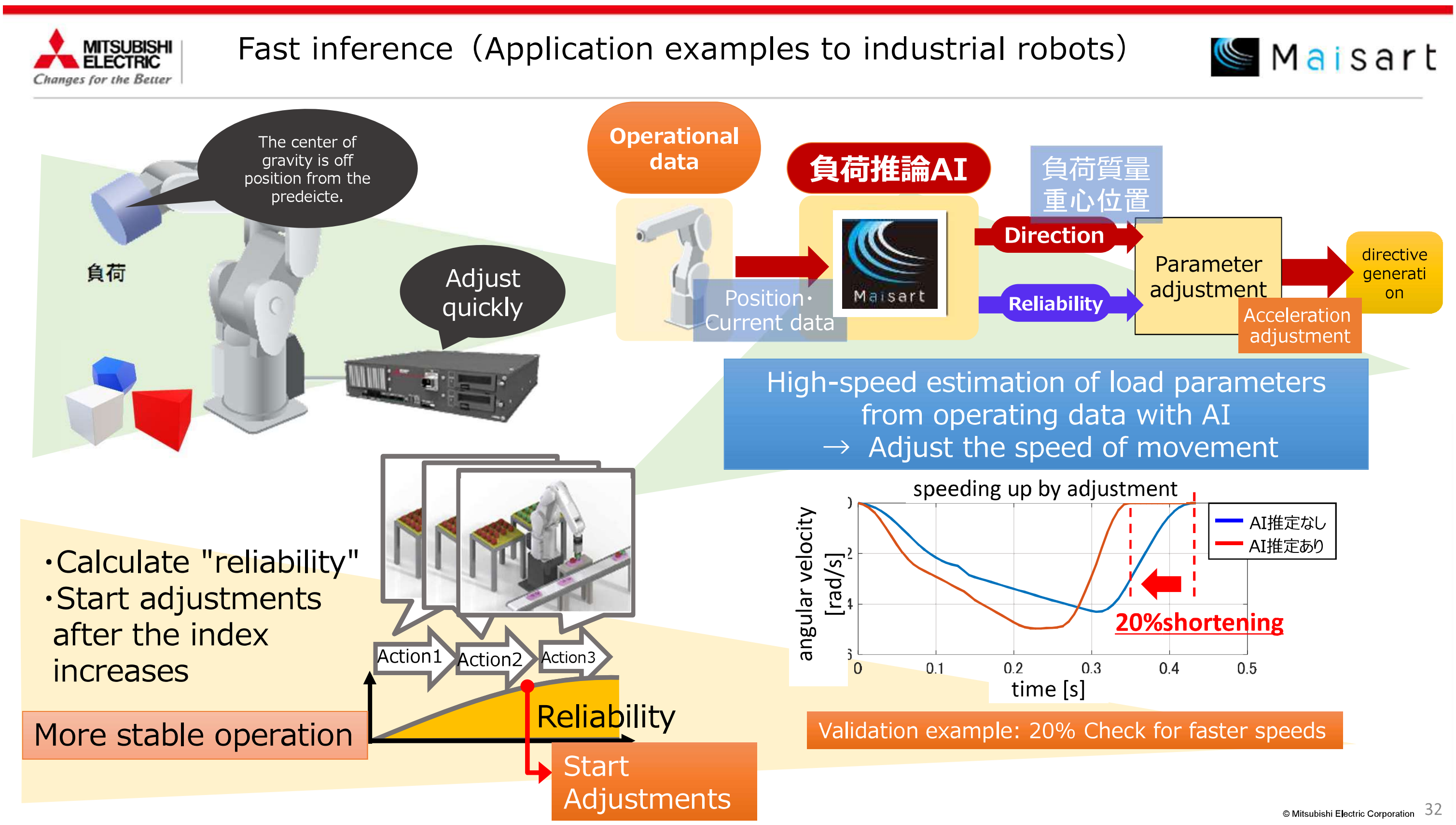
The machining environment changes during machining・・・

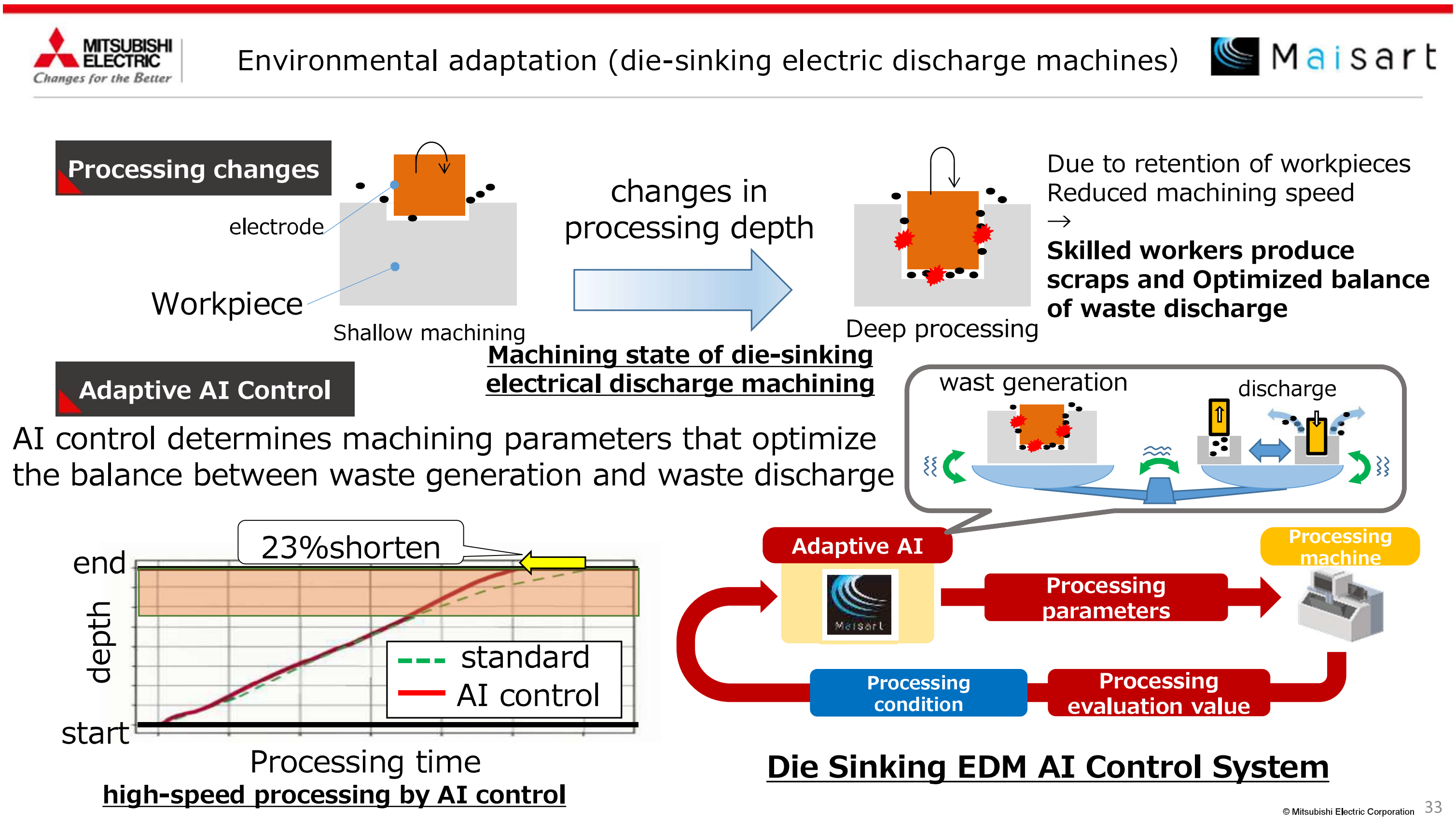
Learns changes in the machining environment during operation and adapts to environmental changes

Joint research with the National Institute of Advanced Industrial Science and Technology

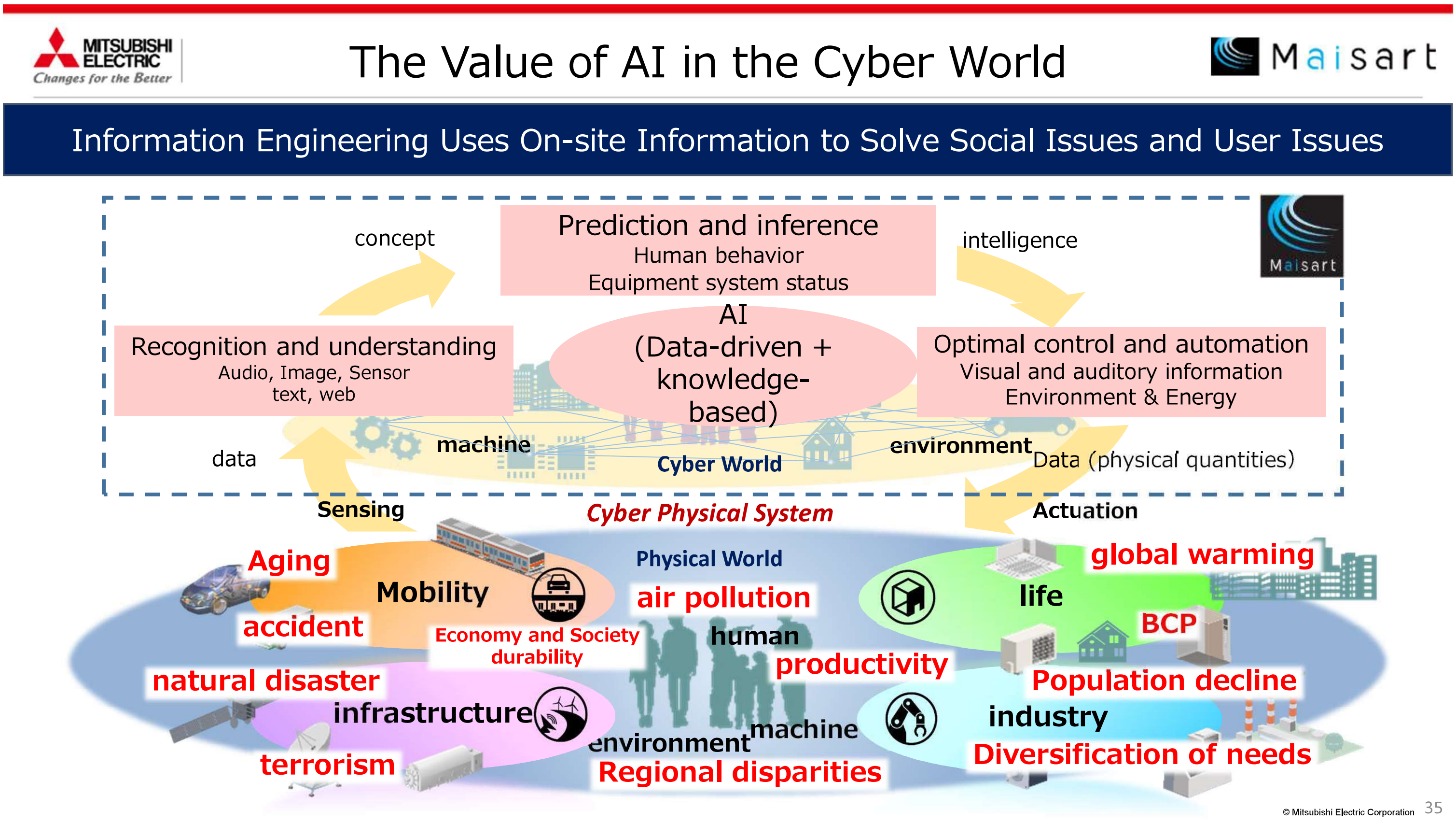
 30


MITSUBISHI ELECTRIC
Changes for the Better
High reliability (CNC cutting machine)
Maisart
Prior
Measure operating errors such as arc trajectories and manually adjust correction parameters
Measuring
processing
Parameter adjustment
Skilled workers
It is difficult to correct with high accuracy with the shape
(Rest)
+10μm
±0μm
-10μm
(Sharpening)
加工精度 7.5μm
After
① AI estimates and corrects machining errors in real time
② Calculate and correct the confidence level of theinference results
Highly accurate correction
Depends on confidence Inference result check
Error correction
Mechanical properties
CNC
AI
Cutting machine
Machining accuracy
3.7μm(51%向上)
© Mitsubishi Electric Corporation 31

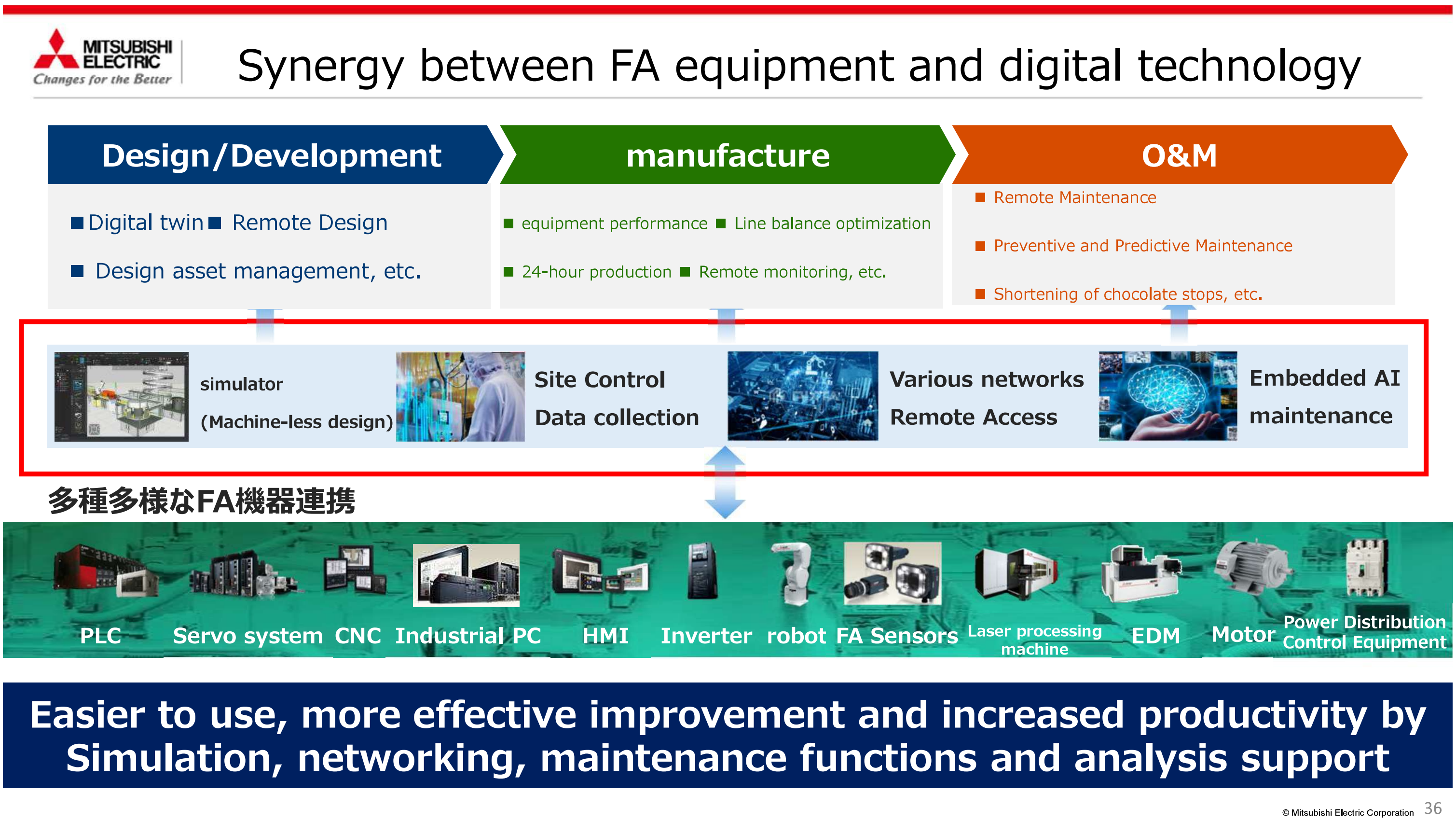

MITSUBISHI ELECTRIC
Changes for the Better
Fast inference（Application examples to industrial robots）
Maisart
The center of gravity is off position from the predeicte.
負荷
Adjust quickly
Operational data
負荷推論AI
負荷質量
重心位置
Direction
Reliability
Position・Current data
Parameter adjustment
directive generation
Acceleration adjustment
High-speed estimation of load parameters from operating data with AI
→ Adjust the speed of movement
speeding up by adjustment
angular velocity [rad/s]
time [s]
AI推定なし
AI推定あり
20%shortening
・Calculate "reliability"
・Start adjustments after the index increases
Action1
Action2
Action3
Reliability
More stable operation
Start Adjustments
Validation example: 20% Check for faster speeds
© Mitsubishi Electric Corporation 32


MITSUBISHI ELECTRIC
Changes for the Better
Environmental adaptation (die-sinking electric discharge machines)
Maisart
Processing changes
electrode
Workpiece
Shallow machining
changes in processing depth
Deep processing
Due to retention of workpieces Reduced machining speed
→
Skilled workers produce scraps and Optimized balance of waste discharge
Machining state of die-sinking electrical discharge machining
Adaptive AI Control
AI control determines machining parameters that optimize the balance between waste generation and waste discharge
wast generation
discharge
23%shorten
end
depth
start
standard
AI control
Processing time
high-speed processing by AI control
Adaptive AI
Processing parameters
Processing machine
Processing condition
Processing evaluation value
Die Sinking EDM AI Control System
© Mitsubishi Electric Corporation
33


chapter 04
Summary



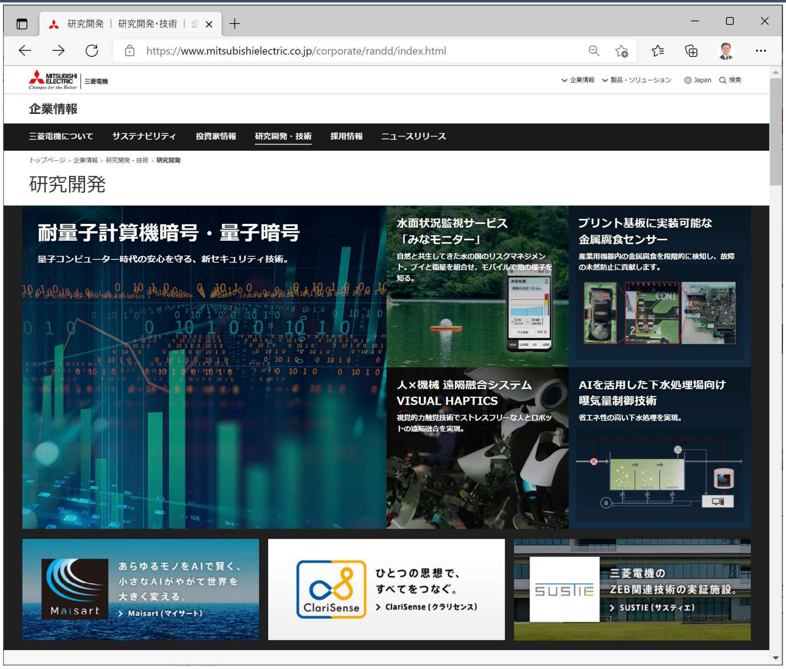
MITSUBISHI ELECTRIC
Changes for the Better
The Value of AI in the Cyber World
Maisart
Information Engineering Uses On-site Information to Solve Social Issues and User Issues
concept
Prediction and inference
Human behavior
Equipment system status
intelligence
Recognition and understanding
Audio, Image, Sensor
text, web
AI
(Data-driven + knowledge-based)
Optimal control and automation
Visual and auditory information
Environment & Energy
data
machine
Cyber World
environment
Data (physical quantities)
Sensing
Cyber Physical System
Actuation
Aging
Mobility
Physical World
air pollution
global warming
life
accident
Economy and Society durability
human
BCP
productivity
natural disaster
Population decline
infrastructure
industry
machine
environment
terrorism
Regional disparities
Diversification of needs
© Mitsubishi Electric Corporation 35



MITSUBISHI ELECTRIC
Changes for the Better
Synergy between FA equipment and digital technology
Design/Development
manufacture
O&M
■ Digital twin ■ Remote Design
■ Design asset management, etc.
■ equipment performance ■ Line balance optimization
■ 24-hour production ■ Remote monitoring, etc.
■ Remote Maintenance
■ Preventive and Predictive Maintenance
■ Shortening of chocolate stops, etc.
simulator
(Machine-less design)
Site Control
Data collection
Various networks
Remote Access
Embedded AI
maintenance
多種多様なFA機器連携
PLC
Servo system
CNC
Industrial PC
HMI
Inverter
robot
FA Sensors
Laser processing machine
EDM
Motor
Power Distribution Control Equipment
Easier to use, more effective improvement and increased productivity by Simulation, networking, maintenance functions and analysis support
© Mitsubishi Electric Corporation 36

Maisart、ClariSense、SUSTIE Web site

Features and examples of Mitsubishi Electric's digital technology



Web Access

「Mitsubishi Electric」
「research and development」

検索

https://www.mitsubishielectric.co.jp/corporate/randd/index.html

 37

MITSUBISHI ELECTRIC
Changes for the Better

・This manual was prepared based on public information, etc. obtained by the Company, but the accuracy and reliability of the manual is guaranteed by the Company. Not In addition, this document does not guarantee future predictions
・We are not responsible for any damage caused in connection with this manual.
・We are not responsible for any lawsuits from third parties that are not the original distribution destination of this book, and we cannot respond to any inquiries.

Session 2:
Intelligent Transportation Systems
(Chair: Kozo Okano)

Relieving Traffic Congestion Stably by Controlling Amount of Vehicles Detouring from Congested Roads

Arisa Kitano†, Takuya Yoshihiro‡

†Graduate School of Systems Engineering, Wakayama University, Japan
‡Faculty of Systems Engineering, Wakayama University, Japan
†e-mail: s236086@wakayama-u.ac.jp
‡e-mail: tac@wakayama-u.ac.jp

Abstract - Automobile traffic jam causes serious problems such as loss of time for drivers on congested roads, traffic accidents, and environmental pollution. Therefore, traffic jam is a problem that needs to be solved. One method to solving these problems is encouraging vehicles to detour around congested roads. In this method, the system detects the occurrence of traffic jams and suggests the drivers a route to detour around the congested road segments. This approach reduces the number of vehicles entering the congested road, thereby relieving the congestion. However, when congestion is relieved by detours, the system terminates the detour or rapidly reduces the number of detours, which increases the number of vehicles entering the road, and causes re-congestion. In this study, as an extension to the conventional methods of detouring vehicles to reduce traffic jams, we propose a new method that balances the number of vehicles entering and exiting a congested road. This proposed method improves the efficiency of the entire traffic network and prevents repeated traffic jams. We implemented the proposed method in a traffic flow simulator and evaluated its performance. The results showed that the average vehicle travel time reduces with the proposed method, indicating that the proposed method controls road traffic more efficiently during traffic jam resolution.

Keywords: automobile, traffic jam, detour route guidance, ITS(Intelligent Transport Systems), traffic control

1 INTRODUCTION

Automobile traffic jam causes problems such as loss of time for drivers on congested roads, traffic accidents, and environmental pollution. Therefore, traffic jam is a problem that needs to be solved. Traffic jam occurs when the traffic demand on a road exceeds the traffic capacity. The causes of traffic congestion may be an increment in traffic demand due to commuting to and from work and school or a reduction in traffic capacity due to traffic accidents or lane restrictions for road construction. One approach to solving such problems is encouraging vehicles to detour around congested roads. In this method, the system detects the occurrence of traffic jams and suggests the drivers a route to detour around the detected road segments. This approach reduces the number of vehicles entering the congested road, thereby relieving the congestion. However, this method has problems. This method temporarily reduces the number of vehicles entering a congestion detected road segment. Then, when the congestion eases, the system terminates the rerouting or reduces the amount of rerouting. However, this approach causes re-traffic congestion when the number of vehicles entering the road exceeds the traffic capacity.Due to mitigate this adverse effects, it is essential to resolve traffic congestion with stable traffic condition transitions, not only when the system detect a traffic jam but also after the congestion relief.

In this study, we expand the conventional methods of resolving traffic jams by detouring vehicles and propose a new method that balances the number of vehicles entering a congested road and the number of vehicles exiting the road. This proposed method improves the efficiency of the entire traffic network from the occurrence of traffic jam to its resolution. Specifically, we propose a method of controlling the number of vehicles that detour around congested roads. The basic idea is to stabilize the change in traffic volume on congested roads when relieving congestion by keeping the number of vehicles entering a congested road at a certain level lower than the number of vehicles exiting the road. This approach provides stable control overall traffic between occurrence of a traffic jam to its resolution and prevents re-traffic congestion and temporary traffic disruptions when the number of vehicles detouring around the congested road is changed. Efficient and stable congestion resolution reduces vehicle travel time due to congestion.

We implemented the proposed method in a traffic flow simulator and evaluated its performance. The results showed that the proposed method was more stable than the conventional method in the process between the occurrence of a traffic jam and its clearing. In addition, the results of the proposed method showed that the average vehicle travel time was lower, indicating that the proposed method made road traffic more efficient during traffic jam resolution.

2 RELATED WORK

Urban Transportation Optimization is one of the major interests in the worlds, and thus a large amount of studies exist in the literature [1]. Two major means to optimize transportation is the traffic signal control [2] and the vehicle route control [3]. The traffic signal control can control all vehicles in the entire network. However, this approach can not reduce traffic on congested roads. In contrast, the route control reduces traffic on congested roads by detouring. Therefore, the route control is more effective in case of congestion because immediate effect can be expected.

Among several existing route control approaches, the route guidance system (RGS) is one of the most promissing one [4]. RGS is basically a reactive system that, once congestion is detected, the system typically computes routes that bypass the congested roads, and guide vehicles to avoid the congestion with those routes. As a considerable amount of traffic avoids the congested routes, the congestion resolves in a short time.

RGS methods are classified into three types, i.e., centralized, distributed, and hybrid ones. In the centralized approach, traffic information is collected by the central server, and the server computes the routes to guide. Several centralized approaches have been proposed [5][12], and several current route map services such as Google Maps [6] and TomTom [7] are implemented as a centralized service. Several fully distributed approach exists, in which vehicles monitors states of roads by themselves, and make their own route decision [8][9]. However, in the distributed approach, it is hard to collect road-state information so that reliable route computation is not possible.

The hybrid approach assumes an infrastructure to collect traffic state information, and the collected information is forwarded to each vehicle, and the vehicle computes their own optimal routes from the obtained information [10][11]. Although those methods sometimes coorporates to distribute traffic to prevent concentration, since essentially the route computation is distributed, the effect inevitably is suboptimal. From the viewpoint of optimization, the central approach achieves the best performance. Although the computational load could be one of the weakpoints, it is still worth exploring the potential of it.

Matsui et al. focus on detouring vehicles far from congested roads compute [12]. While DSP, RkSP, and EBkSP compute detour routes from near the congested road, Matsui et al. considered the possibility of reducing vehicle arrival time by detouring vehicles from roads farther away from congested roads. The method developed by Matsui et al. calculates multiple detour routes from nearby and distant intersections from the congested roads and preferentially guides vehicles to routes with smaller travel-time increases. In addition, because their method calculates detour routes considering the traffic capacity of the roads on the detour route, it prevents secondary congestion from occurring on the detour route. However, this method has the problem that traffic on congested roads becomes unstable because it rapidly decreases the detouring vehicles when congestion eases.

The above methods fail to stabilize traffic on congested roads because they do not properly account for the number of detouring vehicles. The number of detouring vehicles is closely related to traffic stability. When the number of detouring vehicles is high, traffic congestion rapidly relieves. However, when the congestion is relieved and the detour guidance is stopped, traffic is disrupted due to the sudden increase in the number of vehicles on the congested road, resulting in longer vehicle arrival time. On the other hand, if fewer vehicles take detours, traffic is not disrupted by the change in traffic volume. However, there is a risk of worsening traffic congestion due to less traffic reduction on the road. This results in longer vehicle arrival time. Therefore, it is essential to properly determine the number of detouring vehicles when we resolve traffic jams by vehicle detouring.

These methods terminate all detour guidance when there is no more traffic congestion. As mentioned earlier, this leads to an unstable and inefficient traffic state. A smooth change in the number of detouring vehicles would stabilize the traffic condition, but to the best of our knowledge, no method has been proposed in the literature that controls the volume of detouring vehicle flows to stabilize the whole congestion management process. In this regard, this work is new and novel.

3 PROPOSED METHOD

3.1 Overview of the Proposed Method

We propose a method for calculating the number of vehicles detouring around a congested road segment so that balances the number of vehicles entering and exiting that road segment. When traffic congestion is detected, the proposed method calculates the number of vehicles detouring around the congested road so that the number of vehicles entering the congested road is lower than the number of vehicles exiting that road. This approach stabilizes fluctuation in traffic on congested roads and resolves traffic congestion. Even if the detour guidance relieves the traffic congestion and the congestion is no more detected, our method continues the detour guidance. If the guidance is terminated in this situation, the increase in traffic volume on the congested road causes congestion. Therefore, the detour guidance terminates when the cause of the traffic congestion, such as lane closures, is removed and the number of detouring vehicles is reduced to zero. Therefore, the proposed method efficiently resolves traffic congestion and reduces traffic travel time.

3.2 Assumptions

In this study, we envision the situation where a method of vehicle detour guidance resolves traffic congestion. The method detects a traffic jam when the number of vehicles on a road segment exceeds the threshold value, and guides vehicles to detour routes that avoid the congested road segment to relieve the congestion. In this study, we distinguish between the method of computing detour routes and the method of calculating the number of detouring vehicles, and focus on the latter to stabilize and improve the efficiency of congestion resolution. In the computing of detour routes, it is assumed that the method computes more than a single detour route and that road segment as the starting point of the detour route. It is envisioned that roadside units are installed at various locations on the target road network and the data obtained from these units provide the traffic volume for the entire road network. In addition, we focus on the centralized control system in which a central server computes the detour routes and the number of detouring vehicles for the entire target road network, rather than the distributed control system in which each vehicle computes its detour route. In a centralized control system, roadside units transmit the traffic volume to the central server. If the central server detects a traffic jam based on the received traffic volume, it computes the detour plan that consists of the number of detouring vehicles and their detour routes and transmits the plan to roadside units. Then, the roadside units guide the target vehicles based on the detour plan.

3.3 Notation

We define the inflow I_r^t and the outflow O_r^t as the number of vehicles entering and exiting road segment r at time t, respectively. We also define the detour volume D_r^t as the number of vehicles that are proposed to change from a route through road segment r to a route avoiding road segment r at time t. There is more than one road segment that is the starting point of the detour route. Then, when U is the set of those road segments, and d_u^t for $u \in U$ denotes the number of vehicles detouring to the route that starts at road segment u. Therefore, it can be expressed as $D_r^t = \sum_{u \in U} d_u^t$. In addition, it is assumed that we can predict the time vehicles it takes to travel each road segment. When a vehicle change to the route that detours around the road segment r, $travel_{(u,r)}^t$ denotes a predicted travel time of the route from the road segment u, the starting point of the detouring route, to road segment r before the detour. In this case, the effect of the detour volume D_r^t appears at time $t + travel_{(u,r)}^t$.

When detouring vehicles resolve traffic congestion by detouring vehicles, even after the congestion is no longer detected, a certain amount of detouring must be continued to prevent the recurrence of the congestion and to maintain stable road traffic. Therefore, in this study, the resolution of traffic congestion means that the causes of traffic congestion, such as traffic accidents or traffic restrictions, etc., have been removed, the number of detours has been reduced to zero, and traffic has stabilized, not that the congestion is no longer detected.

3.4 State Transition

The central server assigns each road segment a state that transitions according to the road conditions. Figure 1 shows the state transition diagram. There are three states: normal, congested, and adjusted. The road segment given the normal state has no congestion detected, and no guidance is provided to detour around that road. On the road segment given the congested state, traffic congestion is detected. The detour volume D_r^t is calculated so that the inflow I_r^t is a certain amount lower than the outflow O_r^t, and vehicles planning a route through the congested road segment are guided to a detour route based on the detour amount D_r^t. This approach reduces traffic volume below a certain level without disruption. The adjusted state is assigned to a road segment when traffic congestion is no longer detected during detouring vehicles around that road segment. For the road segment given adjustment state, the detour volume D_r^t is calculated to maintain or gradually reduce the current severity of congestion without worsening it until the congestion is resolved.

If the number of vehicles on a congested road segment drops below the congestion detection threshold due to vehicle detours, terminating detour guidance may cause an increase in traffic volume and congestion again. Thus, even though there is no congestion detected on a road segment r assigned the adjusted status, we continue to guide vehicles with the appropriate detour volume D_r^t until it determines that the congestion is resolved and there is no need for detour guidance.

3.5 Calculating the Amount of Detour

The detour volume D_r^t in the congested and adjusted states is computed as shown in Eq. (1).

$$D_r^t = I_r^t - O_r^t + R_r^t \tag{1}$$

Note that we define the reduction R_r^t as the desired difference between the inflow and the outflow at road segment r, and define the detour reservation E_r^t as the traffic volume scheduled to decrease by recent detour guidance. The calculation of the reduction R_r^t is explained in Sec. 3.6.

$I_r^t - O_r^t + R_r^t$ in Eq. (1) controls the relationship between the inflow and the outflow of the road segment r after detour guidance. When the reduction R_r^t is positive and the detour volume D_r^t is $I_r^t - O_r^t + R_r^t$, the inflow of r is reduced lower than the outflow. In addition, the higher the reduction R_r^t is, the larger the difference between the inflow and the outflow of the road segment r after the detour guidance is provided. As this difference between the inflow and the outflow increases, there is a tradeoff: while congestion is rapidly eliminated by lowering the inflow on congested roads, the overall traffic network becomes unstable due to the large disruption in traffic volume.

Note that simply setting the detour volume D_r^t to $I_r^t - O_r^t + R_r^t$ is not enough to calculate the appropriate detour volume D_r^t, because the effect of detouring takes time to be reflected in the congested road r. For example, if the detour volume $D_r^t = I_r^t - O_r^t + R_r^t$ is made at time t, the time it takes for the detour effect to be reflected in r is equal to the time it takes to reach road segment r from detour road segment u, which is a reasonably long time. In other words, there is a delay before the detour effect is reflected on the congested road. This results in a feedback system with a large delay, which generally causes large oscillations in traffic volume on congested roads and results in unstable traffic.

Based on the above, the detour volume D_r^t is calculated in the congested and adjusted states. To balance the inflow I_r^t and the outflow O_r^t of a congested road segment r, the detour volume D_r^t is calculated by $I_r^t - O_r^t + R_r^t$. Traffic control is stabilized by calculating the detour volume D_r^t by considering the feedback time delay.

3.6 Reduction R_r^t

On a road segment r in *congested state*, the inflow I_r^t must be lower than the outflow O_r^t to relieve the congestion. The reduction R_r^t affects the difference between the inflow and the outflow. Then, the reduction R_r^t is equal to α and controls the difference between the inflow and the outflow to always keep a constant α.

The state of a road segment changes from congested state to *adjusted state* when the congestion eases and the number of vehicles on that road segment becomes less than the congestion detection threshold. The adjusted state is intended to resolve traffic congestion stably by maintaining or gradually

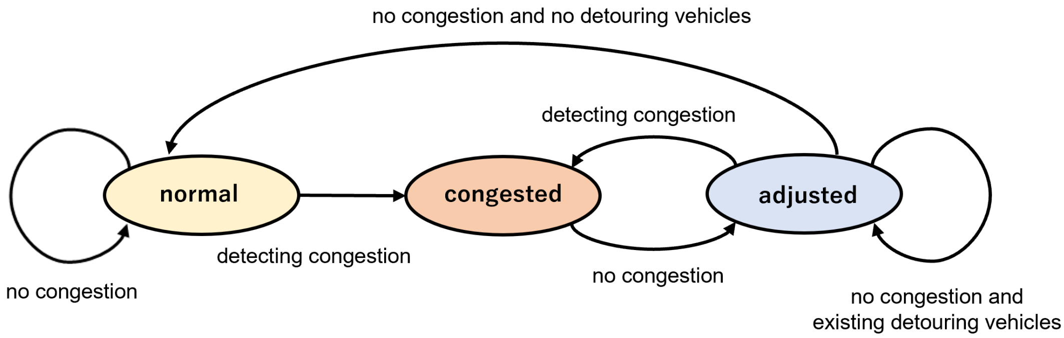


Figure 1: state change diagram

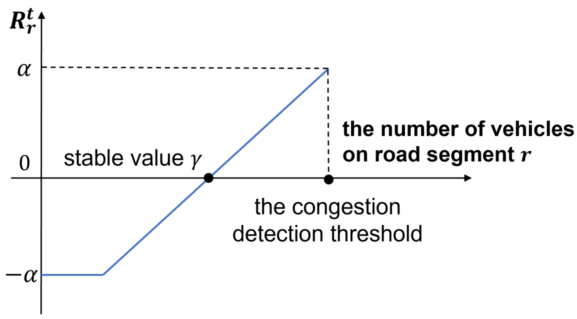


Figure 2: the reduction R_r^t

decreasing the traffic volume so that the traffic volume does not increase and the congestion is not detected again. For this purpose, the reduction R_r^t is adjusted according to the traffic volume on the road segment r. If there is a high probability that congestion is detected again on road segment r, set R_r^t to a positive value to reduce the traffic volume on road segment r by making the inflow on road segment r smaller than the outflow. Conversely, if congestion is easing, set R_r^t to a negative value to increase traffic on r by making the inflow higher than the outflow.

In the adjustment state, the proposed method adjusts the reduction R_r^t according to the number of vehicles on the road segment r used for congestion detection, as shown in Fig. 2. Note that the traffic volume when $R_r^t = 0$ is called the stable value γ, which is set as a parameter value smaller than the congestion detection threshold. Also, α is the same parameter used to calculate the reduction R_r^t in congested. By controlling the reduction R_r^t as shown in Fig. 2, the detour volume D_r^t is adjusted so that the number of vehicles on road segment r is close to the stable value. Then, when the outflow on road segment r increases due to the removal of congestion causes, the detour volume D_r^t decreases to maintain the traffic volume. Finally, the detour volume D_r^t becomes zero and the detour guidance is terminated.

4 EVALUATION

4.1 Method

To confirm that the proposed method reduces the average vehicle travel time, we performed simulations. In our scenario, lane restrictions cause traffic congestion and the proposed method resolves that congestion. By calculating the average travel time for all vehicles, we evaluate that the proposed method can efficiently control not only congested roads but also the entire road network, which is the objective of this study. We implemented Matsui et al.'s method [12] and the proposed method as an extension to it, and compared the performance of the proposed method with that of conventional methods. In the simulation of the proposed method, we use the proposed method for calculating the detour volume, and use Matsui et al.'s method [12] for the congestion detection and the detour route computation. The traffic flow simulator is SUMO (Simulation for Urban MObility) [13], and TraCI (Traffic Control Interface) [14] acquires traffic information and suggests routes for vehicles.

The roadside units acquire traffic volumes on each road at regular intervals, and the central server uses the average of the traffic volumes over the past period of about two traffic-light cycles to compute detour routes and detect traffic congestion because this study assumes the road network with traffic lights. Note that if the central server detects congestion on a road segment with a red light, the congestion may be falsely detected due to an accidental increase in traffic volume. In addition, the detour volume calculated by the proposed method may not be not stable because there is a large difference in traffic volume between roads with green and red traffic signals. Therefore, to avoid such kind of fluctuation on traffic measurement, this simulation calculated the average value for two signal cycles.

4.2 Scenario

In this simulation, we used the grid road map shown in Fig. 3. Each road segment on the road map is a right-of-way with three lanes in each direction, consisting of left-turn and straight-ahead lane, straight-ahead only lane, and right-turn and straight-ahead lane. An enlarged map of the area near the intersections is shown in Fig. 3. Traffic lights is installed at each intersection, and the signal cycle is set to 140 seconds.

Vehicles travel from one end of the road map to the others. For example, the destination of vehicles departing from the bottom of the map is equally assigned to a total of 15 road segments toward the left, top, and right ends of each road network. Therefore, the number of vehicles departing from each

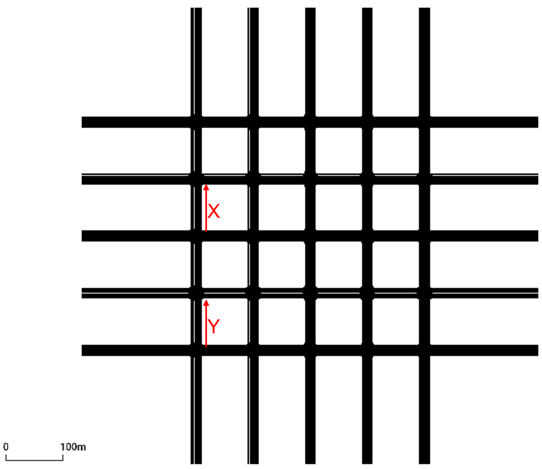


Figure 3: Road maps used in the evaluation

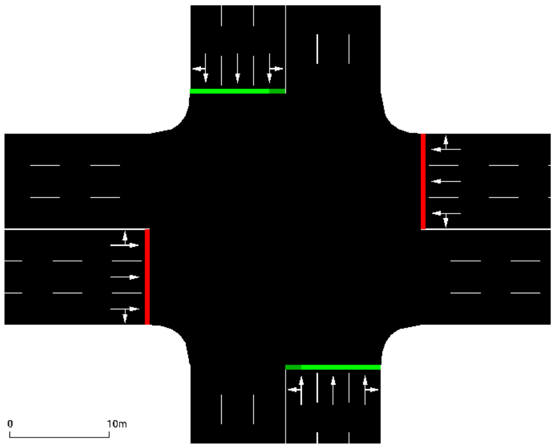


Figure 4: Close-up view of the intersection

road is the same for all roads, and the number of vehicles arriving at each road is the same for all roads. SUMO automatically computes the shortest route for the vehicle using Dijkstra's algorithm [15] based on the origin and destination roads. The weight of each road is its travel time, which is the quotient of the road length and the maximum speed.

In this simulation, we closed two lanes of the busiest road segment to generate traffic congestion. The road segment with the lane closures is shown at point X in Fig. 3. We blocked the two right-hand lanes of the road segment at point X: the straight-ahead lane and the right-turn and straight-ahead lane. Thus, the bottleneck, which is the remaining left lane of the three lanes, causes congestion downstream of point X. Also, the lane restriction prevents vehicles from turning right from point X. During the lane closure, the vehicle planning to turn right from point X re-computes its route when it arrives at the road segment ahead of point X. The start time of the lane closure is set to 3600 seconds after the start of the simulation, and the end time of the lane closure is set to 10800 seconds after the start of the simulation.

The simulation settings are shown in Table 1. The simulation starts at time 0 seconds and ends at time 28800 seconds. Then, we evaluated the vehicles that departed between 1800 and 14400 seconds. We simulated traffic volumes of 4,500 and 5,000 vehicles per hour for the entire road network, respectively. In this simulation, we detect traffic congestion at regular intervals. When congestion is detected, a detour guidance plan until the next congestion detection time is calculated. The congestion detection and traffic volume acquisition are performed every 20 seconds, and the period of the traffic light is 140 seconds. We set the threshold for traffic congestion detection at 0.5 based on the traffic flow model called Greenshields Model [16]. This means that congestion is detected when the number of vehicles on a road is more than half of the maximum number of vehicles on that road. The expansion rate of congestion, a parameter of the comparison method, is set to 1.3 based on Matsui et al.'s paper [12]. The actual probability of following the route guidance is 0.7, i.e., 30 % of vehicles does not follow the route guidance of the system and do not change to follow their own route, and the estimated probability of following the route guidance, described in Sec. 3.5, is 0.8. For both the proposed method and the comparison method, the actual number of guided vehicles is the calculated detour volume divided by this estimated value. As the settings in the proposed method, the stable value γ is set at 0.375, and we simulated the cases where α is 3, 5, 7, 9, 11, and 13, respectively.

Table 1: Simulation setting

Traffic simulator	SUMO [13]
Simulation start time	Time 0s
Simulation end time	Time 28800s
Lane closure time	Time 3600s to 10800s
Vehicle used for evaluation	Vehicle departing time 1800s to 14400s
Traffic volume	4500/h, 5000/h
Signal period	140s
Congestion detection cycle	20s
Traffic acquisition cycle	20s
Congestion detection threshold	0.5
Expansion rate of congestion	1.3
Estimated probability of following route guidance β	0.8
Actual probability of following route guidance	0.7
α	3, 5, 7, 9, 11, 13
stable value γ	0.375

4.3 Results

We calculated the average travel time for all vehicles departing between 1800 and 14400 seconds from the beginning of the simulation. The results for a traffic volume of 4500 vehicles per hour are shown in Fig. 5, and that of 5000 vehicles per hour in Fig. 6. The parameter α of the proposed method is on the horizontal axis. Since the conventional method does not use the parameter α, its average travel time is the same value regardless of the horizontal axis.

Figures 5 and 6 show that the average vehicle travel time

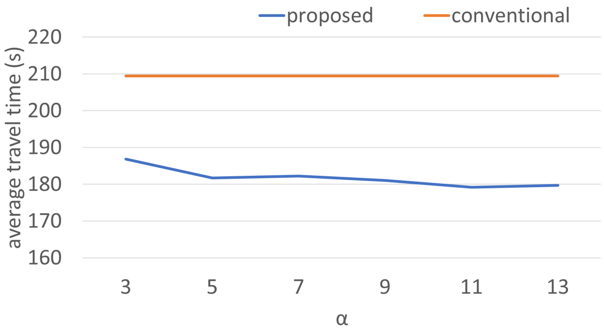


Figure 5: Average travel time for all vehicles (4500 vehicles / hour)

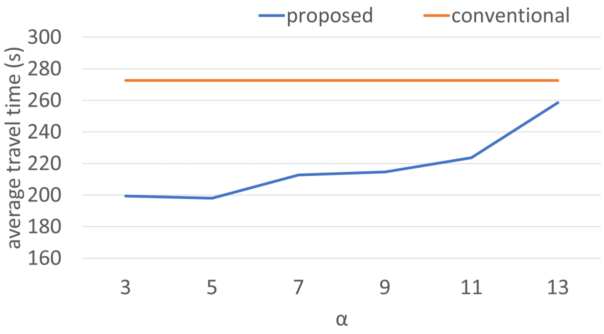


Figure 6: Average travel time for all vehicles (5000 vehicles / hour)

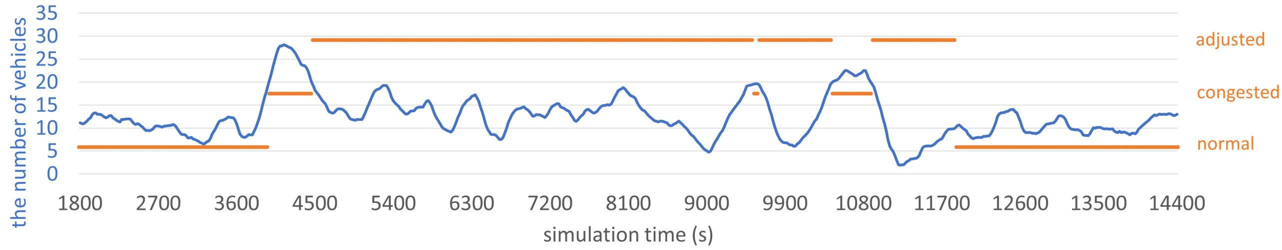
of the proposed method is lower than that of the conventional method. The difference in average travel time between the proposed and the compared methods is about 30 seconds when the traffic volume is 4500 vehicles per hour and about 70 seconds when the traffic volume is 5000 vehicles per hour. Clearly, it can be said that the proposed method reduces the average vehicle travel time. The proposed method performs stable traffic control by maintaining a balance between the number of vehicles entering and exiting congested roads. The results showed that this approach reduces the average travel time of vehicles.

In addition, Fig. 5 shows that the performance is better when $alpha$ is high, while Fig. 6 shows that performance is better when α is low. We suppose that α is related to the fluctuation in traffic volume on congested roads. The larger α is, the larger the difference between the inflow and the outflow on a congested road, and the more rapidly congestion is relieved. On the other hand, the greater the difference between the inflow and the outflow on a congested road, the more turbulent the traffic on that road is. Therefore, when the traffic volume of the entire road network is low, the large α rapidly relieves the traffic congestion and the average travel time becomes low. Conversely, when the traffic volume of the entire road network is high, the large α disrupts the traffic, and the average travel time becomes high.

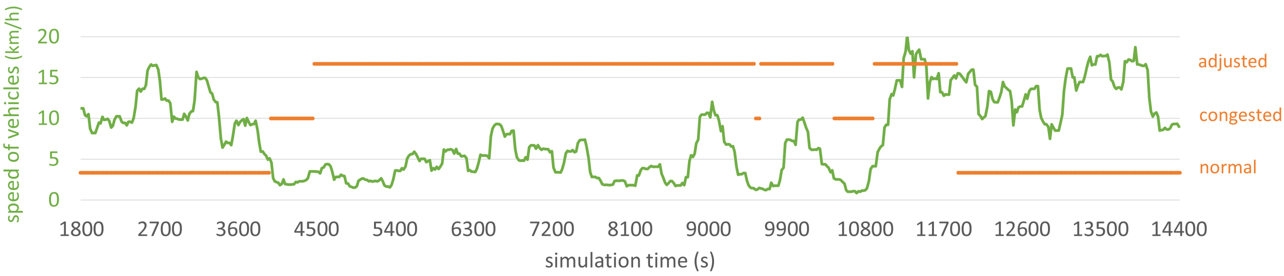
To confirm that the proposed method improves the efficiency of traffic control on congested roads, we consider in detail the road Y in Fig. 3 with a traffic volume of 5000 vehicles per hour and $\alpha = 13$. Road Y is the road where congestion was detected for the longest time. The transition of road Y for the proposed and comparative methods are shown in Figs. 7 and 8, respectively. These figures show the road condition, the average number of vehicles, and the average speed of vehicles on road Y. The number of vehicles and the vehicle speeds are averaged over two signal cycles (140 seconds) to eliminate oscillations in values due to traffic signals. The conventional method does not define the road state. Thus, we regard that the congestion state as the condition in which traffic congestion is detected detour guidance is appled, and the normal state as the condition in which no traffic congestion is detected, and no guidance appled as well. Then, the definition of the congestion state is the same for the proposed and conventional methods since they use the same congestion detection method. In this simulation, traffic congestion is detected based on the number of vehicles. Therefore, when the congested state is assigned to a road, the number of vehicles on that road is exceeding the certain level.

Figures 7 and 8 show that both the proposed and conventional methods detected traffic congestion at around 3600 seconds, which is the start time of lane restrictions, and resolved it at around 10800 seconds, which is the end time of the lane restrictions. Due to the closure of the two right lanes of Road X, a queue of vehicles extends to make the upstram Road Y congested as well. On the other hand, results during the period of lane restrictions showed the differences in performance between the proposed and conventional methods. The results of the conventional methods in Fig. 8 shows that the road state and the average number of vehicles on the road keeps considerably high value. During lane restrictions, detour guidance must actually be continued because traffic capacity on that road is reduced due to the downstream lane restriction. However, the conventional method transits to the normal state and quit detouring when the number of vehicle reduces, which again increases the number of vehicles, leading to the repeated fluctuation of traffic. Compared to the conventional method, Fig 7 shows that in the proposed method, the road state is stabilized and the number of vehicles on the road significantly reduced by controlling the detour volume to maintain a balance between the inflow and outflow volumes on the congested roads. It also shows that the speed of the vehicle in the proposed method is higher than that of the conventional method. The 45 % increase in average vehicle speed with the proposed method reduces the travel time by about 30 seconds per vehicle.

The above results shows that stabilizing traffic on congested roads reduces the average travel time of vehicles in the vehicle route guidance method for resolving traffic congestion. Stabilizing traffic on congested roads prevents traffic disruptions on roads around that congested road. Therefore, the proposed method stabilizes the control of the entire road network and reduces the average travel time for all vehicles. In addition, since the proposed method stabilizes traffic better than the compared method, it suggests that maintaining a balance between the inflow and outflow on congested roads is an effective way to stabilize traffic.

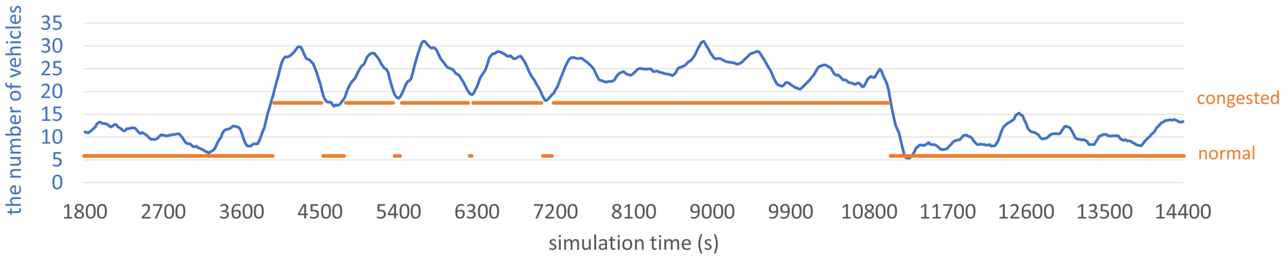


(a) Transition in road state and average number of vehicles



(b) Transition in road state and average speed of vehicles

Figure 7: Simulation of proposed method (traffic volume: 4500 vehicles per hour)



(a) Transition in road state and average number of vehicles

(b) Transition in road state and average speed of vehicles

Figure 8: Simulation of conventional method (traffic volume: 4500 vehicles per hour)

5 CONCLUSION

In this study as an extention to the route guidance method to resolve congestion, we proposed a method for controlling the number of detouring vehicles to stabilize the fluctuation of traffic volume on congested roads. By maintaining the balance between the number of vehicles entering and exiting the congested road, the traffic volume on the congested road is stabilized. This approach stabilizes the traffic control from the occurrence of traffic congestion to its resolution, thereby reducing the average travel time for all vehicles.

We evaluated the proposed method using a traffic flow simulator, SUMO [13]. In our scenario, Lane restrictions on a grid road map leads to traffic congestion, and a route guidance method resolves the traffic congestion. As a route guidance method, we used the method of Matsui et al. [12], and compared the cases where the proposed method is applied and not applied. The results showed that the proposed method reduces the average travel time for all vehicles compared to the conventional method. The results showed that the proposed method As a future issue, it is necessary to improve the calculation method of the reduction R^t_r in the proposed method. The proposed method stably resolves traffic congestion by maintaining a balance between the number of vehicles entering and exiting congested roads. However, in the proposed method for calculating the reduction R^t_r, the appropriate value of α depends on the traffic volume of the entire road network. In addition, the simulation scenarios in this paper are rather simple scenarios for evaluationg realistic of road conditions. Therefore, we will evaluate the proposed method by preparing scenarios similar to the road network to be realized in the future.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 23H03385.

REFERENCES

[1] C. Lin, K.L. Choy, G.T.S. Ho, S.H. Chung, H.Y.Lam, “Survey of Green Vehicle Routing Problem: Past and Future Trends,” Expert Systems with Applications, Vol.41, pp.1118-1138, 2014.

[2] I. Tomar, S. Indu, and N. Pandey. “Traffic Signal Control Methods: Current Status, Challenges, and Emerging Trends,” Proceedings of Data Analytics and Management: ICDAM2021, Vol.1, pp.151-163, 2022.

[3] J. Pan,I. Sandu Popa,K. Zeitouni,and C. Borcea.: Proactive vehicular traffic re-routing for lower travel time, IEEE Transactions on Vehicular Technology, Vol.62, No.8 (2013).

[4] J.M. Sparmann, “Benefits of Dynamic Route Guidance Systems as Part of a future Oriented City Traffic Management System,” Vehicle Navigation and Information Systems Conference 1991, pp.839–847, 1991.

[5] J. Pan, I. Sandu Popa, K. Zeitouni, and C. Borcea, “Proactive Vehicular Traffic Re-routing for Lower Travel Time,” IEEE Transactions on Vehicular Technology, Vol.62, No.8, 2013.

[6] http:// www.google.com/mobile/

[7] http://www.tomtom.com/en_gb/products/mobile-navigation/

[8] A.M. de Souza and L.A. Villas, “A Fully-distributed Traffic Management System to Improve the Overall Traffic Efficiency,” In Proceedings of the 19th ACM International Conference on Modeling, Analysis, and Smulatino of Wireless and Mobile Systems (MSWiM2016), pp.19-26.

[9] Z. Khan, P. Fan, S. Fang, and F. Abbas, “An Unsupervised Cluster-based VANET Oriented Evolving Graph (CVoEG) Model and Associated Reliable Routing Scheme,” in IEEE Transactions on Intelligent Transportation Systems, Vol.20, No.10, pp.3844-3859, 2019.

[10] J. Pan, I.S. Popa and C. Borcea, “DIVERT: A Distributed Vehicular Traffic Re-routing System for Congestion Avoidance,” in IEEE Transactions on Mobile Computing, Vol.16, No.1, pp.58-72, 2017.

[11] B. Liu, J. Long, M. Deng, X. Yang, Y. Shi, “An Adaptive Route Planning Method of Connected Vehicles for Improving the Transport Efficiency,” ISPRS International Journal of Geo-Information, 11(1):39, 2022.

[12] Y. Matsui, T. Yoshihiro,: A Route Guidance Method for Vehicles to Improve Driver ’s Experienced Delay against Traffic Congestion, 18th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp 375–389 (2021).

[13] P.A. Lopez, M. Behrisch, L. Bieker, J. Erdmann, Y. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner.: Microscopic Traffic Simulation using SUMO, 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2575-2582 (2018).

[14] A. Wegener, M. Piorkowski, M. Raya, H. Hellbruck, S. Fischer, and J.P. Hubaux.: TraCI: An Interface for Coupling Road Traffic and Network Simulators, Processing of the 11th communications and networking simulation symposium, pp.155-163 (2008).

[15] E.W. Dijkstra.; A note on two problems in connexion with graphs, Numerische Mathematik, Vol. 1, pp. 269-271 (1959).

[16] B.D. Greenshields.: A Study of Traffic Capacity, Proceedings of the highway research board, Vol. 14, pp. 448-477 (1935).

A Transfer Scheduling Method for ride-sharing Services to Alleviate Traffic Congestion

Shion Suzuki[†], Takuya Yoshihiro[‡]

[†]Graduate School of Systems Engineering, Wakayama University, Japan
[‡]Faculty of Systems Engineering, Wakayama University, Japan
{s246135, tac}@wakayama-u.ac.jp

Abstract - Traffic congestion causes various problems such as lost traffic time and increased exhaust emissions. One solution that has gained attention for alleviating traffic congestion is ride-sharing services. By allowing users to share rides with other users, the cost of transportation per person can be reduced. Replacing traditional private vehicle transport with ride-sharing services can decrease the number of vehicles on the roads, contributing to the resolution of traffic congestion. However, as ride-sharing services become more popular, new transportation demands may arise, leading to changes in traffic patterns and the emergence of congestion due to different factors than those observed at present. One such factor is the increased traffic around destination areas due to the concentration of ride-share vehicles. The aim of this study is to mitigate traffic congestion caused by this increased traffic. The proposed method adds a process to existing ride-sharing scheduling methods, generating transfers in response to congestion. In the case of congestion on a particular road, ride-share vehicles sharing congested routes within the planned route are instructed to transfer during specific time intervals. This transfer allows users already using the ride-sharing service to board vehicles carrying other users, thereby promoting ride-sharing. By facilitating multiple transfers, the number of vehicles heading to the destination is reduced, mitigating congestion caused by vehicle concentration.

Keywords: Intelligent Transportation System (ITS), Ride-sharing, Transfer, Traffic Congestion, Mobility as a Service (MaaS)

1 INTRODUCTION

Traffic congestion is a significant issue in recent years. Traffic congestion is a pressing issue that leads to problems such as fuel loss and the generation of emissions. The effectiveness of ride-sharing services in alleviating traffic congestion has garnered attention. Ride-sharing are transportation services that, like taxis, provide users with transportation from their pickup location to their destination. The main difference from traditional taxis in ride-sharing services is the option to share the ride with other users. Compared to traveling by private vehicle, the use of a ride-sharing service can be expected to reduce traffic volume by allowing multiple people to travel in a single vehicle. In addition to reducing traffic congestion by decreasing overall traffic volume, ride-sharing services offer various other benefits. For users, benefits include reduced fuel costs through ride-sharing, as well as savings in taxes, insurance, and maintenance expenses associated with individual ownership.

In the near future, it is predicted that the widespread of ride-sharing will result in a majority of people traveling to their destinations using ride-sharing services without owning a private vehicle. Furthermore, with the widespread of electric vehicles and autonomous driving technology, it becomes possible to operate ride-sharing services at lower costs due to reductions in labor and fuel expenses. Ride-sharing services will be used not only for travel transportation but also for daily commuting. Since the cost of owning and maintaining a private vehicle will be eliminated, people's expenditures for transportation will be significantly reduced.

Ride-sharing, which is expected to alleviate traffic congestion by reducing the number of vehicles on the road, is anticipated to bring about changes in the current state of transportation. This leads to congestion caused by factors different from the current ones.One of the causes is concentration of ride-sharing vehicles. Demand for transportation to a specific location in a short period of time is likely to be concentrated at popular sightseeing spots on holidays, at event venues that attract many visitors, and at stations during daily commuting to school and work, etc. When ride-sharing services are extensively used, there is a concern that serious traffic congestion may occur in such a busy destination, leading to increased crowding on the roads. This study addresses this issue by encouraging sharing rides among users when traffic congestion occurs. By having more people ride in a single vehicle, the traffic volume around the destination is reduced, contributing to the alleviation of congestion. To the best of our knowledge, method has been proposed that improves traffic congestion in ride-sharing services by transferring ride-sharing vehicles.

2 RELATED WORK

2.1 General ride-sharing

This section describes related work on general ride-sharing services. Ma et al. [1] proposed a method for minimizing the travel distance of vehicles. To increase the profit of the ride-sharing service operator, they compute all combinations of the picking up and dropping off sequences and assign the schedule that minimizes the travel distance to the request. It is confirmed that the proposed method reduces the total distance traveled by the cab by 11% and the price per user by 7% compared to the case without ride-sharing. Okoso et al [2] proposed a route optimization method that prioritizes routes where requests are expected to occur when selecting routes, while allowing users to travel with fewer detours. In an environment where the number of vehicles is insufficient for the demand, this method was found to increase the number of acceptable reservations, suppressing the vehicle travel distance and user waiting time, compared to the conventional method that does not consider other requests when selecting routes.

2.2 Successive Best Insertion (SBI) Method

Noda et al. [3] proposed a vehicle dispatching algorithm for responding to user's transportation demand, called the Successive Best Insertion method. The SBI method is a ride-sharing scheduling algorithm based on a greedy approach. For each request, the SBI method calculates the optimal route and schedules the picking up and dropping off from the current location to the destination. The following shows the operational steps when a user requests a vehicle until a vehicle is dispatched.

1. When the user gets to the boarding location, he/she submits a dispatch request to the system. Each vehicle maintains a stop-point list as a list of the picking-up and dropping-off points for the allocated requests.
2. The system calculates a new stop-point list for all vehicles.These include the new picking-up and dropping-off points specified by the request. The stop-point list for each vehicle corresponds to the arrival time at the user's destination that minimizes the cost function based on user delay time.
3. The system selects the vehicle with the best schedule, i.e. the stop-point list that has the lowest cost among all the vehicle, and assign the request to that vehicle. Namely we adopt the stop-point list of that vehicle to satisfy the request.

The cost function takes into account the time required to fulfill the request and the delay time incurred by other users. Minimizing the cost function allows for faster transportation of users to their destinations.

Due to the nature of the algorithm, it is not possible to consider factors such as the user's boarding location or schedules involving transfers. If there is a sufficient number of vehicles, simulations have confirmed that this method allows users to reach their destinations faster than traditional fixed bus services.

3 PROPOSED METHOD

3.1 Overview

When travel demand concentrates on a single location, ride-sharing vehicles are also concentrated on around that destination, causing congestion. The proposed method aims to alleviate congestion by controlling the movement of ride-share vehicles.controlling the movement of ride-share vehicles is implemented when congestion occurs on a specific road. Transfer promotes to ride a single vehicle with multiple users, and the traffic volume of ride-sharing around the destination is reduced. One transfer reduces by one the number of vehicles going to the destination, which means that, while traffic congestion is continuing, the system needs to instruct a large number of vehicles to transfer.

This study assumes the operation of a ride-sharing service through a centralized control system. The central server controls the routing of all vehicles. When a user arrives at the picking-up point, he/she sends a vehicle request to the central server, including the picking-up location and destination. When the central server receives a request, it selects an assigned vehicle and updates the stop-point so that the vehicle transport the user to the destination. If multiple users are traveling in a single vehicle, the system instructs the vehicle to follow a route that minimizes the total delay time for each user.

Specifically, the proposed method extends the SBI method described in Sec. 2.2. This system typically uses the SBI method to determine the vehicle and route for transporting the user to their destination. The proposed method monitors congestion on each road, and upon detecting congestion, the system transitions to the Transfer Promotion (TP) state. In the TP state, vehicles traveling on congested roads receive transfer instructions from the system, promoting ride-sharing. The proposed method constantly monitors traffic congestion on each road. This control reduces the number of ride-sharing vehicles traveling on the congested roads, thereby alleviating traffic congestion. When the congestion is cleared by this control, the system exits the TP state and returns to the normal state. In the normal state, the system works acording to the SBI method again to compute routes and does not provide any transfer instrucution.

3.2 System Model

Let $G = (I, A)$ be the road network, where I is a set of intersections and A is a set of road segments. There are stop-points on the road network G, and users can board and alight their vehicles at the access points. V is defined as a set of ride-sharing vehicles on road network G. Let V_{len}^{arg} be The average vehicle length. Let μ be the user capacity of a vehicle (excluding the driver).

A user of the ride-sharing service sends a request $q = (o_q, d_q, t_q)$ to the system. o_q is the ride location, d_q is the destination, and t_q is the time when the request is sent. Upon receiving a request, the system assigns a vehicle to the request and updates the vehicle's planned route. The vehicle picks up the user at the picking-up position o_q and delivers the user to the destination d_q according to the planned route. The planned route is determined using the SBI method.

Each vehicle $v \in V$ has a planned route $R_v = (r_1, r_2, \ldots, r_n)$ and moves along R_v. Let $r_k = (a_1^k, a_2^k, \ldots, a_m^k)(1 \leq k \leq n)$ be the route from stop-point p_o^k to p_d^k, where p_o^k is located on a_1^k and p_d^k on a_m^k, respectively. For r_k, p_o^k is called the starting point and p_d^k the ending point of r_k.

The list of stop-points on a planned route R_v is denoted by P_v. Specifically, let $P_v = (p_1, p_2, \ldots, p_{n+1})$ be the stop-point list on the planned route $R_v = (r_1, r_2, \ldots, r_n)$, where starting point of $r_k(1 \leq k \leq n)$ is p_k and the ending point is p_{k+1}. The route from the start point p_1 to the end point p_x included in the planned route R_v is denoted by R_{v,p_x}. That is, R_{v,p_x} refers to the route until reaching the stop-point p_x

Algorithm 1 Incremental Optimal Insertion Method

1: **for** v in V **do**
2: $\quad C_q^v, R_v' \leftarrow$ minInsertCost(P_v, q)
3: **end for**
4: $v \leftarrow \text{argmin}_{v \in V} C_q^v$
5: $R_v \leftarrow R_v'$

Function 1 minInsertCost$(P_v, q = (o_q, d_q))$

1: **for all** $(p_v^i, p_v^j) \in \dot{P}_v \times \dot{P}_v$, $i \leq j$ **do**
2: $\quad P_v^{ij} \leftarrow$ insertPositon$(P_v, o_q, d_q, p_v^i, p_v^j)$
3: $\quad R_v^{ij} \leftarrow$ calcualteShortestRoute(P_v^{ij})
4: $\quad C_q^{vij} \leftarrow$ calculateAssignCost(R_v^{ij})
5: **end for**
6: $x \leftarrow argmin_{ij} C_q^{vij}$
7: **return** C_q^{vx}, R_v^x

in the planned route R_v. For simplicity, we sometimes refer the planned route R_{v,d_q} to the destination d_q of a request q as $R_{v,q}$. R_v is sometimes referred as $R_v = (a_1, a_2, \ldots, a_n)$ as a list of road segments along the route.

Also, let $R_{v,a}$ be the route to reach road segment a. For each road segment $a \in A$, let δ_a^t be the estimated passing time at time t. This is the time estimated at time t to take from entering until exiting a. Similarly, we define the estimated transit time $\delta_r^t = \sum_{a \in r} \delta_a^t$ and $\delta_R^t = \sum_{r \in R} \delta_r^t$ for r and R. If the current time is t, the estimated transit time that vehicle v travels along the planned route $R_{v,p}$ and arrives at the stop location p can be denoted as $t_p^v = t + \sum_{a \in R_{v,p}} \delta_a^t$. For a stop-point list $P_v = (p_1, p_2, \ldots, p_{n+1})$ of vehicle v, let $\dot{P}_v = (p_0, p_1, \ldots, p_{n+1})$ denote the stop-point list that includes the previous stop-point p_0. Here, p_0 is the stop-point that the vehicle has just passed, and vehicle v is on the road between P_0 and P_1 at that time.

The SBI method assigns a new request q to the vehicle that minimizes the cost increased by accepting q. The cost is the sum of the delay time when q are allocated to the vehicle. Specifically, for a vehicle $v \in V$, let Q_v be the set of accepted requests, and the cost is calculated as follows.

$$C_q^v = \delta_{R_{v,q}}^t + \sum_{q' \in Q_v} (\delta_{R_{v,q'}}^t - \delta_{R'_{v,q'}}^t), \tag{1}$$

where R_v' denotes the planned route of v before query q is added. That is, R_v is obtained by merging query q to the planned route R_v', the first term $\delta_{R_{v,q}}^t$ denote the arrival delay of q to its destination o_q, and the second term denotes the increase in delay time of each accepted query $q' \in Q_v$ by inserting q into the schedule.

Algorithm 1 shows the request assignment algorithm of the SBI method. Whenever the system receives a new request q, it executes Algorithm 1 and assigns q to one of the vehicles $v \in V$. In lines 1-3, the cost of assigning q to each vehicle v is calculated according to formula (1). $minInsertCost(\cdot)$ is calculated by adding stop-points for q to the arbitrary locations in the stop-point list $P_v = (p_1, p_2, \ldots, p_{n+1})$ for vehicle v, computing the cost of every case based on formula (1), and finding the minimum among them. In other words, this algorithm computes planned routes $R_v^1, R_v^2, \ldots, R_v^{\frac{(n+2)(n+3)}{2}}$ and its cost C_q^v for each $P_v^1 = (o_q, d_q, p_1, p_2, \ldots, p_{n+1}), P_v^2 = (o_q, p_1, d_q, p_2, \ldots, p_{n+1}), \ldots, P_v^{n+3} = (p_1, o_q, d_q, p_2, \ldots, p_{n+1}), P_v^{n+4} = (p_1, o_q, p_2, d_q, \ldots, p_{n+1}), \ldots, P_v^{\frac{(n+2)(n+3)}{2}} = (p_1, p_2, \ldots, p_{n+1}, o_q, d_q)$, finds the minuimum-cost R_v^x at tine t^q, and return R_v^x and its cost C_q^v After that, in line 4, we select the vehicle v with the minimum cost. For the selected vehicle v, in line 5, we assign request q and update the planned route of v to the new route R_v.

Function 1 shows The procedure of the function $minInsert$ $Cost(\cdot)$ to find the optimal planned route for vehicle v to process query q. Lines 1-5 creates all combinations to insert picking-up point o_q and destination d_q of query q. Line 2 creates a planned route with o_q inserted immediately after P_v^i and d_q immediately after P_v^j. Here, as we mentioned before, $\dot{P}_v = (p_0, p_1, \ldots, p_{n+1})$ is the stop-list list of v, with the last stopped point p_0 added as the first element. The line 3 finds the planned route R_v^{ij} corresponding to P_v^{ij}. Line 4 computes the cost of R_v^{ij} according to formula (1). In line 6, let x be the one with the smallest cost among all combinations of i and j. Finally, line 7 returns the cost C_q^{vx} and the planned route R_v^x.

Up to this point, it has been the same ride-sharing model as related work. Based on the SBI method, it is possible to transport multiple individuals using a single vehicle. In the proposed method, transfer processes are added to reduce the number of vehicles traveling on the road. When a traffic congestion is detected on any road segment $a \in A$ on the road network G, the model transitions to the Transfer Promotion (TP) state. Once congestion is relieved, and it is confirmed that there is no congestion on all roads, the system returns to normal state. In the TP state, the system alleviates congestion by triggering transfers for vehicles scheduled to travel on congested roads, thereby reducing the number of ride-sharing vehicles on congested roads. The transfer instruction method is described in the next section.

3.3 Determination of Transfer Vehicle and Transfer Point

When traffic congestion occurs on a road segment a_c, the system transitions to a TP state. Let $R_{v,a_c} = (a_1, a_2, \ldots, a_n)$ be the route of vehicle v to the congested road and δ_{R,a_c}^t be the time taken to reach the congested road a_c. In the TP state, At regular intervals, the system instructs the vehicles to initiate transfers. The transfer schedule is calculated as the planned route for each vehicle, including transfers. In other words, the system modifies the SBI-based planned route of each vehicle described above to incorporate a transfer schedule for some vehicles. The transfer schedule calculation first determines the two vehicles that will transfer, and then calculates the transfer point and the routes so that their users can transfer between them. The first step is to find transfer vehicle pairs that can transfer in a low cost.

The transfer vehicle pairs are calculated as follows. Let a_c be the road where congestion is detected when the system transition to the TP state. The TP state persists until congestion at a_c is resolved. At regular intervals, the system gets a set of vehicles scheduled to pass through a_c. The system

Algorithm 2 Transfer Scheduling Algorithm

```
1: V_s ← emptyList()
2: for v in V do
3:     if a_c ∈ R_v then
4:         append(V_s, v)
5:         compute δ^t_{R_{v,a_c}}
6:     end if
7: end for
8: sortVehicles(V_s)
9: for i = 1 to |V_s| do
10:    for j = i to |V_s| do
11:        p_s ← getTransferPlace(v_1, v_2)
12:        C_{v_1,v_2} ← calcTransferCost(v_1, v_2, p_s)
13:        if C_{v_1,v_2} < γ then
14:            R_{v_1} ← changeDest(P_{v_1}, p_s)
15:            R_{v_2} ← insertPlace(P_{v_2}, p_s)
16:            remove (V_s, v_1, v_2)
17:        end if
18:    end for
19: end for
```

sort these vehicles by $\delta^t_{R_{v,a_c}}$, which are their scheduled arrival time to a_c, and create a list of vehicles $V_s = (v_1, v_2, \ldots, v_n)$ in the ascending order. Let t be the current time, and $\delta^t_{R_i,a_c} \leq \delta^t_{R_j,a_c}$ meets for any $i < j$. A nested loops are executed on V_s to retrieve pairs (v_i, v_j) in the order of the arrival time to a_c, and if the transfer condition is satisfied, the transfer point and transfer schedule are calculated and applied. The transfer conditions are that, first, the destinations are the same, and second, that the transfer cost to the users si lower tha a predetermined threshold. The transfer cost is explained in the next section.

After determining the transfer route for the vehicle pairs (v_i, v_j), it is necessary to calculate where the transfer will take place. Let R_{v_i,a_c} and R_{v_j,a_c} be the planned routes of v_i and v_j, respectively. Both of these planned routes include a_c and therefore two vehicles will eventually converge. Therefore, the vehicle pair shares a common route. Among these, select the road segment a_s that is the furthest from a_c and on which no congestion is detected, and set the appropriate transfer place on a_s as the transfer point. If there is no suitable transfer point that meets the conditions, the vehicle pair (v_i, v_j) is not instructed to transfer.

Algorithm 2 shows the algorithm for computing the transfer schedules. In lines 1-7, we create a list V_s of all vehicles whose planned route includes the congested road segment a_c, and compute the arrival time to a_c of each vehicle $v \in V_s$. In line 5, t is the time when the Algorithm 2 is executed. In line 8, V_s is sorted based on arrival time to a_c, and in lines 9-10, all transfer vehicle pairs in V_s are generated with the priority given based on the arrival time. In line 11, the transfer point is calculated, and in line 15, caluculated the transfer cost. In lines 12-16, when the transfer cost is less than the threshold value γ, a transfer schedule is applied. In lines 14-15, a new planned route is created by inserting a transfer point into each of the planned routes of v_1 and v_2. In line 16, v_1 and v_2 are removed from V_s so that they will not appear in the transfer vehicle calculation. After Algorithm 2 finishes, each vehicle is notified of the updated planned route and is instructed to transfer.

3.4 Calculation of Transfer Cost

The system calculates the transfer cost, which is the delay time incurred by the transfer, for each of the transfer vehicle pair (v_1, v_2). When the transfer cost is less than threshold γ, the system conducts transfer. Transfers are made such that v_1 first arrives at the transfer point p_s, drops off all users, v_2 arrives there, and picks up all of them. Therefore, the transfer cost is determined by the total increase in arrival time for each user.

Let n_1 and n_2 be the number of boarding users in vehicles v_1 and v_2, respectively. Since the planned routes include p_s for both v_1 and v_2, transfers do not alter the route of the vehicle. Let δ_{tran} be the time required for users boarding and alighting at p_s. If the current time is t, the time when v_1, v_2 arrive at p_s be denoted as $\delta^t_{R_{v_1,p_s}}$ and $\delta^t_{R_{v_2,p_s}}$,respectively. Therefore, the transfer cost can be calculated as follows.

$$C_{v_1,v_2} = (n_1 + n_2)\delta_{tran} + n_2(\delta^t_{R_{v_2,p_s}} - \delta^t_{R_{v_1,p_s}}). \quad (2)$$

4 CONCLUSION

Compared to traveling by private vehicle, using ride-sharing services allows multiple people to travel in a single vehicle, leading to an anticipated reduction in traffic volume. In the near future, in a society where ride-sharing services are widely adopted, the proportion of people using ride-sharing services without owning private cars is expected to significantly increase In this study, we propose a method to mitigate congestion by reducing the number of ride-share vehicles on roads near the destination and issuing transfer instructions when congestion is detected. This approach aims to address congestion issues in a society where ride-sharing services are prevalent.

REFERENCES

[1] Ma Shuo, Zheng Yu, and Wolfson Ouri: Real-time City-ScaleTaxi ride-sharing, IEEE Transactions on Knowledge and Data Engineering VOL27, NO.7, pp. 1782-1795 (2015).

[2] Ayano Okoso, Keisueke Otaki, Satoshi Koide, and Tomoki Nishi: Path Optimization with Future Demand for Ride-sharing(in Japan), Transactions of Information Processing Society of Japan Vol.60 No.10, pp. 1653-1661 (2019).

[3] Noda Itsuki, Ohta Masayuki, Shinoda Kousuke, Kumada Yoichiro, and Nakashima Hideyuki: Evaluation of usability of dial-a-ride systems by social simulation, Lecture Notes in Computer Science, pp. 167-181 (2003).

Development of a Framework for Jaywalking Risk Map to Reduce Pedestrian-to-Vehicle Accidents

Yuichi TOKUNAGA*, Atsushi Yamamoto**, and Masashi SAITO*

* Faculty of Graduate School of Engineering, Kanazawa Institute of Technology, Japan
** Department of Management Systems, Kanazawa Institute of Technology, Japan
{y.tokunaga, msaito}@neptune.kanazawa-it.ac.jp

Abstract - Crossing outside of a crosswalk or ignoring a traffic signal is called jaywalking. Jaywalking is a major road safety issue because even vehicles with safety features have difficulty predicting jaywalking suddenly crossing the road from the shadow of an obstacle. We hypothesize that the possibility of jaywalking is related to the road and its surrounding environment and propose a method to quantify the risk using a statistical method and convey it to traveling vehicles. We have developed a framework for generating risk potentials using Bayesian networks to realize this proposal. If our hypothesis is correct, we should be able to estimate the risk potentials of jaywalking using our generated model outside of the observed location. We plan to verify the accuracy of the risk maps generated by this framework by comparing them with actual measurements.

Keywords: Jaywalking, Risk map, Bayesian network.

1 INTRODUCTION

In recent years, the number of accidents involving automobiles has been on a downward trend due to the improvement of the road traffic environment, the spread of traffic safety awareness, and improvements in safety performance of vehicles. However, the number of accidents involving people versus vehicles has remained flat in recent years [1]. In particular, the number of people killed in accidents while walking continues to be the leading cause of death by condition and has begun to increase slightly. In addition, crossing violations and disregard for traffic signals account for more than 45% of the causes [2]. These dangerous behaviors of pedestrians are called "jaywalking" and countermeasures against them have become an issue in road traffic, including overseas [3]. It is assumed that elderly pedestrians jaywalk due to the overconfidence brought on by years of experience despite their declining physical ability and sensory perception. Although automobile pedestrian detection has improved over the years, the braking distance is insufficient for an emergency stop against a Jaywalker emerging from between vehicles in the oncoming lane. Prioritizing safety requires a significant reduction in speed, resulting in traffic congestion. As a result, it is difficult to avoid jaywalkers who may cross from anywhere at any time.

Common countermeasures against jaywalking include the installation of anti-crossing barriers (guardrails) and regulatory signs. Anti-crossing barriers are installed between the roadway and sidewalk to physically deter people from crossing, and are considered to be an effective measure to prevent jaywalking. Regulatory signs can be installed in various locations due to their ease of installation, but they are not very effective, and neither of them is a fundamental solution to the problem.

Therefore, instead of eliminating jaywalking, we should consider foreseeing pedestrian’s jaywalking and avoiding them on the vehicle side. As mentioned above, it is difficult for vehicles to be prepared for jaywalking from all points, but by identifying points with a high probability of jaywalking, it is possible to reduce speed only there, thereby achieving both safety and avoidance of traffic congestion. To achieve this, this paper proposes to statistically determine the risk of jaywalking on a road.

Section 2 introduces related works on jaywalking countermeasures, Section 3 describes the proposed framework based on the hypotheses derived from these studies, Section 4 describes its implementation in detail. Finally, in Section 6, we discuss the results and future works.

2 RELATED WORK

More than 70% of those killed in walking accidents are elderly with older than 65 years old [2]. According to an analysis of accidents involving elderly people crossing the road by the Traffic Accident Analysis Center [4], 51% of elderly people who had accidents while crossing the road did not notice the approach of a car because they did not check for safety or check insufficiently. A typical accident scenario is shown in Figure 1. The pedestrian starts crossing immediately after A1's car passes, continues crossing without checking left and right directions, and crosses without noticing B1's car approaching from the opposite direction.

Technology to detect and avoid invisible pedestrians is being considered as the features of Advanced Driver Assistance System (ADAS) for automobiles [5]. For example, research is being conducted on technologies to recognize the

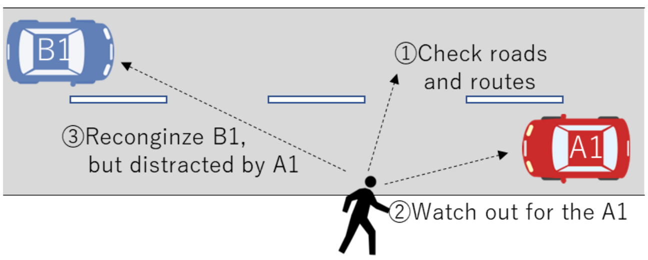


Figure 1: Mechanism of accident while crossing

movement of invisible objects from images reflected in curved mirrors and reflective objects [6][7], and to share sensing information of other vehicles, GPS information of pedestrians, and image or sensing information from infrastructure facilities through V2V, V2P, and V2I communications respectively and to detect the danger range [8]. However, all these technologies are dependent on outdoor reflective objects, sensors and cameras, and pedestrian communication devices, and thus lack versatility and do not provide fundamental solutions to the problem.

The use of risk potential maps has been proposed to reduce accidents caused by unseen hazards [9]. A risk potential map shows not only the hazards physically present in the driving path, but also the degree of danger in a space based on past cases and probability, using numerical values and colors like contour lines. For example, an intersection without a signal, which is a blind spot at a wall, has a higher danger level because of the risk of ejection. The risk map generated in this way is used to realize safe driving by presenting the driver using Head-Up-Display (HUD) and Augmented Reality (AR) or by referring in the automatic driving control to plan the low-risk routes or to control speed. Several studies and demonstrations have been conducted, but the challenge is how to narrow down the risk; if the risk map is applied to jaywalk, generalized to roads without pedestrian crossings, the jaywalk risk is high, and cars are always forced to drive at low speeds. On the other hand, if only points where jaywalking has occurred in the past are considered high risk, accidents caused by jaywalking at other points cannot be prevented. It is necessary to narrow down the risk value with more accurate information.

On the other hand, there have been many studies on pedestrian crossing behavior from the viewpoint of road design in the field of civil engineering planning [10][11]. Hamamoto et al. studied pedestrian behavior at a national road intersection leading to a station connection [12], and explained that pedestrians want to take the shortest route possible, and if there is a possible crossing on the route, they will cross even if it is outside the pedestrian crossing area. The study notes that jaywalking is more likely to occur at locations where crossing cannot be physically deterred, such as street intersections, entrances to roadside facilities, and openings in anti-crossing barriers. Takehira also studied the relationship between roadway facilities and jaywalking on several roads in densely populated urban areas [13]. Unlike national roads, anti-crossing barriers on urban streets are often characterized by disconnection at the approaches to stores and parking. In all the study locations observed, more than 90% of the disorderly crossings occurred at locations where there were openings at both the start and end points.

Thus, it is shown that the potential risk of jaywalking may be inferred from the road environment, and anti-crossing barriers are an important clue in estimating jaywalk risk. On the other hand, many community streets and suburban roads are not equipped with anti-crossing barriers, and the presence or absence of such barriers alone cannot narrow down the potential risk area of jaywalking. It is essential to discover environmental features other than anti-crossing barriers for risk mapping that aims at both safety and traffic efficiency.

3 PROPOSAL OF THE FRAMEWORK

3.1 Target Setting

Related works have shown that the opening of an anti-crossing fence affects the probability of jaywalking occurring. So, if we create a model to estimate the probability of jaywalking based on these road environment characteristics in a certain limited area and expand it to the whole country using map information and this estimation model, we can create a national jaywalking risk map. However, to narrow down the risk area, it is necessary to discover features other than anti-crossing barriers.

Therefore, we formulate a hypothesis based on use cases. Facilities where people gather, such as train stations and schools, are often located on main streets, and pedestrians targeting the facility flow into the main street from side streets near the facility, bypassing nearby pedestrian crossings to follow them to the facility. If there are no crosswalks nearby, the behavioral psychology of choosing the shortest route is at work, and pedestrians are more likely to cross at the point where they exit the side street onto the main street. Based on this use case, it is highly likely that the mutual location of the facility where people congregate, the influx of side streets in the vicinity, and the crosswalks affect the potential risk probability of jaywalking. The goal of this study is to substantiate this hypothesis.

To achieve this goal, we will generate a jaywalking estimation model based on the location of facilities and other features and crosswalks and evaluate whether the risk map derived from the model is valid outside of the observation points.

3.2 Framework Construction

Creating a jaywalking estimation model requires the development of factors such as facility type and feature extraction, as well as observational data. On the other hand, whether an appropriate risk map can be obtained depends on the hypotheses based on use cases, and there is no guarantee that the results will be worth the effort of creating the model. Therefore, the initial stage is to generate a model from a minimal dataset to prove the hypothesis, while building a framework to extend the missing variables and data.

An overview of the framework is shown in Figure 2. First, location information of features in the road environment that induce jaywalking, such as facilities and pedestrian crossings, are automatically obtained using open data. Next, location dataset is created, and an estimation model is generated. Utilizing the created estimation model, the probability of occurrence of jaywalk is estimated. Finally, a heat map is superimposed on the map using the occurrence probability as the risk value to create a jaywalking risk map. Each of these functions is described below.

3.2.1. Extraction of location information on the road environment

Google Maps is used to automatically acquire location information of features on the road environment that trigger

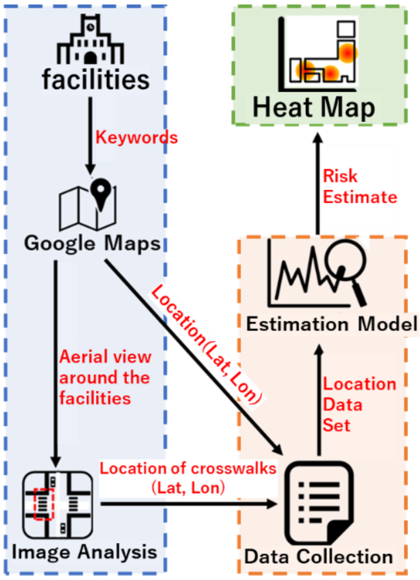


Figure 2: Framework overview

jaywalk. Facilities such as stations and schools are listed by the search function and their latitude and longitude information is obtained. On the other hand, pedestrian crossings cannot be extracted by the search function. As a method to address this problem, a study has been reported that analyzes aerial photo images to detect pedestrian crossings [14][15], and we will use the same method. The method of extracting crosswalks is explained in 4.1.2.

3.2.2. Generating an Estimation Model

Next, a location dataset is created from the calculated latitude and longitude information. The presence or absence of jaywalking is observed in advance from visual surveys, and the data is created by calculating the distance and direction for each latitude and longitude of the observation location. The created data is compiled as location data, and an estimation model is created using a Bayesian network.

Bayesian network is a probabilistic inference model that shows multiple causal relationships in a weighted graph structure (network) and indicates each causal relationship by probability [16]. Even when the full picture of causal relationships is not known, the estimation accuracy can be improved by accumulating findings obtained through observation as partial causal relationships. We believe that this is the best method for this project to clarify the causal relationship of the jaywalking outbreak through repeated hypotheses and observations.

Bayesian network can be expressed by the following equation :

$$P(x_1, \dots, x_N) = \prod_{i=1}^{N} P\left(x_i \middle| parent(x_i)\right) \qquad (1)$$

where x_1 denotes each event and $parent(x_i)$ denotes the upper event in the relationship. The probability of occurrence of a given combination of events is expressed as the product of the conditional probabilities of each causal relationship.

From the hypotheses set up in this paper, x_i indicates the relationship between the location of the facility at the observation point, the road structure, and the location of the pedestrian crossing. The model is generated by data on the jaywalking situation at each observation point, analyzing the dependencies between each causal event and determining the conditional probability of each relationship.

Bayesian estimation was used to create the estimation model. Details of the location dataset and estimation model are described in 4.2.2 and 4.2.3.

3.2.3. Generating a Jaywalking Risk Map

Finally, a jaywalking risk map is created by overlaying a heat map on a map with the probability of jaywalking as the risk value. The colors displayed on the heatmap should be separated according to the risk values so that the information can be obtained visually.

4 IMPLEMENTATION

The proposed framework automatically obtains the latitude and longitude of structures from Google Map searches and detects pedestrian crossings from aerial images to obtain location information. Based on these location relationships and the observed information on the availability of jaywalking, a Bayesian estimation model is created, and a jaywalking risk map is generated. This section describes the acquisition of crosswalk location information, creation of the Bayesian estimation model, and generation of the risk map.

4.1 Extracting Crosswalk Locations

4.1.1. Image Collection

Using the location data (latitude and longitude) of the facility retrieved by Google Maps, aerial images centered on the facility (or its center of gravity in the case of multiple facilities) are collected from the Google Maps Platform (hereinafter referred to as "GMP"). GMP outputs a 70-meter square aerial image by specifying the location data of the center of the image to be acquired. Assuming a 1 km square risk map, the location data sets of 15 x 15 = 225 images must be specified to GMP. To minimize gaps and overlaps at image boundaries, the location data of the center point of each image was determined using the Vincent method from the latitude and longitude of the center point of the target area, and the distance and direction of each image. These data sets were input into GMP to obtain the images.

4.1.2. Extraction of crosswalks

Although several methods have been proposed for extracting pedestrian crossings from aerial images, we apply YOLO v5, which is widely used for general object detection applications. We trained on 300 crosswalk images extracted from aerial photographs and obtained a high detection accuracy with an F value of 0.95 by setting confidence>= 0.8.

Using the acquired crosswalk coordinate information and the latitude and longitude information of the images, the

Figure 3: Observation points of jaywalking

latitude and longitude of the crosswalk are calculated YOLO outputs the coordinate information of the bounding box of the detected object. Since the latitude and longitude information of the center of the original image is known, the latitude and longitude information of the center of the bounding box was calculated from this information, and this was used as the location of the crosswalk.

4.2 Creating a Bayesian network Model

4.2.1. Collection of training data

To obtain the training data, a visual survey of jaywalking was conducted. Observations were made for in an area centered in front of the main gate of the university, mainly during the hours when the school attendance was concentrated, and the starting point of the jaywalking was recorded. Although the percentage of pedestrians killed in accidents while walking is higher among the elderly, we observed jaywalking among all age groups of pedestrians, as jaywalking is done regardless of age. The overview of the study area is shown in Figure 3 and the time trends of observed jaywalking are shown in Figure 4. Red boxes in Figure 3 indicate locations where jaywalks were observed, and black boxes indicate locations where a particularly large number of jaywalks were observed. Many jaywalkers were observed at the intersection of side streets to the main street, and many pedestrians were observed jaywalking not only to and from the university but also to and from the supermarket.

4.2.2. Creation of Dataset

The dataset consists of five items: the number of jaywalkers, the distance and direction to the nearest facility, and the distance and direction to the nearest pedestrian crossing. Road structure was excluded from the item because the criteria for categorization had not been clarified. This is one of the issues to be addressed in the future.

The visual survey results are entered into the number of jaywalking occurrences in this dataset. The survey in this study covered an area of 80 meters from north to south and

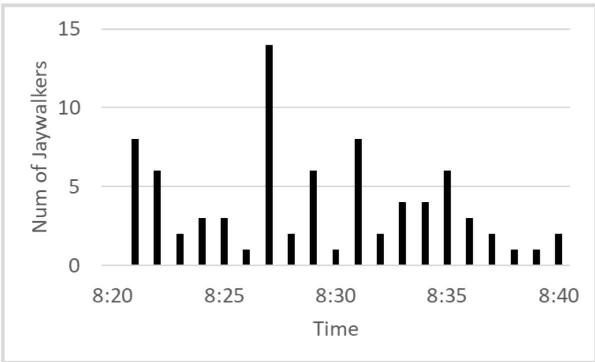


Figure 4: Trends in occurrence of Jaywalking

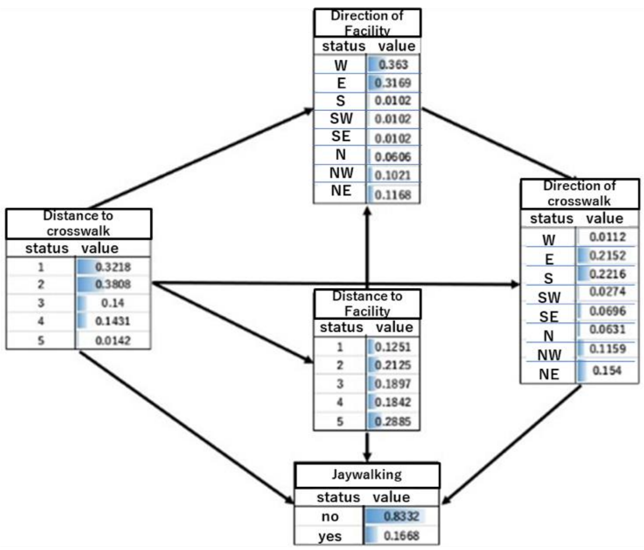


Figure5: Created Bayesian network

800 meters from east to west. This corresponds to 80 mesh x 800 mesh = 68,000 records.

4.2.3. Estimation Model Generation

BayoLinkS [17], software for building Bayesian network, was used for model generation. This software can learn structures from data, extract dependencies (Bayesian networks), and generate estimation models. Since the data needed to be discretized, the distances in the dataset were equally divided every 50 meters, and the directions were labeled in eight levels: east, west, south, north, and south. The number of divisions for discretization has a significant impact on model performance, so the appropriate number of divisions is an issue to be considered in the future.

The Bayesian network constructed by structural learning is shown in Figure 5. The solid line shows the dependencies, and the table below the items shows the probability of occurrence of each state relative to jaywalking probability. In the current study, the distance to the facility and the crosswalk and the direction of the crosswalk are directly causally related to the occurrence of jaywalking, while the direction of the facility is causally related to the direction of the crosswalk. This is considered to be a result of the influence of the learned terrain and is a challenge when generalization, i.e., when the model is used to estimate other areas.

4.3 Risk Map Generation

The risk maps are generated in QGIS, using the estimation model generated by BayoLinkS to obtain the probability of jaywalking occurrence for each mesh location, which is then converted into a heat map. Since we want to show a risk map for a road, we obtain road information from OpenStreetMap and display the probability of occurrence only for the mesh location that overlaps the road. The created risk map is shown in Figure 6. The circled dots represent pedestrian crossings.

The green boxes indicate visual survey locations, i.e., areas of correct data. The upper left corner of the map indicates a high-risk area. This area is considered high-risk because it is a residential area and there are no pedestrian crossings. On the other hand, there are no areas near pedestrian crossings that are indicated as high risk, indicating that the distance from the pedestrian crossing has a significant impact.

Figure 6: Jaywalking risk map

5 CONCLUSION

This study developed a jaywalking risk map that visualizes the probability of jaywalking occurrence as a potential risk with the aim of preventing jaywalking, one of the causes of human-vehicle accidents. Based on related works, we hypothesized that the mutual location of facilities where people gather, the influx of side streets around them, and pedestrian crossings affect the occurrence of jaywalking, and aimed to create an environment that can be used as a framework to materialize this hypothesis. We extracted the location of keyword facilities and crosswalks from maps and aerial imagery, generated a Bayesian network risk estimation model along with observation data, and displayed the risk of jaywalking on the road in a QGIS heat map.

We generated an estimation model using this framework based on observations of signal neglect in the vicinity of a university facility, and confirmed that it is possible to generate a Jaywalking Risk Map for an area that is larger than the observed area. Areas without pedestrian crossings in residential neighborhoods show a high level of danger, while areas near pedestrian crossings show a low level of danger, as expected. In the future, additional jaywalking observations will be conducted for the risk map areas, and the accuracy will be verified by comparing the generated maps with actual measurements.

ACKNOWLEDGEMENT

This research was supported by a Grants-in-Aid for Scientific Research(C) numbered 21K11856.

REFERENCES

[1] Cabinet Office, FY2022 Status of Traffic Accident and Current State of Traffic Safety Measures(in Japanese), WHITE PAPER ON TRAFFIC SAFETY IN JAPAN 2023, pp. 50-70 (2023).

[2] National Police Agency, Traffic Accidents in 2022 and Other Matters(in Japanese), https://www.npa.go.jp/news/release/2023/20230302jiko.html (19 July 2023)

[3] Millington, G., Jaywalking, In Shades of Deviance, Routledge, pp. 17-20 (2014).

[4] H. Shibasaki, Accidents Involving Elderly People Crossing the Street (in Japanese), Institute for Traffic Accident Research and Data Analysis, https://www.itarda.or.jp/presentation/19 (31, Jan 2023)

[5] L Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, W. Shi, Computing Systems for Autonomous Driving : State of the Art and Challenges, IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6469-6486 (2021)

[6] Y. Shi, X. Ying, and H. Zha, Unsupervised Domain Adaptation for Semantic Segmentation of Urban Street Scenes Reflected by Convex Mirrors, IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 24276-24289 (2022)

[7] R. Sato , and K. Onoguchi, Scene Recognition for Blind Spot via Road Safety Mirror and In-Vehicle Camera, 26th International Conference on Pattern Recognition (ICPR), pp. 3714-3720 (2022)

[8] A. Kabil, K. Rabieh, F. Kaleem, and M. A. Azer, Vehicle to Pedestrian Systems: Survey, Challenges and Recent Trends, IEEE Access, vol 10, pp. 123981-123994 (2022)

[9] P. Raksincharoensak, T. Hasegawa, and M. Nagai, Motion Planning and Control of Autonomous Driving Intelligence System Based on Risk Potential Optimization Framework, International Journal of Automotive Engineering, vol. 7, no. 1, pp.53-60 (2016)

[10] Y. Nagawaki, and K. Suzuki, Empirical Analysis of Factors for Jaywalking and Its Preventive Measures on Urban Road, JSTE Journal of Traffic Engineering, vol. 8, no. 2, pp. A_149-A_158 (2022).

[11] K. Adachi, K. Suzuki, Y. Okada, and H. Ito, Empirical Analysis of Jaywalking Behavior on Two-stage Crosswalk, JSTE Journal of Traffic Engineering, vol. 7, no. 2, pp. A_102-A_109 (2021).

[12] K. Hamamoto, and T. Murayama, The Safety Evaluation Analysis of the Walker in the Crosswalk and Station Junction, JSCE, vol. 39, No. 402, pp. 1-4 (2009)

[13] S. Takehira, Study on Design and Operation of Urban Arterials Based on Nature of the Pedestrian Behavior, University of Tokyo (2014)

[14] Y. Ishino, and H. Saji, Extraction of Road Markings from Aerial Images, 2008 SICE Annual Conference, pp. 2180-2183 (2008)

[15] A. Maeda, Automatic Generation of a Pedestrial Road Network Using Aerial Photographs and Maps, Journal of Information Processing, vol.61, no.2, pp. 262-273 (2020)

[16] D. Heckerman, A Tutorial on Learning with Bayesian Networks, Innovations in Bayesian Networks: Theory and applications, pp. 33-82 (2008)

[17] NTT DATA Mathematical Systems Inc., "BayoLinkS", https://www.msi.co.jp/english/index.html (11 June 2023)

Obstacle avoidance with extended velocity obstacles algorithm

Takumi Seita*, Shunsuke Michita*, Seiji Komiya* and Toshihiro Wakita*

*Graduate School of Engineering, Kanagawa Institute of Technology, Japan
s2384004@cco.kanagawa-it.ac.jp

Abstract - Many traffic accidents occur on community roads where pedestrians and vehicles coexist, therefore safe and smooth autonomous driving is expected. Based on this situation, this paper proposes a path planning method for autonomous vehicles by extending the Velocity Obstacles algorithm (VO). VO is commonly used in mobile robotics. Three points were extended to apply the algorithm.

The first point was to use 2D TTC to reduce unnecessary avoidance. It was decided that avoidance would not be performed if there is room for it based on the 2D TTC calculation of collision time. The second point established motion constraints by assuming avoidance is performed with steady-state circular motion because vehicles have more motion constraints than robots. The third point was to change the collision detection method from circle approximation. Considering the size and shape of the vehicle, it was decided to draw tangents from the four corners of the vehicle to the obstacle and combine them.

Using the above algorithm, avoidance paths were generated and better performance than VO method.

Keywords: Autonomous Mobility, Pedestrian prediction

1 INTRODUCTION

In recent years, although the reduction of traffic accident fatalities has been decreasing, the decreasing trend of the number of traffic accidents on community roads has remained relatively low. According to data from the Ministry of Land, Infrastructure, Transport and Tourism, in 2021, there were approximately 80,000 traffic accidents on community roads, compared to approximately 220,000 accidents on arterial roads [1]. In terms of the reduction rate since 2004, arterial road accidents have decreased by about 70%, while community road accidents have decreased by about 60%. Therefore, in order to reduce traffic accidents on community roads as well as arterial roads, the realization of safe and efficient autonomous driving is desired.

However, community roads have distinct characteristics such as the lack of white lanes, pedestrians freely walking along various paths, and narrow road widths, making a significantly different traffic environment compared to arterial roads. Because there are numerous problems that differ from traditional autonomous driving, it difficult to achieve autonomous driving on community roads.

Autonomous navigation in complex environments like community roads has been extensively studied in the context of mobile robots.

Some representative methods that have been traditionally studied include the Dynamic Window Approach (DWA), the risk potential method, and the Velocity Obstacle (VO) method.

DWA [2] is an approach that generates trajectory candidates by combining translational and rotational velocities. It evaluates each generated trajectory using an evaluation function. Following that, the trajectory with the highest evaluation value is selected, and the robot follows that trajectory. One of the problems of this method is that it does not consider dynamic obstacles, which may result in insufficient avoidance of dynamic obstacles such as pedestrians.

The risk potential method [3] defines repulsive potentials and an attractive potential to the target reaching position to obstacles. It generates target velocities and target paths based on the calculated potential gradients. One of the problems of this method is that the generated potential field is a virtual quantity without physical meaning. Due to the lack of physical significance, parameter adjustments are required, making it difficult to achieve proper control of the trajectory, velocity, and adaptation to various traffic environments.

Velocity Obstacle (VO) [4] is a method for motion planning of a robot in a dynamic environment with moving obstacles. This method defines a "velocity obstacle" as the velocity space within which a robot may have potential collisions based on the current positions and velocities of the robot and obstacles. By calculating velocity obstacles for all obstacles and selecting a velocity that does not belong to any velocity obstacle, this method generates obstacle-avoiding trajectories. As the latest algorithm in VO, Adaptive Velocity Obstacle (AVO) [5] has been proposed. This method is proposed to solve that if the robot chooses the critical velocity of the collision cone to minimize the time consumption, the error of the sensor will lead to the collision. This approach shows that the expansion radius of the moving obstacle is adjusted adaptively according to the relative velocity and distance between the robot and the obstacle. After, the DWA is used to get the local path through optimizing the velocity outside the collision cone. Regarding this method's issues, one could point out the limited range of selectable speeds when the robot approaches obstacles and the expansion radius increases. Another algorithm Ellipse-based Velocity Obstacle (EBVO) [6] has been proposed. This approach first selected the linear velocity with the EBVO, and then calculated the angular velocity proportional to the change rate of the boundary line of the EBVO with respect to the orientation of the robot. However, since the robot's angular velocity was depending on the user-defined parameter, the robot oscillated or collided with obstacles at worst if the parameter was set in the wrong way. And there is an improved version addressing the issues of this EBVO [7]. A common issue among these algorithms is the lack of consideration for the kinematic constraints of non-holonomic robots.

Among these methods, it was considered that VO could be applicable to autonomous driving on diverse community roads where numerous pedestrians act as moving obstacles and there is a wide range of traffic environment variations. However, applying VO to autonomous driving of vehicles on community roads presents several problems.

The first problem is the existence of numerous pedestrians. On community roads, there are often many pedestrians walking in various directions along narrow roadways. When applying VO to such situations, the velocity space may become entirely covered by velocity obstacle regions, leaving no viable velocity options to choose from.

The second problem is related to achieving the selected velocity. In VO, the velocity pair v_x, v_y that can reach the goal fastest among velocities not belonging to the velocity obstacle region is chosen. Many mobile robots are equipped with differential drives or omnidirectional wheels, making it relatively easy to achieve the selected velocity. However, in case of automobiles, which mostly adopt front-wheel steering and have longer wheelbases, there are significant constraints on achievable velocities. As a result, there can be a significant difference between the selected velocity and the realized velocity, potentially leading to inappropriate trajectory generation.

The third problem relates to road width. Generally, community roads have narrow roadways, which limits the freedom of path selection. On the other hand, in VO, there are cases where a viable velocity is not selected despite being capable of driving on narrow roads. One reason is that VO approximates both pedestrians and the ego vehicle with circles for collision detection. Since the shape of automobiles often differs significantly in terms of width and length, there can be a significant difference between the actual collision detection and the circle approximation. Additionally, the circular approximation tends to generate avoidance paths that take unnecessarily large avoidance distances from obstacles. As a result, on narrow community roads, there can be situations where no velocity candidates are available.

To address these problems, this study attempts to achieve autonomous driving on community roads by extending the capabilities of VO. The extended method is referred to as VO-drive in this research.

2 METHODS

2.1 VO

VO is a method for motion planning of a robot in a dynamic environment with multiple moving obstacles. In this algorithm, based on the current positions and velocities of the robot and obstacles, the velocity space of the robot that may lead to future collisions is defined as the "velocity obstacle." The positions of the robot and obstacle are presented as **A** and **B**, respectively, and their velocities as $\mathbf{v_A}$ and $\mathbf{v_B}$. The size of the robot and obstacles are approximated by circles circumscribing them. Each radius is presented R_A and R_B, respectively. The circle centered at **B** with a radius of $R_A + R_B$ is defined as the collision circle. In this case, the set of relative velocities $\mathbf{v}_{A,B}$ that would result in a collision between **A** and **B** forms a collision cone $\boldsymbol{CC}_{A,B}$ (Figure 1).

The collision cone $\boldsymbol{CC}_{A,B}$ is a planar region bounded by two tangents, λ_f and λ_r, from **A** to the collision circle. In this region, the relative velocity $\mathbf{v}_{A,B}$ between the robot and the obstacle would result in a collision. Thus, the region of relative velocities $\mathbf{v}_{A,B}$ that lead to a collision is determined.

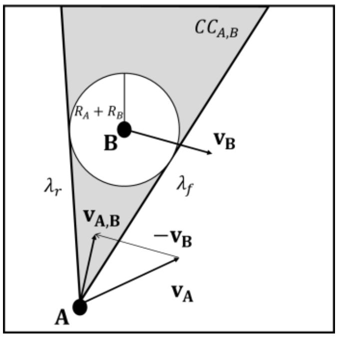


Figure 1 Relative distance $\mathbf{v}_{A,B}$ and collision cone $\boldsymbol{CC}_{A,B}$

Next, the region of robot velocities $\mathbf{v_A}$that would result in a collision is calculated. The obstacle velocity $\mathbf{v_B}$ is added to each velocity in the collision cone $\boldsymbol{CC}_{A,B}$. This is equivalent to translating $\boldsymbol{CC}_{A,B}$ by the velocity $\mathbf{v_B}$ (Figure 2). If the robot velocity $\mathbf{v_A}$ lies within this translated collision cone, a collision would occur. Therefore, this translated collision cone is referred as the "velocity obstacle."

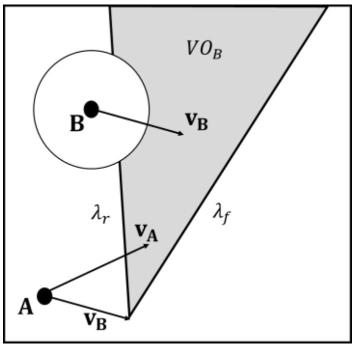


Figure 2 velocity obstacle

The selection of velocity for robot motion is performed as follows. First, an initial velocity is calculated based on a predetermined magnitude of velocity, assuming no obstacles. If the initial velocity lies outside the velocity obstacle, it is directly used for motion. If the initial velocity lies within the velocity obstacle, the velocity closest to the initial velocity on the outside of the velocity obstacle is selected and used for motion.

2.2 VO-drive

To adapt VO to community roads and autonomous driving, three extensions were made.

2.2.1. Limitation of the target object

The first extension addresses the issue of handling a large number of pedestrians. In community roads, there may be many pedestrians walking in various directions along narrow roadways. When applying VO to such a situation, the velocity space is covered by the velocity obstacle region, and no available velocities become exist. To resolve this, a method can be considered to limit the pedestrians (obstacles) that need to be avoided using some criteria. Possible criteria include distance, direction, etc. In this study, a two-Dimensional Time To Collision (2D TTC) based on the time margin until collision is used [5]. It is known that 2D TTC is related to pedestrians' sense of safety in collision avoidance for mobile robots. Therefore, pedestrians with a 2D TTC value above a certain threshold can be considered to have a lower risk of collision and can be excluded from collision detection.

Time To Collision (TTC) is a physical indicator used in Autonomous Emergency Braking (AEB) systems in vehicles. This indicator represents the time until collision with a leading vehicle if the current relative velocity between the ego vehicle and the leading vehicle is maintained. In the world coordinate system, the ego vehicle's front-end position and velocity are presented as x_e and v_e, respectively, while the leading vehicle's rear-end position and velocity are presented as x_l and v_l, respectively. In this case, the relative distance, d_x, and relative velocity, v_x, between the host vehicle and the leading vehicle can be calculated as $x_l - x_e$ and $v_l - v_e$, respectively (Figure 3).

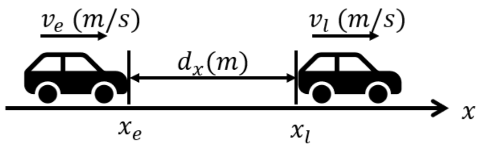


Figure 3 TTC outline figure

Therefore, the value of TTC, presented as t_x, can be expressed by the following equation (1).

$$t_x = -\frac{d_x}{v_x} = -\frac{x_l - x_e}{v_l - v_e} \quad (1)$$

In this study, the 2D extension of TTC, referred to as 2D TTC [8] was used. This index takes into account the positional relationship in a 2D space considering the vehicle's longitudinal direction (x-axis in the vehicle coordinate system) and lateral direction (y-axis in the vehicle coordinate system), whereas the previous TTC was calculated based on the positional relationship along the vehicle's longitudinal direction (x-axis) only. In the world coordinate system, the position of the ego vehicle is represented as (x_a, y_a), the position of the obstacle is represented as (x_b, y_b), the velocity of the ego vehicle is presented as (v_{ax}, v_{ay}), the velocity of the obstacle is presented as (v_{bx}, v_{by}), the size of the ego vehicle is presented as r_a, and the size of the obstacle is presented as r_b. Therefore, the relative distance is calculated as $\sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$, and the relative velocity is calculated as $\sqrt{(v_{bx} - v_{ax})^2 + (v_{by} - v_{ay})^2}$ (Figure 4).

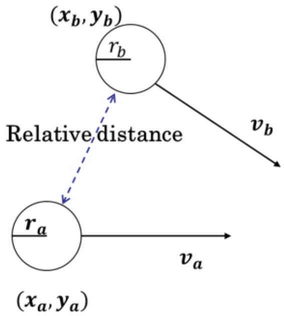


Figure 4 2D TTC outline figure

Therefore, the value of the 2D TTC is given by the following equation (2).

$$TTC_{2d} = \frac{\sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} - (r_a + r_b)}{\sqrt{(v_{bx} - v_{ax})^2 + (v_{by} - v_{ay})^2}} \quad (2)$$

Pedestrians with a 2D TTC value equal to or greater than a certain threshold are considered not requiring avoidance and are therefore excluded from the velocity obstacle calculations.

2.2.2. Velocity selection based on vehicle motion constraints

The second point is addressing vehicle motion constraints. Unlike mobile robots, automobiles have significant motion constraints, so an extension was made to select velocities achievable by car steering.

First, assuming the vehicle is either moving straight or performing a steady-state circular motion, the trajectory is calculated [9]. The illustration of this is shown in the figure below (Figure 5). Let θ be the angle between the X-axis and the vehicle's longitudinal direction (yaw angle), β be the angle between the vehicle's direction of travel and its longitudinal direction (slip angle), r be the angle between the X-axis and the vehicle's direction of travel, and V be the vehicle's velocity. The slip angle is determined by Equation (3), and the angle between the X-axis and the vehicle's direction of travel is given by Equation (4).

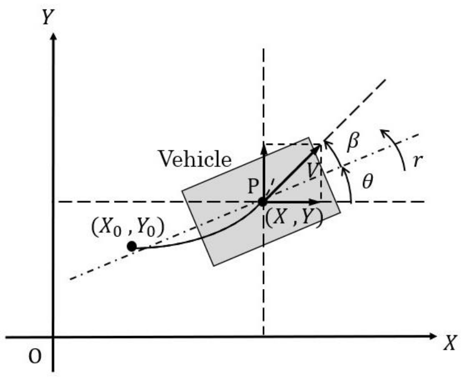


Figure 5 Vehicle trajectory prediction

$$\beta = \left(\frac{1 - \frac{m}{2l}\frac{l_f}{l_r K_r} V^2}{1 - \frac{m}{2l^2}\frac{l_f K_f l_r K_r}{K_f K_r} V^2} \right) \frac{l_r}{l} \delta \qquad (3)$$

$$r = \left(\frac{1}{1 - \frac{m}{2l^2}\frac{l_f K_f - l_r K_r}{K_f K_r} V^2} \right) \frac{V}{l} \delta \qquad (4)$$

Therefore, the trajectory of the moving vehicle's center of gravity is considered. The positions and velocities of the vehicle and the obstacle in the world coordinate system are shown in the figure below (Figure 6). Let (x_e, y_e) and (v_{ex}, v_{ey}) represent the position and velocity of the vehicle, respectively, and (x_p, y_p) and (v_{px}, v_{py}) represent the position and velocity of the obstacle. The obstacle is assumed to move in a straight line at a constant speed, and its trajectory is predicted.

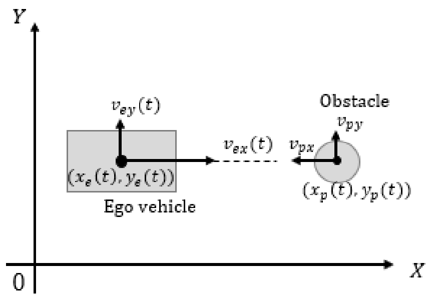


Figure 6 Each variable in the world coordinate system

Given the initial position (X_0, Y_0) and the initial yaw angle θ_0, the position (X, Y) and the yaw angle θ at any given time t can be calculated using equations (5), (6), and (7).

$$X = X_0 + V \int_0^t \cos(\beta + \theta)\, dt \qquad (5)$$

$$Y = Y_0 + V \int_0^t \sin(\beta + \theta)\, dt \qquad (6)$$

$$\theta = \theta_0 + \int_0^t r dt \qquad (7)$$

The actual calculation process is outlined below.

Step1 Provide steering angle for ego vehicle.

Step2 Calculate β and r using the steering angle from Step1 according to equations (3) and (4).

Step3 Using equations (5) and (6) from Step2, calculate the reachable points.

Step4 Calculate the velocity required to reach the speed in Step3.

Step5 Convert the velocity in Step4 to the velocity within the VOmap.

Step6 Prepare a map similar to VOmap, and substitute the value of the steering angle at the velocity position in Step5.

Step7 Repeat Step 1 to 6.

Here is the illustration of these steps (Figure 7).

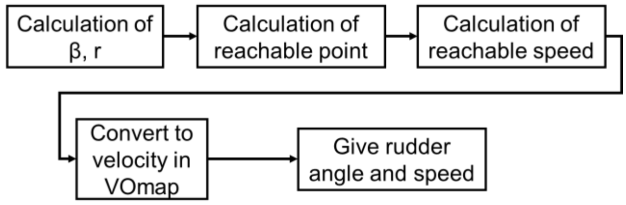


Figure 7 Outline of trajectory prediction

However, there is a difference when combining the VO method with trajectory prediction. VO assumes both the ego vehicle and the obstacles move in constant linear motion and generate avoidance paths based on the assumption that the current velocity will continue. On the other hand, trajectory prediction for constant circular motion involves nonlinear motion with constant translational and rotational velocities, making it incompatible to directly apply the VO method.

Therefore, in this case, an attainment point based on the current velocity that would result in a constant circular motion was assumed. The line connecting the starting point and the attainment point was approximated.

The generated trajectory prediction points for the ego vehicle are shown in the figure below (Figure 8).

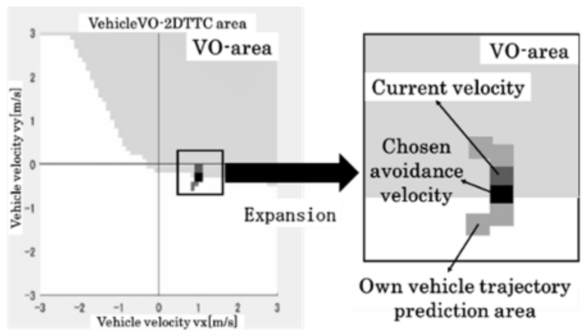


Figure 8 Example of trajectory prediction

2.2.3. Collision detection based on the vehicle contour

The third point addresses the issue of narrow road widths. In the VO approach, both pedestrians and ego vehicles are approximated as circles for collision detection. However, this approximation can result in significant differences between the actual collision detection and the circle approximation. Using circular approximation often leads to unnecessarily large avoidance routes.

According to the Japanese Cabinet Office, a "community road width" is defined as a road with a carriageway width of less than 5.5 meters in urban areas. If the avoidance route is too large, there is a risk of veering off the road or even the possibility of being unable to avoid the obstacle altogether. To achieve a compact and safe avoidance route, a collision detection method that takes into account the vehicle's contour was adopted (Figure 9).

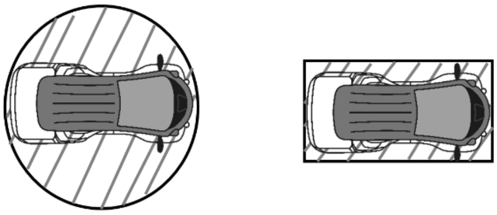


Figure 9 Comparison between circle approximation and contour-based methods

By utilizing the vehicle's contour, performing VO calculations from the four corners of the vehicle to the obstacles. Figure 10 shows velocity obstacle area. The original VO area is shown in gray and extended method area shown in black. This shows that it can be observed that extended method results in a smaller area compared to the original method. This approach allows for more precise collision detection and avoidance planning, taking into account the actual shape and size of the vehicle.

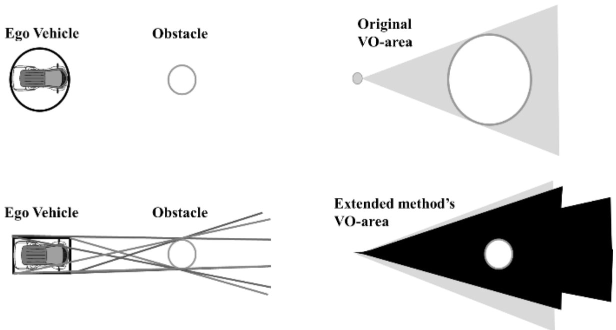


Figure 10 Caomparison of calculation methods between original VO and extended method

When actually performing circular approximation in a scenario simulating a community road environment, as shown in Figure 11, attempting to avoid obstacles leads to deviating from the lane due to the constraints of the road width. Therefore, by performing calculations from the four corners of the vehicle, it becomes possible to avoid obstacles while staying within the lane width, as shown in Figure 12. This approach ensures that the vehicle navigates safely within the available space and minimizes the risk of deviating off the road.

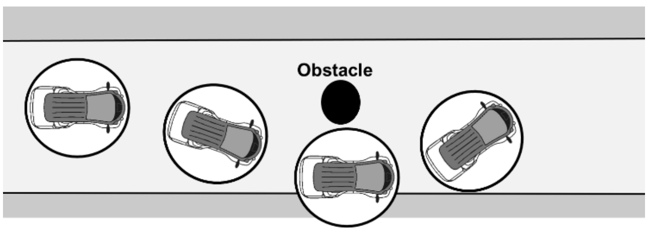


Figure 11 Example of avoidance using circle approximation

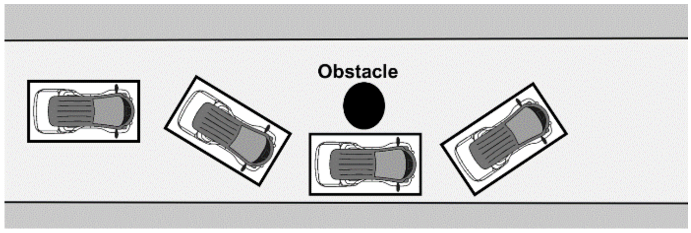


Figure 12 Example of avoidance using vehicle contour

3 SIMULATION EXPERIMENT AND RESU LTS

3.1 Experimental conditions

Using the VO-drive method proposed in this study, a simulation experiment was conducted to compare it with the original VO (Velocity Obstacle) method.

The simulation was performed using MATLAB. A time step of 0.1 seconds was used, and the initial value of the vehicle speed was set to 15 km/h. When the path generation method instructed a specific speed, it followed the generated speed by each route generation accordingly.

For VO-drive, the threshold for 2D TTC was set to 6 seconds as a limitation condition for the target obstacle. Only obstacles with TTC values below this threshold are considered for avoidance. And, the attainment point was defined as the point reachable 0.1 seconds later. For obstacle behavior prediction, the walking speed of the tracked obstacle was calculated using a Kalman filter, and the predicted position was calculated based on the calculated walking speed [10].

3.2 Comparison between VO-drive and VO

First, a comparison was conducted between the original VO and the extended VO-drive to confirm the effectiveness of the extension.

A comparative experiment was conducted on the limitation of the target obstacle. The results are shown in Figure 13. The ego vehicle starts from the position (-15,0) in the world coordinate system and moves straight at a speed of 15 km/h. An experimental scenario was designed where the vehicle needs to avoid a stationary obstacle located at (15,0). From Figure 13, it can be observed that the extension leads to a delay in the timing of avoidance initiation.

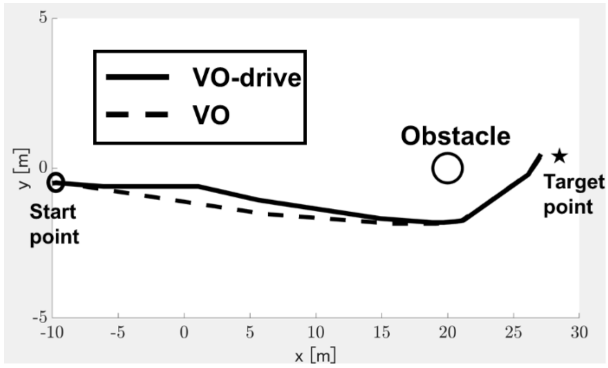


Figure 13 Comparison of target obstacle limitation

Second, a comparative experiment was conducted on velocity selection based on vehicle motion constraints. The results are shown in Figure 14. The ego vehicle starts from the position (-12,0) in the world coordinate system and moves straight at a speed of 15 km/h. Two obstacles are introduced, each moving diagonally at a speed of 0.5 m/s, creating a crossing scenario. From Figure 14, it can be observed that the proposed method achieves smoother and more efficient avoidance, while the original VO approach exhibits sharp turns.

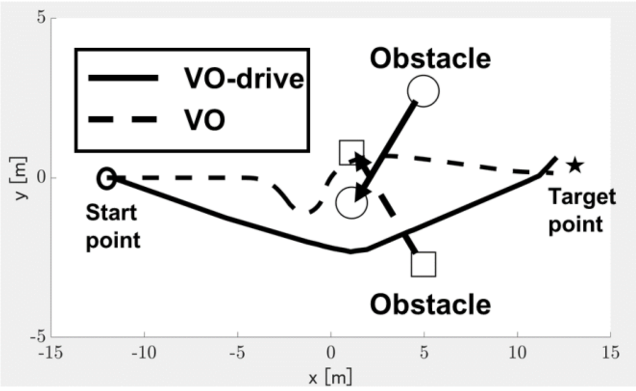


Figure 14 Comparison of vehicle motion constraints

Lastly, a comparative experiment was then conducted on collision detection based on the vehicle contour. The results are shown in Figure 15. The ego vehicle starts from the position (-12,0) in the world coordinate system and moves straight at a speed of 15 km/h, aiming to compare whether it can navigate between the obstacles. From Figure 15, it can be observed that the conventional circle approximation method fails to navigate between the obstacles and takes a significantly larger avoidance path, while the extension allows the vehicle to pass between the obstacles successfully.

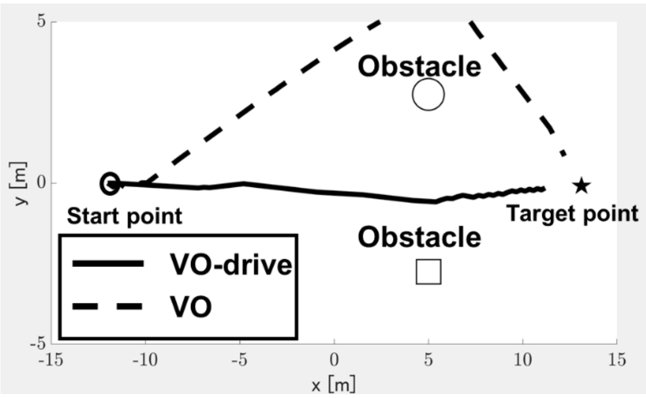


Figure 15 Comparison of collision detection

4 CONCLUSION

By conducting a comparison of methods using simulations, it was possible to demonstrate the usefulness of the proposed VO-drive approach for autonomous driving on urban roads. After conducting simulations, actual experiments were carried out using a vehicle, and successful avoidance routes were performed.

5 FUTURE WORKS

Currently, the real vehicle experiments are still ongoing. In the future, it is plan to conduct real vehicle experiments, increasing the vehicle's speed and introducing moving obstacles to create an environment closer to real-life urban road conditions. In addition, by having pedestrians actually walk, it is aim to realize avoidance that makes them feel safe.

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism, Community road traffic safety measures portal, https://www.mlit.go.jp/road/road/traffic/sesaku/anzen.html.

[2] D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” IEEE Robot. Autom. Mag., vol. 4, no. 1, pp. 23-33(1997).

[3] Michael T. Wolf, Joel W. Burdick, “Artificial Potential Functions for Highway Driving with Collision Avoidance”, IEEE International Conference on Robotics and Automation (2008).

[4] Fiorini P, Shiller Z, “Motion Planning in Dynamic Environments Using Velocity Obstacles”, The International Journal of Robotics Research, vol.17, no.7, pp.760-772 (1998).

[5] Mengyuan Gu, Yunzhi Huang, “Dynamic Obstacle Avoidance of Mobile Robot Based on Adaptive Velocity Obstacle”, 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC), IEEE (2021).

[6] Jeon, J. D., Lee, B. H. “Ellipse-based velocity obstacles for local navigation of holonomic mobile robot”, Electronics Letters, 50(18), 1279-1281 (2014).

[7] B. Lee, J. Jeon, J. Oh, “Velocity Obstacle based local collision avoidance for a holonomic elliptic robot”, Autonomous Robots, vol.41, no.6, pp.1347-1363 (2017).

[8] Yuto Yada, Shunsuke Michita, Seiji Komiya and Toshihiro Wakita, “Path Planning considering Pedestrian Characteristics”, International Workshop on Informatics (IWIN 2022).

[9] Masato Abe, Automotive Vehicle Dynamics Trajectory and Applications, Tokyo Denki University Press (2012).

[10] Takumi Goto, “Pedestrian behavior prediction and forecast circle generation using Kalman filter”, Kanagawa Institute of Technology, Undergraduate Thesis (2019).

Title: Autonomous driving on community roads using small mobility -Route generation using trajectory prediction and 2D TTC-
Author: Takumi Seita, Shunsuke Michita, Seiji Komiya, Toshihiro Wakita
Corresponding Author: Takumi Seita
Tel: 046-291-3091
Fax:046-242-6806
E-mail: s2384004@cco.kanagawa-it.ac.jp

A LiDAR Point Cloud Data Fusion from Connected Autonomous Vehicles to Reduce Network Traffic Congestion at Server

Pasupuleti Haripriya, Yoshinao Wada, Kenjiro Yamada, Takuro Sone, and Tomoya Kitani

Graduate School of Integrated Science and Technology, Shizuoka University, Japan
{haripriya-p, y-wada, k-yamada, t-sone, t-kitani}@kitanilab.org

Abstract - In a road traffic setting, autonomous vehicles continuously generate LiDAR point cloud data, which is useful in generating and maintaining dynamic map of on-road objects. This data can be used to form a Static Object Map (SOM) that identifies static objects that are useful in identifying irregular objects like road defects, road construction and so on. However, transmitting large raw LiDAR data poses challenges due to server-level network congestion. This paper proposes a solution by developing communication and upload protocols, along with an aggregation algorithm, to remove redundant data before transmission. Simulations demonstrate significant reductions in data size, alleviating network congestion at the server, and improving efficiency compared to conventional methods.

Keywords: LiDAR point cloud data, dynamic map, network traffic congestion, communication protocol, aggregation algorithm

1 Introduction

This paper explores key components in developing autonomous driving applications, which includes dynamic maps, high precision 3D maps, LiDAR sensor technology, and efforts to reduce data size through V2V communication, difference extraction and aggregation. It emphasizes the importance of Static Object Map (SOM) for not only identifying shape of static objects like traffic poles, road shape and so on, but also helpful in recognizing irregular objects in the road environment like road construction, road defects and so on. The focus is to collect the LiDAR point cloud data acquired by autonomous vehicles at server by updating SOM in autonomous vehicles based on the frequency of applications, such as static, semi-static, and semi-dynamic information layers of dynamic map[1], as shown in Fig. 1. The paper delves into LiDAR data, particularly its 3D perception capabilities, and the challenges possessed by the large size of LiDAR data collected by autonomous vehicles. To address this, we proposed a method for V2V communication of LiDAR point cloud data, utilizing a voxel-based map of SOM and acquired LiDAR point cloud data by autonomous vehicles to reduce the data size. Evaluations demonstrate a significant reduction in unstructured LiDAR data, alleviating network traffic congestion at the server with the help of differece extraction, aggregation and shared LiDAR point cloud data among neighbouring vehicles. This approach proves more efficient compared to conventional methods of individually sending pre-processed LiDAR data to the server.

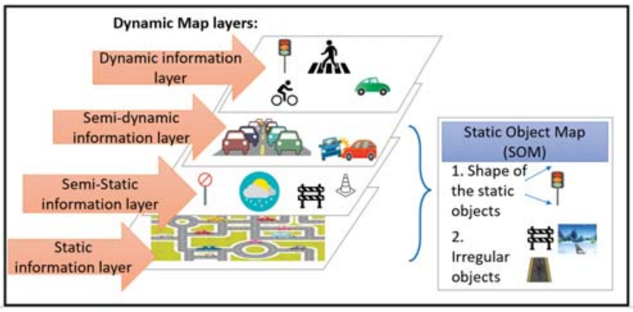


Figure 1: Dynamic map layers and contents of SOM

2 Related work

In our research on road traffic networks, we aim to alleviate congestion and reduce server load through communication protocols and aggregation algorithm.

Shagufta Ali et al. [2] discuss wireless communication technologies in V2V communication, including DSRC, 4G-LTE, and 5G. DSRC excels in mobility but has scalability issues, while 4G-LTE faces challenges in high traffic conditions. 5G-V2X, leveraging fifth-generation technology, offers enhanced communication and coverage.

Chen Qi et al. [3] proposed a LiDAR-based 3D object detection method with data fusion for cooperative perception. Hongyu Li et al. [4] introduced a two-phase point cloud registration mechanism, and Anand et al. [5] have researched the issue of the massive size of point cloud data produced by LiDAR and have suggested Octree-based compression technique. In point cloud registration, the Iterative Closest Point (ICP)Algorithm [6] and Normal Distribution Transform (NDT) Algorithm [7] are widely used, but challenges persist in accurately registering point clouds from vehicles due to high outlier ratios caused by different views and obstructions [4].

3 Target problem

3.1 Target system architecture

In this section, we describe the target VANET model. Our target system architecture consists of a road environment scenario with cars, buildings, roads, and other infrastructures. We consider each car is an autonomous driving vehicle with a LiDAR sensor that can communicate with other vehicles and infrastructure, and we also assume that every car has same version of SOM (in voxel form) stored in it's database. The target system architecture is illustrated in Fig. 2. Initially, autonomous vehicles acquire LiDAR point cloud data. Subsequently, this newly acquired LiDAR point cloud data is converted into voxels and compared to the stored Static Object Map (SOM) in the vehicle to identify differences, termed as ”Difference Extraction.” All other autonomous vehicles undergo a similar process. The LiDAR point cloud data extracted during difference extraction is exchanged between vehicles for aggregation, aiming to eliminate redundant data. The final aggregated data is then uploaded to the server.

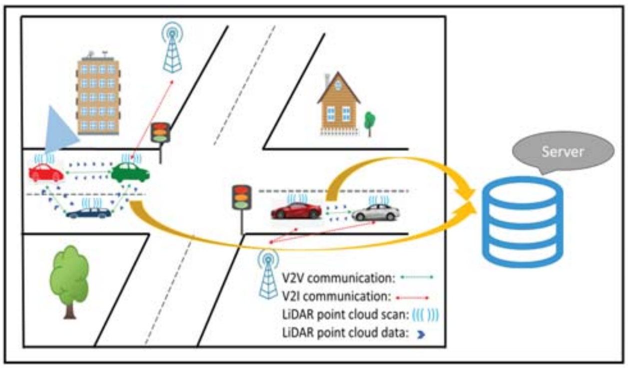


Figure 2: Target System Architecture

3.2 Assumptions

3.2.1 Road network

We assume the following road traffic network as shown in Fig. 3.

Vehicle ID’s: sender vehicle (V_1), receiver vehicle (V_2)

Intersection ID’s: {J1, J2, J3, J4}

Edge ID’s: {E1, -E1, E2, -E2, E3, -E3, E4, -E4, E5, -E5, E6, -E6, E7, -E7}

V_1**’s route path (R1):** {-E0, -E1, -E2, -E3, E5}

V_2**’s route path (R2):** {-E0, E4, E5}

Overlapped route path (R3): {-E0, E5}

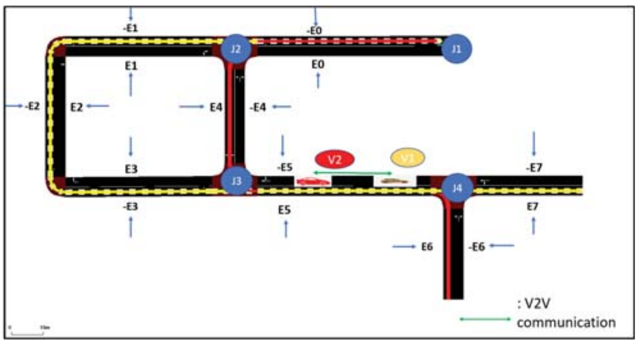


Figure 3: Road Network Scenario

3.2.2 Communication technologies

Various communication technologies are utilized in vehicular communication networks. For the communication between vehicles, DSRC technology can be used as its range is limited, and for the communication between vehicles and servers, cellular technology is capable of transferring data to remote areas such as servers.

3.3 Target object model of urban space

3.3.1 Data structure of target LiDAR device

In this experiment, we used LIVOX AVIA as LiDAR sensor[1]. The specifications of the LIVOX AVIA LiDAR sensor are described in the Table. 1 as shown below. LIVOX AVIA can

Table 1: LiDAR sensor specifications

Specification	Details
Field of View(FOV)	Non-repetitive $70.4° \times 77.2°$
Point Rate	240Kps
Range Accuracy	2cm
Angular Precision	$< 0.05°$
Return Mode	First and Strongest

measure 240,000 points per second, and the bit rate is calculated from this value. To consider the bit rate of specific data, we will refer to the LVX format of point clouds that can be obtained from LIVOX AVIA.

According to “LVX Specifications” [2], a document published by LIVOX, the LVX format stores the values of the Cartesian coordinates of the data in 4-byte increments. Since there are three items (X, Y, Z), this requires a total of 12 bytes of data. In addition, in LIVOX AVIA, a Time Stamp value is assigned to each of the 96 points, requiring a data volume of 8 bytes. Therefore, there are 240,000 points with 12 bytes of coordinate value (X, Y, Z) data and 25,000 points with 8 bytes of time information (TimeStamp) data, resulting in a total bit rate of approximately 2.94MB/S.

The fact that the coordinate values (X, Y, Z) are expressed in 4 bytes each and the time information in 8 bytes is a general value, not limited to the LVX format of point clouds that is acquired by LIVOX AVIA. This data volume is also used in the LAS format [9], which is one of the point cloud file formats. The LAS format is also used in “VIRTUAL SHIZUOKA” [10], which is an open data point cloud database published by Shizuoka Prefecture and is a common point cloud format.

If the file format obtained from LiDAR was in LAS format, there would be 240,000 points with a total of 20 bytes of X, Y, Z, and $TimeStamp$, resulting in a bit rate of approximately 4.58MB/s.

[1] Avia LiDAR sensor, https://www.livoxtech.com/avia

[2] LVX Specifications V1.1.0.0, https://www.livoxtech.com/3296f540ecf5458a8829e01cf429798e/downloads/Livox\%20Viewer/LVX\%20Specifications\%20EN_20190924.pdf

3.3.2 Voxelization

Voxelization refers to the process of converting 3D objects or scenes into a structured grid-like representation into cubic volumetric elements called as voxels. Voxels are similar to pixels in images, but voxels deal with 3D representation and visualization of objects or scenes, whereas pixels deal with 2D images. According to [8], LiDAR point cloud data have the most accurate 3D description, but its unstructured data organization can cause data density variation, noise & outliers, occlusions & shadows, misalignment & registration, and segmentation & object extraction. However, voxel grids of LiDAR point cloud data can represents not only organized spatial information from point clouds but also provides more efficient storage, processing, and analysis. An example of voxelized urban area is shown in Fig. 4

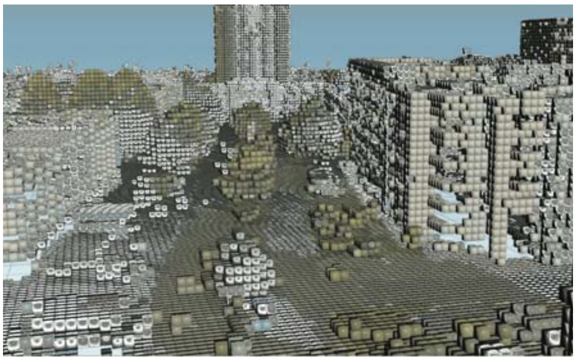


Figure 4: An example of voxelized urban area

3.3.3 SOM (Static Object Map)

A High-precision 3D map typically contains static information such as road link information, lanes, guardrails, road signs, and pedestrian crossings at more accurate locations. It gives the position of the road equipment and infrastructure, but it lacks the shape information of the static objects. So, we proposed a target object model of urban space called as a Static Object Map (SOM) that provides the structure of static objects in the static information layer. We extended the SOM mechanism to semi-static and semi-dynamic information applications in identifying irregular objects such as road construction, flooded areas, road defects, and so on. In this paper, we represented the static object map (SOM) as a set of structured voxel grids with static, semi-static and semi-dynamic information, which refers to parked cars, that are indicated by red circle as shown in Fig. 5.

4 Proposed method

4.1 Overview

Figure 6 shows the overall system flow of the proposed method to maintain and update SOM by acquiring LiDAR point cloud data from autonomous vehicles with less communication traffic at the server.

Figure 5: An example of voxelized SOM

4.2 Sensing and Data Compression phase

We assume that all the autonomous vehicles and server has the same version of SOM in voxel form. In sensing and data compression phase, all the vehicles acquire LiDAR point cloud data along it's travelled route path and convert into voxels through voxelization process. Then each vehicle performs difference extraction. In which, the vehicle extract differences with the stored SOM and acquired LiDAR point cloud data. The detailed explanation of voxelization process and difference extraction is explained in section 4.5. If there is any difference extraction, then the vehicle triggers control message exchange protocol. Otherwise, there will be no communication established between vehicles.

4.3 Design of Control Message Exchange protocol

The vehicles in the road traffic environment communicate with each other within the IEEE 802.11p DSRC protocol. After the sensing and data compression phase, if there is any difference extraction available at the sender vehicle, then the control message exchange protocol is performed. The sender vehicle broadcasts the difference extraction LiDAR point cloud data Edge ID's of road_id list. The vehicle ID's, edge ID's and road_id list of route paths are derived in section 3.2.1. The flow of the control message exchange protocol is shown in Fig. 7.

4.4 Design of a LiDAR point cloud data exchange decision algorithm

During the sensing and data compression phase, the receiver vehicle is also able to extract the difference in SOM and acquired LiDAR point cloud data. The receiver vehicle receives and checks the control message. The receiver vehicle compares own road_id list with sender vehicle's road_id list to identify any overlapping road_id's. Then the receiver vehicle requests the sender vehicle for LiDAR point cloud data of the overlapping road_id list as shown in Fig. 8. Next, the sender vehicle sends only the difference extraction LiDAR point cloud data of overlapping road_id list to receiver vehicle. We consider IEEE 802.11p communication technology to exchange LiDAR point cloud data between vehicles.

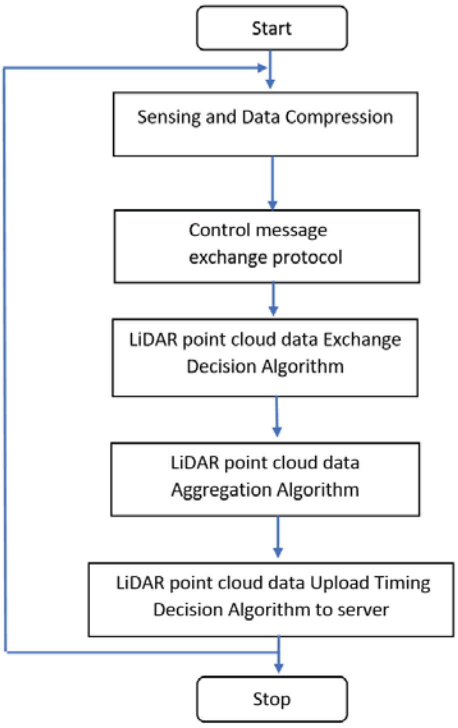


Figure 6: A flowchart representing the overview of the proposed method

4.5 Design of LiDAR point cloud data aggregation algorithm

To aggregate point clouds measured from multiple vehicles, data reduction and fusion must be performed due to limited line speeds. Algorithms for data reduction include voxelization and comparison with SOM, while point cloud synthesis is an algorithm for fusing multiple point clouds.

In this study, voxelization is used to reduce the amount of data in a point cloud and makes it easier to exchange point clouds. Voxelization reduces the coordinate resolution of a point cloud. By reducing the resolution, the absolute number of points in a point cloud can be reduced by combining neighboring points into a single point cloud. It is possible to eliminate duplicate point clouds. However, doing so may result in a decrease in accuracy.

Let us consider the specific amount of point cloud deletion on the surface area. Suppose that LiDAR, which can measure 240,000 points per second, evenly measures a surface with an area of $24/m^2$. The original point density was one point per cm^2, but the data size of the point cloud was reduced by constructing voxels with a side of 2.5 cm. This means that for $1/m^2$, there are 1600 points, which means that for $24/m^2$, there are 38,400 points. In other words, for a surface filled with a point cloud of cm^2, the point density can be reduced to about 1/6 with a side of 2.5cm. We regard this process as voxelization.

To give you an idea of the voxelization process, we prepared a point cloud of the area near the fountain on the Hamamatsu campus of Shizuoka University. The original point cloud data visualized by Houdini which is a 3D computer graphics (CG) software is shown in Fig. 9.

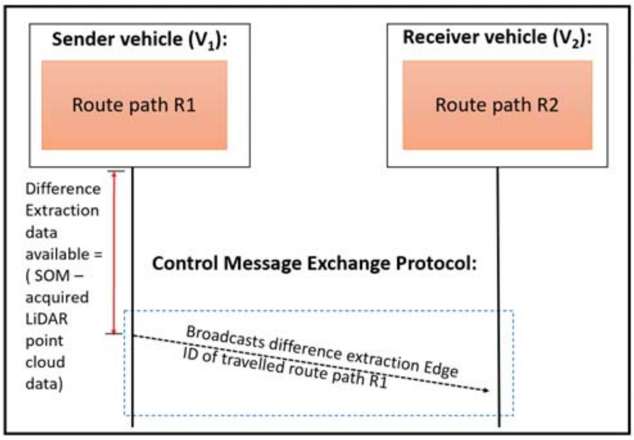


Figure 7: Control Message Exchange Protocol Flow

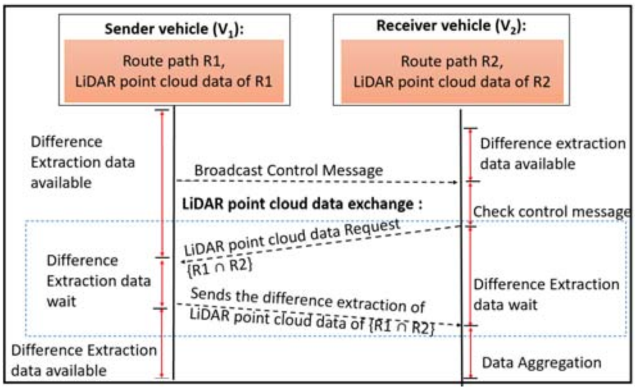


Figure 8: LiDAR Point Cloud Data Exchange Decision Algorithm Flow

The point cloud in Fig. 10 is voxelized every 25 cm by Houdini.

In this case, voxelization reduces the amount of data since any number of points in a 25 cm square can be represented by a single cube.

In addition to voxelization, difference extraction from the SOM is used as an algorithm for data reduction. The original point cloud is as shown in Fig. 11 (a). Figure 11 (b) shows the extracted differences from the SOM. By extracting the differences, point clouds that overlap with the SOM can be found. Therefore, point cloud data can be reduced.

Point cloud merging is a method of aligning and merging point clouds measured from multiple vehicles. When point clouds are merged, they combine into one single point cloud. This allows multiple point clouds to be brought together and consolidated into a master vehicle. Specific methods for merging point clouds include the ICP algorithm, which matches the nearest points of two point clouds.

To compare the SOM and the synthesized point cloud with minimal error, we did not use any synthesis algorithm that is dependent on initial values or introduces errors but simply captured the two point clouds from the same location and superimposed them to reproduce the synthesized point cloud. Since the ICP algorithm and other positioning methods do

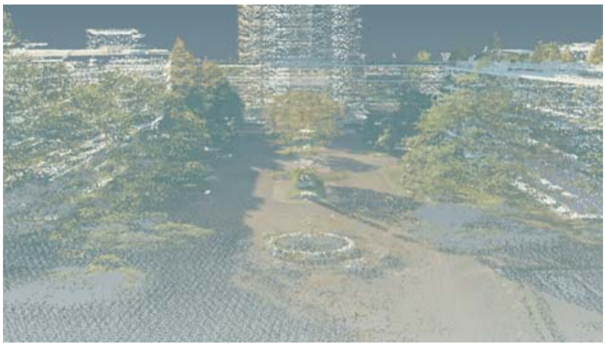


Figure 9: Raw LiDAR point data

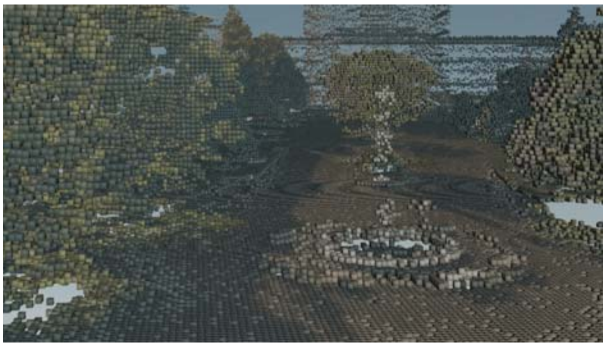


Figure 10: 25cm-cube Voxelized data of Fig. 9

not remove duplicate point clouds, the point cloud synthesis by superposition in this experiment does not differ from general registration methods, such as the ICP algorithm in terms of the amount of data.

4.6 Design of LiDAR point cloud data upload timing decision algorithm

At this stage, the receiver vehicle contains the final aggregated data of difference extraction LiDAR point cloud data of both sender vehicle and receiver vehicle. Finally, the aggregated data is sent within a timeout of (t_0) to the server as shown in Fig. 12, via a cellular communication technology due to its large bandwidth and wide coverage area. The reason for timeout (t_0) is because the maximum update frequency varies for static, semi-static and semi-dynamic applications. If the aggregated data is not uploaded to the server within the designated timeout (t_0), it will be considered as expired and eventually be discarded. So, we set a timeout (t_0) to upload the aggregated data to server.

5 Evaluation

To evaluate the proposed method, we have conducted an experiment that involves the test data collected at Shizuoka University, Hamamatsu campus. The collected test data is processed through computer simulations. A detailed explanation of the experimental environment and configuration is given in the following subsections.

(a) before extracting the differences

(b) after extracting the differences

Figure 11: Extracting the differences between two point cloud data

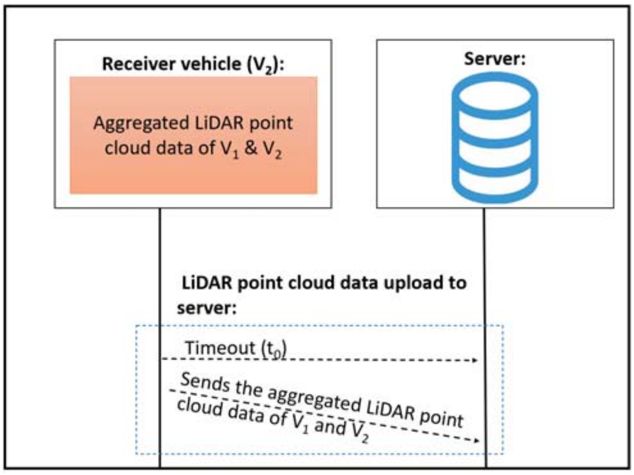


Figure 12: LiDAR Point Cloud Data Upload Timing Decision Algorithm Flow

5.1 LiDAR point cloud data aggregation

5.1.1 Experimental environment and configuration

We conducted experiments on voxelization and difference extraction from the SOM to decrease the data volume of the point cloud.

The scanning time was set to 10.6 seconds and incremented every 0.1 seconds for a total of 106-point cloud samples. The reason for using 0.1 seconds as the unit is that the FOV coverage of a typical LiDAR reaches its upper limit after 0.1 seconds. According to Fig. 13 in the LIVOX AVIA User Manual published by LIVOX [3], the scan pattern of LiDAR with the conventional mechanism shows that the FOV coverage reaches its upper limit at 0.1 seconds.

In this study, we utilized the LIVOX AVIA scan pattern, which differs from conventional patterns. To adapt to the conditions, we divided the point clouds into 0.1-second increments. Although LIVOX AVIA has the same pattern as the conventional scan pattern, it is difficult to adjust the angle of view and include the entire object because of the narrow FOV. Therefore, a non-repeating scanning method with a wide FOV coverage was adopted.

For the experiment, LiDAR was initially positioned in a stationary location, and point cloud V_0, which solely contains stationary objects, was created as a SOM. Additionally,

[3] LIVOX AVIA User Manual, `https://terra-1-g.djicdn.com/65c028cd298f4669a7f0e40e50ba1131/Download/Avia/Livox%20Avia%20User%20Manual%20202204.pdf`

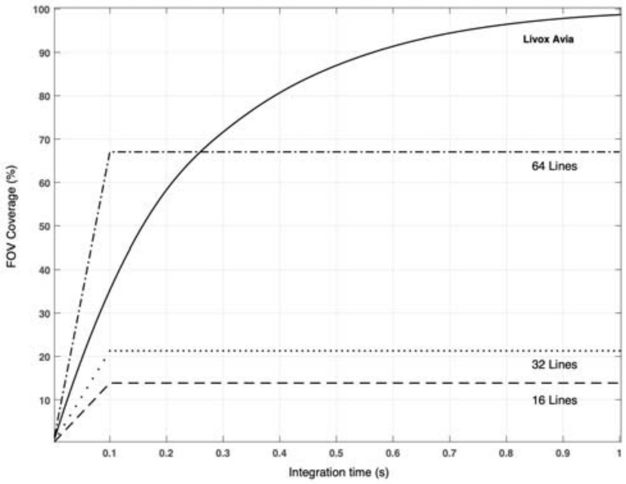


Figure 13: Scaninng time comparison (from LIVOX AVIA User Manual, p. 5)

LiDAR point clouds of V_1 and V_2 were prepared to include stationary objects that were not present in the SOM V_0. LiDAR point cloud V_2 includes irregular objects in addition to the stationary objects found in V_1. And we estimate that V_1 is a semi-static and V_2 is a semi-dynamic data. The environment of the experiment is as follows: an original LiDAR point cloud as SOM, LiDAR point cloud with a new box, and LiDAR point cloud with a new chair is shown in Fig. 14.

(a) V_0: original as SOM (b) V_1: new box adding to V_0 (c) V_2: new chair adding to V_1

Figure 14: Example measurement environment

The voxelization process utilized a LiDAR measurement error margin of 2.0 cm, with each voxel having a side length of 2.5 cm. The center of gravity of the point cloud existing within the 2.5 cm unit delimitation was used as the representative value of a single voxel.

The comparison was performed using an algorithm that removes all point clouds within 2.0 cm of the SOM (V_0) from V_1 and V_2.

The LiDAR has a measurement error of 2.0 cm. Any points beyond this distance can be considered non-existent in the SOM (V_0). In other words, there is a high possibility that objects that do not exist in the SOM are being measured, and differential detection of point clouds is possible.

Ideally, the experiment should be conducted while in a moving vehicle with LiDAR sensor to accurately measure the point clouds, which are closely linked to real-world traffic conditions. However, the point cloud difference extraction method used in this experiment is based on the SOM point cloud and uses a simple method of deleting the point clouds measured from vehicles that are within the threshold value. The point cloud gathered by LiDAR while in motion may contain inaccuracies from the IMU, GPS, and other sources, in addition to errors from the LiDAR sensor itself. Therefore, if a point cloud is extracted using a threshold value of 2cm measurement error of LiDAR along with the errors of IMU and GPS, it is likely that a considerable number of point clouds will be excluded.

Hence, for this instance, we immobilized both the LiDAR and the object, thereby constraining the error factor exclusively to the LiDAR itself. This allows us to estimate the amount of data to be deleted from the point cloud measured from the vehicle.

5.1.2 Acquired LiDAR point cloud data and aggregated results

First, each point cloud was voxelized. After performing voxelization, the average value of the total data count excluding point group V_0 (which represents the data for SOM) and the reduction ratio can be summarized as follows. To calculate the reduction ratio, we divided the number of points left after voxelization by 24,000, which represents the number of samplings done in a 0.1-second time frame. As shown in Table 2, voxelization reduced the point cloud by more than 27.0%.

Table 2: Results after conducting voxelization

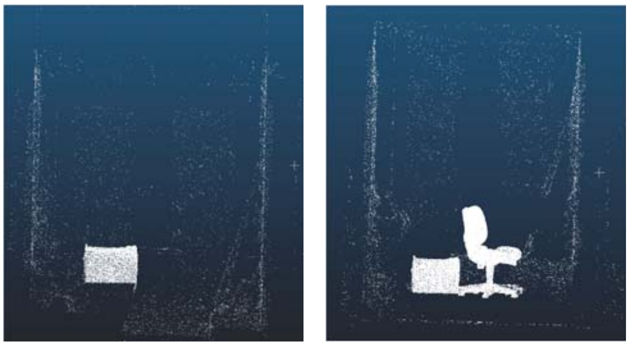
	remaining points	reduced ratio
voxelization V_1	17333	27.8%
voxelization V_2	17169	28.5%

Next, a comparison was made between the point groups V_1 and V_2 with V_0, which is the SOM, to extract the differences. The average of the total number of point clouds and the reduction ratio was as shown in the following Table 3. By comparing the voxelized point cloud with the SOM, we reduced the data amount by more than 95.0%. The reduction was calculated by dividing the number of voxelized points by the number of points remaining after comparing with the SOM. The results of the difference extraction between V_1 and V_2 using SOM can be seen in Fig. 15. To enhance a clear visible distinction, both V_1 and V_2 point clouds showcase a 10.6-second duration instead of the previous 0.1-second representation.

Table 3: Results after differences extraction

	remaining points	reduced ratio
diff extraction V_1	241	98.6%
diff extraction V_2	755	95.6%

From here, both point clouds(voxel) V_1 and V_2 are further compared and the remaining point clouds are extracted. The average of the total data amount and the reduction ratio was



(a) V_1 result (b) V_2 result

Figure 15: difference extraction result

as shown in Table 4 below. The number of point clouds that could be eliminated by performing the difference extraction between V_1 and V_2 exceeded 84.0%. The reduction ratio was calculated by dividing the number of points remaining after the comparison of V_1 and V_2 by the total number of points in V_1 and V_2 after the comparison with the SOM.

Table 4: Overlapped points of V_1 and V_2

	remaining points	reduced ratio
V_1 and V_2	154	84.6%

As a result, the total number of original point clouds, V_1 and V_2, consisting of 48,000 points, were reduced to 154 points through voxelization, SOM, and comparison with the aggregated point cloud. In other words, the data size could be reduced to about 0.3% of the original data size.

Note that voxelization and difference extraction depends on the size and shape of the object, so it is not possible to say that they reduce the specific data size. In this experimental data, the object volume occupied by the SOM is small, and points within 2.0 cm of the object being compared were removed during difference extraction. Therefore, the data size may tend to be smaller than the actual voxelization and difference extraction performed.

In fact, the remaining points should be 241 as a result of Table 4, but are smaller than 241 for the aforementioned reasons.

In addition, the point cloud aggregation algorithm used in this experiment is a simple superposition. Therefore, in a real environment, there is a possibility that errors may occur in the superimposed point cloud, and the detection distance and other factors must be adjusted when performing difference extraction.

6 Conclusion

Our proposed method not only reduces the network traffic congestion, that is the communication load at the server but also reduces the computational load on the server parallelly. We fuse LiDAR point cloud data to not only remove the duplicate, redundant data but also increase the accuracy of the collected data from the neighboring vehicles and piece the missing data by exchanging data between other vehicles. We believe that in future work, SOM can be further extended to the dynamic information layer of a dynamic map. In our proposed method, we only considered two vehicles' data, but in future work, this scope can be expanded to identify the right and wrong information which is an error data by judging the collected data from more than two vehicles. Moreover, we collected the test data from the University campus. As a future work, We plan to expand this model by collecting the test data from the real road scenario with 'n' number of vehicles.

REFERENCES

[1] Autononous Driving Lab's Editors, "What is Dynamic map? It is useful for Autonomous Driving (2023)," `https://jidounten-lab.com/y_6059` (in Japanese)

[2] Ali, Shagufta, "Vehicle to Vehicle communication," 10.13140/RG.2.2.24951.88487, 2019

[3] Chen, Q., Tang, S., Yang, Q., & Fu, S, "Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds," *Proc. in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, (pp. 514-524). IEEE, July 2019.

[4] H. Li, H. Liu, H. Lu, B. Cheng, M. Gruteser and T. Shimizu "Bridging the Gap Between Point Cloud Registration and Connected Vehicles," *Proc. in IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 178-192, doi: 10.1109/OJVT.2022.3165930, 2022.

[5] B. Anand, V. Barsaiyan, M. Senapati and P. Rajalakshmi, "Real Time LiDAR Point Cloud Compression and Transmission for Intelligent Transportation System," *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia*, pp. 1-5, doi: 10.1109/VTCSpring.2019.8746417, 2019.

[6] P. J. Besl and N. D. McKay "Method for registration of 3D shapes," *Proc. in Proc. Sensor Fusion IV: Control Paradigms Data Structures, International Society for Optics and Photonics*, vol. 1611, pp. 586 – 606, 1992.

[7] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *Journal of Field Robotics,*, 24(10), 803-827, 2007.

[8] Xu, Yusheng, Xiaohua Tong, and Uwe Stilla, "Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry," *Automation in Construction,*, 126 (2021): 103675, 2021.

[9] LAS Specification 1.4 - R15, *LAS Specification* `https://www.asprs.org/wp-content/uploads/2019/07/LAS\_1\_4\_r15.pdf`

[10] "VIRTUAL SHIZUOKA," Shizuoka Prefectural Government, Japan, `https://www.pref.shizuoka.jp/machizukuri/1049255/1052183.html` (in Japanese)

[11] Bentley. J. L , "Multidimensional binary search trees used for associative searching" , Commun. ACM, vol. 18, number. 9, pp. 509 – 517, 1975

Session 3:
Network and Security
(Chair: Tomoya Kitani)

A Scheduling Method for LoRaWAN Assuming a Mixture of Incompatible Devices

Tabasa Ishibashi† and Takuya Yoshihiro‡

†Graduate school of Systems Engineering, Wakayama University, Japan
‡Faculty of Systems Engineering, Wakayama University, Japan
†e-mail : s246015@wakayama-u.ac.jp
‡e-mail : tac@wakayama-u.ac.jp

Abstract -
IoT(Internet of Things) is being used as a system to collect and utilize the vast amount of data from numbers of things. LPWA (Low Power Wide Area) is payed attention as a communication technology for IoT and amang them, we use LoRaWAN. LoRaWAN has a problem of increasing frame collisions as number of devices in the network increase. Frame collisions happen when several devices transmit messeages to base station at the same time and are more likely to occur as the number of devices increases. Existing studies have proposed a method for scheduling the transmission timing of devices to solve the problem. However, it is difficult to transit to a new method because existing studies require all devices in the network to support the new protocol. Therefore, in this paper, we present a scheduling method for LoRaWAN to support a mixture of compatible and incompatible devices by extending the MAC protocol of LoRaWAN. The mechanism estimates the time of incompatible devices transmitting messages and schedule transmissions sent by compatible devices avoiding the time. The proposed method thus aims to improve the scalability of the whole network.

Keywords: IoT, LPWA, LoRaWAN, scheduling, communication protocol

1 INTRODUCTION

Recently, IoT(Internet of Things) is being called as a system to collect and utilize a vast amount of data from many devices. IoT devices require low energy consumption to run on battery power and long-distance wireless communication to perform small and infrequent packet exchanges. This type of wireless communication technology is generally called as LPWA (Low Power Wide Area), which includes several specific standards such as NB-IoT [1], SIGFOX [2], and LoRaWAN [3], etc. Among them, LoRaWAN is suitable to study new protocols because of its open deployment policy. LoRaWAN is a communication protocol established by LoRa Alliance, has an open specification, and uses unlicensed bands. In the current protocol, LoRaWAN uses an ALOHA-based protocol, which does not schedule transmissions. Class-A devices send a data frame immediately without carrier sensing when a data frame to send is ready. If there are a small number of devices in a network, the probability of a frame collision is low. However, if the number of devices in a network increases, the probability of frame collision increases because of the high density of transmissions. Due to the nature of the ALOHA-based MAC protocol, the scalability of LoRaWAN is limited [4].

Therefore, several methods have been proposed for LoRaWAN to optimize communication networks by scheduling data frames. In these methods, all devices in the network are required to implement the new protocol, so that it is hard to transit from the current protocol to the new protocols.

In this paper, we propose a scheduling method for LoRaWAN to support a mixture of compatible and incompatible devices by extending the MAC protocol of LoRaWAN. The system first estimates the time when incompatible devices send their data frames based on the records of past data frame receptions. Next, the timing and the channel of transmissions of compatible devices are scheduled by systems and are notified to the devices with ACK frames. The proposed method thus aims to improve the scalability of the whole network by controlling the transmission timings of devices to avoid collisions.

The remainder of this paper is organized as follows. In Section 2, we give an introduction to LoRaWAN networks. Section 3 describes the details of our scheduling scheme. Finally, Section 4 concludes the paper.

2 LORAWAN

In the LoRaWAN protocol, many end devices communicate with NS (Network Server) through GW (Gateway). GW and end devices in a network of LoRaWAN make the star topology of the network. The devices in a network seen are listed as follows. End devices are placed in the network to send data frames to GW. GW receives data frames from end devices, then forwards them to NS as IP packets. NS manages end devices in the network according to the LoRaWAN protocol.

In the LoRaWAN network, end devices communicate with NS through GW. The physical layer on which end devices and GW communicate is LoRa modulation, and the physical layer on which GW and NS communicate is Ethernet. All communications can be bi-directional, but mainly most of them are uplink transmissions sent from end devices to NS. The timing at which end devices listen to the channel is defined, at which GW transmits downlink messages to end devices.

In the LoRaWAN protocol, the operation of end devices is classified into three types, class-A, class-B, and class-C. Class-A is the standard class, in which all end devices must operate. End devices in class-A communicate with GW interactively in which each end device's uplink transmission is followed by two downlink receive windows. At each of the downlink receive windows, GW sends an ACK frame when confirmation is required. NS is allowed to send a MAC command with the ACK frames to manage end devices. Class-A operation has the lowest power consumption among all classes, in which end devices of class-A run on battery power so that they sleep most of the time. Class-B is the operation that extends class-A, and end devices of class-B sleep most of the time similarly to class-A. In addition to Class-A, Class-B operation forces end devices to receive downlink communications at regular intervals, which allows NS to periodically control end devices, thereby the power consumption with class-B operation is higher than class-A operation. Thus, class-B operation seems to assume that the end devices are equipped with a solar panel at least. Class-C is the operation that further extends class-B. End devices of class-C always listen to the channel, thereby class-C operation requires the highest power consumption. Thus end devices of class-C are assumed to have a power source. In this paper, we suppose the most basic deployment scenario of LoRaWAN, and all end devices are assumed to implement class-A operation.

3 SCHEDULING TRANSMISSIONS

3.1 Premise of The Method

In the proposed method, we assume that the LoRaWAN network consists of end devices that are incompatible or compatible with the proposed method, and a single GW which can receive uplink messages in multiple channels. In this paper, for the sake of conciseness, we assume that all end devices and GW use the same Spreading Factor (SF) value, and that the transmission channels are limited to two of them, which all LoRaWAN end devices must implement.

3.2 Overview of Sheduling Transmission Method

The proposed method aims at reducing the number of frame collisions with time-slotted scheduling. Time slots are defined by dividing time per a certain length. In this paper, the length of the time slot is determined according to the time of transmitting an uplink message, i.e., end devices send one uplink message at one time slot. Transmissions of compatible end devices are scheduled to avoid collision with those of other end devices in the network.

The first step, which is shown in Sec. 3.4 in detail, is to observe incompatible end devices transmitting the uplink messages and to estimate the timing at which they will send uplink messages. Incompatible device is any device that is not compatible with the proposed method. With a certain amount of time drift, incompatible end devices transmit uplink messages periodically. Thereby, it is possible to estimate the timing at which incompatible end devices send the uplink messages, according to the records when GW receives the frames. The second step, which is shown in Sec. 3.5 in detail, is to determine when and on which channel, compatible end devices transmit their uplink messages. After being informed of the schedule computed by GW, they transmit uplink messages according to the schedule. Then, the probability of frame collision reduce because the schedule enables compatible end devices to avoid frame collisions.

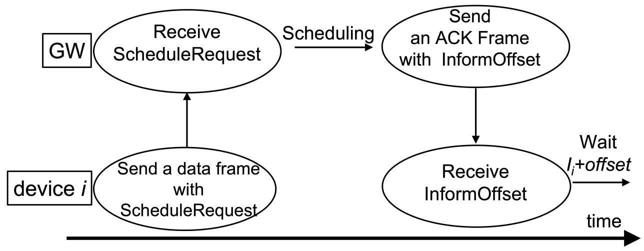


Figure 1: Scheguling flow of compatible end devices over time.

3.3 Alteration of Communication Procedures and Format

In the proposed method, the downlink messages from GW to compatible devices are extended to support scheduling functions. Therefore, GW must distinguish compatible end devices from incompatible devices. A schedule request as the MAC command from an end device indicates it is compatible, whereas no schedule request from the end device indicates that it is incompatible with the proposed method. Sending the MAC command to GW, the compatible end devices inform their transmission interval and request to schedule their transmissions.

Figure 1 shows the process in which compatible end devices inform the transmission interval to GW and obtain their schedule in return. Note that the end devices are assumed to have completed the join procedure before this trasnaction. A compatible end device i transmits a data frame attached with the MAC command, which is newly defined in the proposed method, to inform its transmission interval. When GW receives the data frame, GW schedules the transmissions of device i, and replies the transmission schedule expressed as the offset, where offset is the value to shift the next transmission timing. The MAC command, ScheduleRequest, is newly defined, by which the compatible devices inform their transmission intervals. Figure 2 shows the frame format of ScheduleRequest command. The transmission intervals of the compatible end devices are assumed 10 minutes, 15 minutes, 20 minutes, and 30 minutes, and those four kinds of transmission intervals are denoted using 2-bit. According to the transmission interval of the device, GW determines the transmission schedule.

A compatible end device which completed the join procedure to the network sends data frames with ScheduleRequest. If this is the first time transmission for the device, it sends the ScheduleRequest at an arbitrary time according to the LoRaWAN protocol. After GW receives the ScheduleRequest, the transmission schedule is determined according to the method presented in Sec. 3.5. After the transmission schedule is determined, GW supplies the schedule for the compatible end device using the MAC command called InformOffset.

Bit#	7⋯6	5⋯0
ScheduleRequest Payload	period	RFU

Figure 2: Frame format of ScheduleRequest

size(bytes)	2	1	3
InformOffsetPayload	OffsetSeconds (16bits singed integer)	OffsetFractionalSecond (8bits unsinged integer)	Freq

Figure 3: Frame format of InformOffset

The InformOffset is also a newly defined MAC command to inform the offset and the frequency channel of the compatible end devices. The offset is the value to force the end devices to change the next transmission timing. The offset, o, is determined as follows.

$$o = t_{next} - (t_r + I_i), \tag{1}$$

where t_r is the last receive time from the compatible end device, t_{next} is the estimated next transmission time and I_i is the transmission interval of the device i, respectively.

Figure 3 shows the frame format of InformOffset command. In the InformOffset command, the pair of the OffsetSeconds and the OffsetFractionalSecond fields stand for o by 3 bytes. The OffsetSeconds field represents the integer part of o rounded down to the nearest whole number. The OffsetSeconds is a positive number when o is a positive number, and a negative number when o is a negative number. The OffsetFractionalSecond field denotes the decimal point of o in 256 steps. On the other hand in the LoRaWAN protocol, the frequency channels are denoted by 3 bytes. Using this representation, the Freq field denotes the frequency channel by 3 bytes in which the end device transmits the uplink messages. When the compatible end devices receive the InformOffset, they set the channel and the time of the next transmission according to the InformOffset. The time interval until the next transmission is I_i+o and after that the intervals are o, i.e., o is effective until the device receives another InformOffset. The devices do not change the transmission channel until the device receives another InformOffset, too. The compatible end devices act like incompatible end devices if the devices do not receive the InformOffset in responce for their ScheduleRequest commands consecutively, regarding the scheduling process as wrong. Even if the compatible end devices transmit when the time is not in the schedule, GW computes their new schedules based on the latest receive time and replies the InformOffset.

3.4 Estimate Transmission Timing of The Incompatible Devices

To avoid collisions between transmissions of compatible and incompatible end devices, it is necessary to estimate the time at which the incompatible end devices transmit. Generally speaking, end devices get a large time error while they sleep. This time error causes the end devices to shift their transmission time. Under the effect of the time errror, GW performs the following procedure to estimate the time at which the incompatible end devices transmit as accurate as possible.

When GW receives the uplink messages from the incompatible end devices, GW stores the reception time on the reception record table. The number of reception time records per device is N, and so if the records exceeds N, the earliest record is deleted. If incompatible end device's reception records are more than k, the average interval of the reception time from the device is computed. This is to reduce the affection of the time error.

The average interval of reception time from the device i is defined as the estimated transmission interval and expressed as I_i. With this value, we estimate the transmission timings until T minutes away fram the current time. Given the latest reception time being t_r, GW estimates the device's transmission schedule k times as $t_r+I_i, t_r+2I_i, ..., t_r+kI_i$, where k is the quotient of T divided by I_i.

In the proposed method, the schedule array is used to manage the above transmission schedule of all the end devices. Let the length of a time slot be t_s, then the number of slots M in T is expressed as $M = \frac{T}{t_s}$. In the schedule calculation, we assume that s_0 corresponds to the latest reception time, and that the slots are given as $S = \{s_0, s_1, \ldots, s_{M-1}\}$. In the schedule array, each slot s_j holds the list of the device IDs at which the device is estimated for transmission. If there is no device to be transmitted in slot s_j, the corresponding list is $null$. After estimating the transmission slots of the incompatible end device,GW uodates the transmission schedule array, by first deleting all the past estimation and then adding the newly estimated slots to the array. In addition, GW sends the ACK message one or two seconds later after receiving the uplink message. If the data frame requires GW to send the ACK frame to the end device, the slots at which GW must transmit the downlink messages are scheduled, too.

3.5 Schedule The Transmission of The Compatible Devices

After receiving an uplink message from a compatible end device, GW calclates the transmission schedule of that device, and sends an ACK frame with the offset, which is the value indicating how much time that device must shift the next transmission. Receiving the offset with the ACK frame, that device changes the scheduled time of the next transmission accordingly.

On calculating the transmission schedule for the end device, GW selects the best offset having the lowest cost among all candidate offsets. We let the required transmission interval of compatible end device i is I_i, the time duration to calculate schedule is T, the length of the slot is t_s, the number of slots in I_i is $J_i = \frac{I_i}{t_s}$ and that in T is $M = \frac{T}{t_s}$. The offset candidate sets $O_{cand} = \{\lceil -\frac{J_i}{2} \rceil, \ldots, -1, 0, 1, \ldots, \lfloor \frac{J_i}{2} \rfloor\}$ represents the starting time of slots in a single transmission interval I_i from half before to halfvafter one cycle with respect to the first slot s_0, i.e., the current time. The schedule S_o is the set of the transmission slots at which the device transmits the uplink messages every J_I after the current time, defined for each offset candidate o. Since the current time expresses

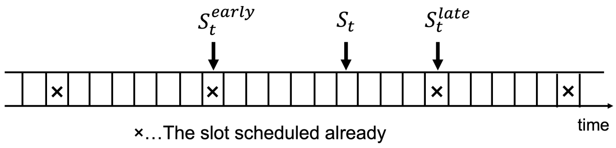


Figure 4: The example for s_t^{early} and s_t^{late}

slot s_0, it is defined as $S_o = \{s_{o+kJ_i}\}$, where $k = 1, 2, 3, \ldots$ and $0 \leq o + kJ_i \leq M$.

The cost for the offset candidate $o \in O_{cand}$ has two indicators, one evaluates the risk of collisions with other transmissions for each transmission, and the other evaluates the degree of time difference the device is forced to change its transmission time. The candidate offset with the lowest cost is adopted for the transmission schedule for the device. The cost $C(o)$ for the offset candidate, o, is defined as follows.

$$C(o) = \alpha C_{col}(o) + (1 - \alpha) C_{int}(o), \tag{2}$$

where $C_{col}(o)$ is the collision cost and $C_{int}(o)$ is the interval change cost, and α is the constant given in advance, which takes balance between those two costs.

The collision cost $C_{col}(\cdot)$ is aimed at preventing to allocate slots whose nearby slots have already been used by another device. In the LPWA network, most end devices may transmit uplink messages out of the allocated slots because of their low accuracy clock. Thereby, it is desirable that the slots scheduled for the devices are far from each other to decrease the probability of frame collisions. Terefore, $C_{col}(\cdot)$ is defined to take value if the allocated slots are closer to other occupied slots. The collision cost $C_{col}^{s_t}$ of the slot s_t is defined as follows.

$$C_{col}^{s_t} = \frac{T - kI_i}{T} \left(\frac{1}{s_t - s_t^{early} + 1} + \frac{1}{s_t^{late} - s_t + 1} \right), \tag{3}$$

where s_t^{early} is the closest occupied slot earlier than $s_t (= s_{o+kJ_i})$, s_t^{late} is that later than s_t. Figure 4 shows the example s_t^{early} and s_t^{late}. There are four occupied slots, and s_t^{early} and s_t^{late} are the closest ones earlier and later than s_t.

By summing the collision costs of the slots corresponding to an offset, the collision cost for the offset candidate o is defined as follows.

$$C_{col}(o) = \sum_{s_t \in S_o} C_{col}^{s_t}. \tag{4}$$

On the other hand, $C_{int}(\cdot)$ is the interval change cost aiming at avoiding large data interval alteration for each device in case of resheduling. It is desirable that the transmission interval of end devices does not change largely because IoT devices are required to transmissions regularly. Since larger offset means larger change in interval, the interval change cost $C_{int}(\cdot)$ should also be larger. The interval change cost is defined as follws.

$$C_{int}(o) = |\frac{o}{J_i}|. \tag{5}$$

The cost is defined as the ratio of the offset value o to the number of the slots in the transmission interval J_i. Since J_i varies for each device, this cost function takes into account the difference of transmission intervals among end devices. Finally, the offset o adopted as the transmission schedule for the device is defined as follows.

$$o^{ad} = \underset{o \in O_{cand}}{\operatorname{argmin}}\, C(o). \tag{6}$$

After determining the optimal offset o^{ad}, GW reflects the scheduling result on the schedule array. First, the device ID i are all removed from the array and write the transmission schedule on the array, where the schedule o^{ad} is $S_{o^{ad}} = \{s_{o^{ad}+kJ_i}\}, k = 1, 2, 3, ...$ Second, it is necessary to take into account the timings at which GW sends the downlink messages, which are one and two seconds later. Thereby, for each $s_t \in S_{o^{ad}}$ added into the schedule array, we add the device ID i to slots s_{t_1} and s_{t_2}, which are corresponding to Rx1 and Rx2 respectively. The schedule is computed in this manner and the offset o for the device is determined. This value of o is sent to device i in the InformOffset command attached to the ACK frame.

4 CONCLUSION

In this paper, we proposed a scheduling method for the LoRaWAN network which supports a mixture of incompatible and compatible devices with the proposed method. In the current protocol, since the ALOHA-style protocol does not control end devices to avoid collision, the devices transmit the uplink messages at arbitrary timings.

However, frame collisions are more likely to occur as the number of the end devices increased. The proposed method is aimed at improving the scalability of the LoRaWAN network by using the schedule array with time slots and by supplying the transmission schedule to compatible devices with the MAC command. First, the network server estimates the timings at which each incompatible device transmits and GW receives the uplink messages from the incompatible device in the future. Next, GW determines the transmission schedule of each compatible device when GW receives the uplink messages from the compatible device. When GW computes the transmission schedule for each compatible device, it considers the fluctuation of the transmission interval and the risk of frame collisions. The proposed method thus aims to improve the scalability of the whole network by controlling the transmission timings of devices to avoid collisions.

Acknowledgement

This work is partially supported by the Cooperative Research Project Program of the Research Institute of Electrical Communication,Tohoku University.

REFERENCES

[1] 3GPPHP, https://www.3gpp.org/news-events/3gpp-news/nb-iot-complete,Last visited February 2023.

[2] sigfoxHP, https://www.sigfox.com, Last visited February 2023.

[3] LoRaAllianceHP, https://lora-alliance.org, Last visited February 2023.

[4] Jetmir Haxhibeqiri, Eli De Poorter, Ingrid Moerman, Jeroen Hoebeke: A survey of LoRaWAN for IoT: From Technology to Application, Sensors 2018, 2018.

Comparison on function and performance between MQTT and DDS for IoT DEP

Masaki Mitsuuchi [*], Kazuhiro Kosaka [*], Tetsuya Yokotani[**], and Koichi Ishibashi[**]
[*]Graduate School of Engineering, Kanazawa Institute of Technology, Japan
{c6301158, c6201098}@st.kanazawa-it.ac.jp
[**]College of Engineering, Kanazawa Institute of Technology, Japan
{yokotani, k_ishibashi}@neptune.kanazawa-it.ac.jp

Abstract - Internet of Things (IoT) services are currently being introduced in various fields. This has led to a marked increase in the number of connected IoT devices. To accommodate them efficiently, a platform that lightens processing is considered necessary. One such platform is the IoT Data Exchange Platform (IoT DEP) standardized by ISO/IEC JTC1/SC41, which is a platform architecture that actively utilizes Pub/Sub type communication. This paper compares Message Queuing Telemetry Transport (MQTT) and Data Distribution Service (DDS), which are representative of Pub/Sub type communication, for the realization of IoT DEP. Comparison items are shown qualitatively in terms of protocol positioning (scope of specification), popularity, components, performance, quality control, security, etc. The results of the performance evaluation when discards were given showed that MQTT showed superior values in terms of both traffic volume and traffic stability compared to DDS. In terms of quality control, DDS is superior to MQTT because, in addition to guaranteed delivery, it is possible to set the timing of transmission and the deadline for transmission. Similarly, for security, DDS, which is configured with common access control settings, is considered superior to MQTT. Based on these results, MQTT is applied to End Point connections, prioritizing dissemination and lightweight. For interworking points, it is appropriate to apply DDS, which does not require a broker, and to operate as a distributed broker for MQTT.

Keywords: IoT DEP, IoT Platform, message type communication, MQTT, DDS

1. Introduction

In recent years, Internet of Things (IoT) services have been introduced in various fields. This has led to a significant increase in the number of connected IoT devices, and with the increase in IoT devices comes the need for the ability to synchronize state and data among multiple IoT devices [1]. For example, in a typical IoT architecture, it is commonly considered to apply an Internet to the Network layer. A typical IoT architecture model is shown in Fig. 1. In conventional Internets, TCP/IP is used for network access, and HTTP or HTTPS is generally applied to the application layer protocol. In this case, TCP/IP, which is a connection-based communication, requires the execution of 3-way handler education, which complicates the sequence and thereby significantly increases the required bandwidth. In addition, the conventional Internet requires a communication sequence that includes data and address resolution by the Domain Name System (DNS) each time an end device sends data. In addition, the operational burden of managing communication routes from a huge number of end devices also increases. From these perspectives, it can be said that the conventional Internet is not necessarily suitable for IoT communication. In addition, if minute data sent from IoT devices is transmitted by conventional methods, the control data becomes large and inefficient. To accommodate these efficiently, a platform that is lightweight in processing is considered necessary. To support this, topic-based publication/subscription (aka pub/sub model) is considered the most widely used communication method, and several implementations are available for field applications [2]. One such implementation is the IoT Data Exchange Platform (IoT DEP), which is being standardized at ISO/IEC JTC1/SC41 under the initiative of this university. This study compares Message Queuing Telemetry Transport (MQTT) and Data Distribution Service (DDS), which are representative of Pub/Sub type communication, aiming to realize IoT DEP.

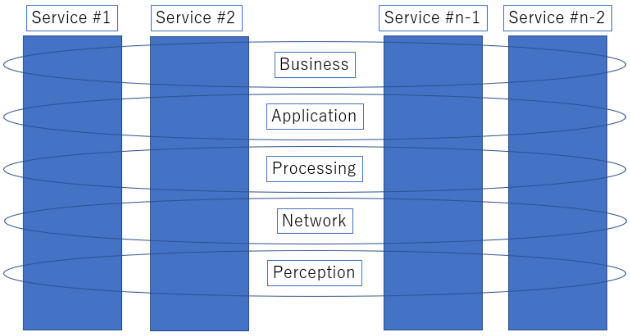


Fig.1 Typical architecture model

2. Overview of IoT communication and related areas

In this section, we will organize the situation surrounding IoT communications and provide an overview of the conditions and requirements for IoT communication applications and the relevance of the IoT reference model and platforms.

2.1. IoT Reference Model and Platforms

Various IoT reference models have been discussed in various countries, including RAMI4.0 (German: Reference Architecture Model for Industrie 4.0), IIRA (U.S.: Industrial Internet Reference Architecture), and IVI (Japan: Industrial Value chain Initiative). These overviews are publicized in [4] and elsewhere. In this context, the IoT reference model was specified by JEITA in 2019 for a wide range of purposes, including the dissemination of IoT systems, standardization of IoT architectures, and dissemination of IoT platforms.

Cite [3]. The model consists of a Value Chain axis (X-axis) indicating the purpose and business flow of the entire IoT service, an Interoperability axis (Y-axis) indicating the connection of information. The model consists of the Interoperability axis (Y-axis), which indicates integration, and the Zoon axis (Z-axis), which indicates the spread of services. Among these, the Y-axis and Z-axis, which are the key axes for a bird's-eye view of the IoT platform, are used to classify the IoT platform as shown in Fig.2.

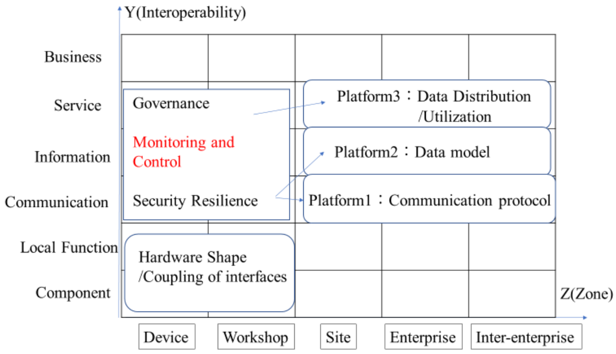


Fig.2 Positioning of IoT Platforms

Fig.2 shows the part of [3] that is necessary to discuss IoT platforms. As shown in [3], IoT platforms exist for Site (within a company), Enterprise (between sites), and Inter-enterprise (between organizations), and can be divided into Communication (network: communication protocol), Information (information processing: data model), and Service (service: data distribution/utilization). Information processing: data model), and Service (provisioning service: data distribution/utilization). In this report, Platform 1, Platform 2, and Platform 3 will be referred to as Platform 1, Platform 2, and Platform 3, respectively, with Platform 1 being the main focus.

2.2. Requirements for IoT communication

The current Internet mechanism incurs overhead in data transfer. When the frequency of data transfer is high, it causes significant problems for the communication network. Lightweight communication methods need to be introduced to solve this problem. Many IoT applications require public networks, and these applications need to coexist with traditional applications such as e-mail, video distribution, and IP telephony. However, IoT applications can benefit from lightweight communication architectures and protocols. Therefore, approaches and solutions for traditional and IoT applications should be specified. Many IoT applications require the efficient transfer of a huge number of small packets with small data blocks over the network. Furthermore, the required communication quality also depends on the type of application: application layer protocols such as HTTP require a TCP 3-way handshake and IP-based routing. Therefore, these protocols are not suitable for the widespread use of IoT.

There are requirements for the protocols to be used for data communication. One is lightweight protocol overhead: IoT applications require high-speed processing of tiny data generated by a huge number of devices. This means that a lightweight protocol must be used to transfer those data. In addition, IoT applications may require real-time rather than reliable transfers. As with lightweight protocol headers, a simplified communication sequence is desirable. Next is the QoS control function for each IoT application. Since each IoT application may require different QoS, it is necessary to provide a flexible QoS control function. When multiple applications are multiplexed, the current Internet provides these functions to each IP flow. Finally, reliable transmission independent of TCP connections. The current Internet can reliably transmit data even when some of the applications multiplexed onto the connection require real-time transmission rather than reliable transmission. One of the factors for this is the establishment of TCP connections. However, when TCP connections are established, large delays occur when real-time transmission is required. By making it selectable for each application, highly reliable transmission independent of TCP connections can be realized.

3. Overview of IoT DEP

This section provides an overview of the IoT DEP and describes the relationship between this standardization initiative led by the University and the verification of principle that is underway in parallel. We also identify issues for IoT DEP implementation and select topic-based Pub/Sub type communication to be addressed in this report.

3.1. Standardization on IoT-DEP

The following is an overview of the IoT data exchange platform (IoT DEP) that the author and others are promoting as an example of Platform 1. For more information on IoT DEP, please refer to [5]-[7], etc.

The IoT, in which all things are connected to the Internet, is a technology that is currently attracting a lot of attention, but as IoT devices become more widespread, the current Internet communication standards will become inconvenient. However, as IoT devices become more widespread, the current communication standards for the Internet will become inconvenient. In addition, an analysis of IoT use cases has been conducted, and the results indicate the need for a platform that is lightweight in its communication processing and abstracts the physical network. Based on the results of this study, the IoT DEP is the specification of a new platform that can be commonly applied to applications.

IoT DEP is a platform architecture that actively utilizes Pub/Sub communication, an IoT-specific communication method that enables more efficient data transfer than conventional communication methods using IP addresses and DNS servers. To achieve lightweight communication, Information Centric Network (ICN) technology, which is not bound to IP addresses, is applied to end devices and servers, while dedicated paths connect multiple Nodal Points in the IoT DEP network. The IoT DEP network consists of dedicated paths connecting multiple Nodal Points. IoT DEP functions are implemented in Nodal Points. This enables the IoT DEP to operate on the existing Internet and to be compatible with existing protocols such as Hypertext

Transfer Protocol (HTTP) and File Transfer Protocol (FTP), and to realize the transfer of a huge amount of data.

To briefly summarize the advantages of IoT DEP, access to a DNS server is not required, and because ICN technology is used, less header information is required, resulting in smaller communication data volume. As a result, low latency can be achieved. It is also possible to operate on the existing Internet.

IoT DEP is being standardized by ISO/IEC JTC/SC41 as ISO/IEC30161 series. Table.1 summarizes the relationship between the promotion of this standardization and the verification of principle that is being conducted in parallel with it in Fig.3.

Table.1 Scope of ISO/IEC30161 series

Part	Contents
Part1	・Requirement ・Architecture ・Functional block definition
Part2	・Functional definition of Nodal Point ・Coordinated control between Nodal Points

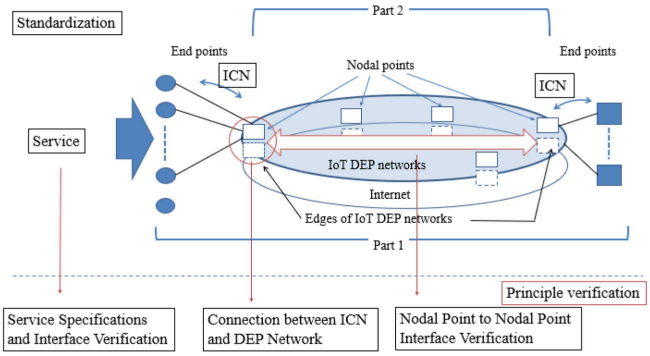


Fig.3 International standardization and proof of principle of IoT DEP

In Table.1, these standards provide requirements and architecture. Detailed protocols are positioned as implementation issues. However, we are currently verifying the principles within the framework specified in these standards.

The following three points are the key points of the proof-of-principle. In the verification of principle, priority is given to the application of MQTT to the ICN portion of Fig.3.

(i) **Service specification and interface verification**: Information from various sensors, etc., is analyzed and transferred to the payload of MQTT's Publish message. The format of information from sensors, collection period, etc. are parameters.

(ii) **Connections between ICN and DEP networks**: Connections between MQTT and the edge of the DEP network (implementation of gateway functionality).

(iii) **Interface verification between Nodal Points**: The connection between Nodal Points is between the IoT DEP middleware, which applies a simpler scheme than IP-based routing in the Internet. The schemes are summarized in [8] and can be categorized as shown in Table.2.

Table.2 Classification of Nodal Point-to-Nodal Point Control

		Dynamic	Static
(iv)		Approach1	Approach4
(v)	Direct	Approach2	Approach5
	Indirect	Approach3	Approach6

In Table.2, (iv) is a method that transfers the information request (equivalent to "Subscription" in the case of MQTT), and (v) is a method that transfers the information itself (equivalent to "Publication" in the case of MQTT). In the case of response-driven, the method in which the information itself is transferred directly into the network is called Direct, while the method in which only the ID of the information is transferred and the information itself is offloaded by a network other than the IoT DEP (e.g., existing Internet) to reduce the amount of traffic is called Indirect. In Table 2, Approach 4 and 5 are shown in [9], Approach 6 in [10], and Approach 2 in [11].

3.2. Future challenges for IoT DEP

As described above, the IoT DEP so far has applied MQTT, a typical Pub/Sub type communication between ICNs, to achieve higher speed and lower latency compared to the existing Internets. However, there are still some issues to be solved in the coordination and control between Nodal points. For example, the connection between Nodal Points of the IoT DEP in [9] is realized by connecting distributed brokers of MQTT with a virtual ring. However, this method increases the delay time on the virtual ring, and the problem between Nodal Points needs to be solved to implement IoT DEP. Therefore, this report presents a proposal for practical application of IoT DEP by qualitatively comparing MQTT, which has already been applied as a connection method between Nodal Points in IoT DEP, and DDS, a communication protocol that does not require a Broker.

4. Summary of MQTT and DDS

This section provides an overview of Pub/Sub type communication, and outlines MQTT and DDS.

4.1 Publish/Subscribe type communication

Publish/Subscribe type communication refers to communication for sending and receiving messages with the sender as the Publisher and the receiver as the Subscriber. Conventionally, to send and receive messages, it is necessary to prepare a Broker to mediate messages between the Publisher and Subscriber, and MQTT uses this method. Fig.4 and Fig.5 show the configuration diagrams of MQTT and DDS.

Fig.6 shows a configuration diagram of the communication method using HTTP. In the conventional communication method, it is necessary to convert a URL to IP, go through a DNS server, and then connect to a data server. In MQTT, however, a Broker is placed between the Publisher and the Subscriber, and connection control between the Publisher and the Subscriber can be simplified through the Broker. DDS does not require a Broker, but instead allows data servers to communicate directly with each other. This eliminates the need to access a DNS server, which is necessary with conventional communication methods, thereby preventing delays and congestion. This means that Pub/Sub communication is lighter in weight than conventional communication methods.

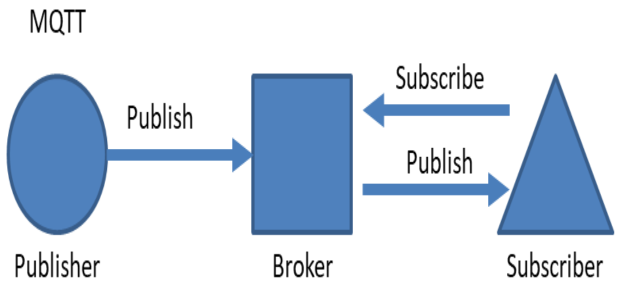


Fig.4 MQTT communication

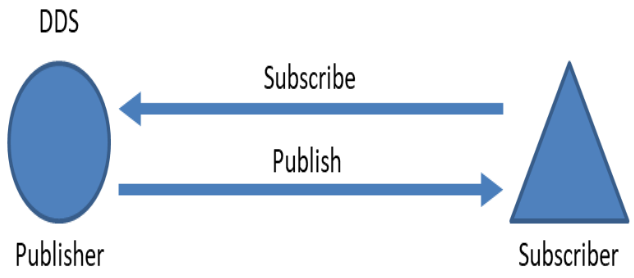


Fig.5 DDS communication

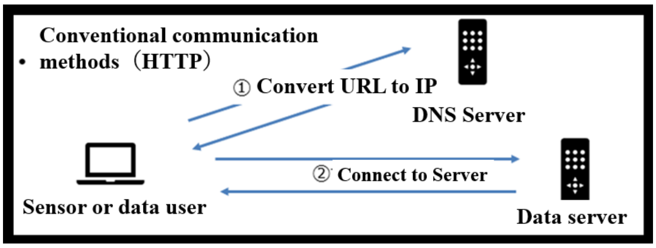


Fig.6 Configuration diagram for HTTP communication

4.2. MQTT

The Message Queuing Telemetry Transport (MQTT) protocol [12], developed by IBM and now an OASIS standard, is a lightweight message-oriented protocol [13] and a topic-based Pub/Sub type communication. MQTT can operate over TCP/IP and requires a Broker to exchange data. In practice, the Broker is a server to which clients can publish/subscribe topics, and message traffic goes through it. Clients can also authenticate to the Broker by accessing with a user name and password. SSL/TLS [14] encrypts the security of messages [15].

MQTT communicates by specifying the unique name of the information to be acquired. This unique name is a hierarchical name address called a Topic, and MQTT communicates by specifying a unique name called a Topic. Topics have a hierarchical structure, which can be created by separating each Topic with "/" when describing it. For example, KIT/GSE/LAB can be used to create a hierarchical structure as shown in the figure. In this case, use the "#" and "+" symbols. The following is an example of a topic to be specified.

Exact match：KIT/GSE/LAB
Left-hand match：KIT/GSE/＃
Partial match：KIT/＋/Floor

The system is designed to receive only those items specified in this way.

MQTT implements three guaranteed delivery functions. QoS0 (non-guaranteed delivery), QoS1 (guaranteed delivery), and QoS2 (non-overlapping guaranteed delivery) can be set, respectively. Fig.7 shows the communication sequence for each QoS function.

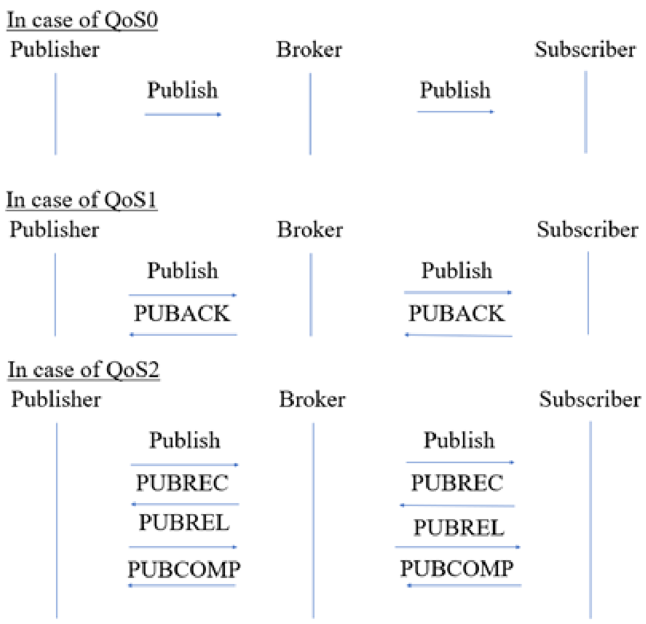


Fig.7 Communication sequence of MQTT QoS function

QoS0 has no delivery guarantee and only sends a Publish message once, which is suitable for frequent transmission of sensor data over time. QoS1 provides more reliable communication than QoS0 by sending a delivery confirmation message called PUBACK to the Publish message and guaranteeing delivery once. QoS2 provides more reliable communication than QoS1 and is the most reliable in MQTT. Compared to QoS0 and QoS1, QoS2 is used when the number of communications is higher than that of QoS0 and QoS1, but when data is to be sent reliably. MQTT is used when you want to send data reliably. In this way, MQTT provides delivery assurance with three QoS functions.

4.3. DDS

Data Distribution Service (DDS) is data-centric publication and subscription middleware for highly dynamic distributed systems [19], standardized by OMG (Object Management Group) [16]. Data is published to a DDS domain, and subscribers can subscribe to data from that domain without knowing the source or structure of the information. DDS offers a wide range of QoS parameters such as durability, lifetime, presentation, reliability, and delivery time [17]. According to the OMG website, DDS is one of many protocols used in industrial sectors such as railway networks, air traffic control, smart energy, medical services, military and aerospace, and industrial automation [18].

Like MQTT, DDS is a topic-based Pub/Sub type communication, and has in common that the quality control function called QoS can be used to set the guaranteed delivery level and that it is implemented by middleware [21]. One difference is that MQTT operates over TCP/IP, while DDS can operate over UDP and TCP. Another difference in design is that MQTT requires a Broker, whereas DDS does not require a Broker and allows direct communication between Publishers and Subscribers. DDS can also provide real-time, many-to-many managed connections [20].

5. Comparison of MQTT and DDS for connection between Nodal Points

In this section, we will qualitatively compare MQTT and DDS and describe the comparison results with references and other information. Based on the comparison results, advantages and disadvantages in MQTT and DDS will be identified.

5.1 Comparative verification

Comparison of MQTT and DDS. In Table.3, we show the comparison results in MQTT and DDS based on the literature in [24] and the results of the performance evaluation conducted in this study. The number of + indicates a superior characteristic.

Table.3 Comparison MQTT and DDS

	MQTT	DDS
Layer	Application Layer	Application Layer, Data Model
Popularity	+++	+
Architecture	Pub/Sub /Broker	Pub/Sub
Performance	++	+
QoS	++	++
Security	+	++

In Table.3, with respect to penetration, MQTT is one of the most used messaging protocols by developers, whereas DDS is not supported on existing IoT Platform platforms [21]. A clear difference between DDS and MQTT in terms of components is that MQTT has a Broker and DDS does not; MQTT ensures security by requiring a Broker, whereas DDS is configured with common access control settings, The absence of a Broker does not undermine security [22]. In terms of quality control, DDS is superior to MQTT in that all data requires a QoS contract, and in addition to guaranteed delivery, the timing of transmission and the deadline for transmission can also be configured. One of the factors that make DDS superior in terms of security is that it is equipped with access control plug-ins, such as Mosquitto, which implements access control by user/password or RSA authentication for the Subscriber and Publisher of a specified topic. Although some MQTT protocols, such as Mosquitto, implement access control by user/password or RSA authentication for the Subscriber and Publisher of a given topic, the protocol itself does not specify access control, and is considered less secure than DDS [23].

6. Performance evaluation

In this section, performance evaluation is conducted under the following conditions in order to evaluate the "performance" item shown in Table.4 on actual equipment. The evaluation results are also discussed and explained.

6.1. Configuration

Next, performance evaluation is performed under the following conditions: MQTT is based on TCP, DDS is based on UDP and TCP, and QoS is set to QoS0 (non-guaranteed delivery) for MQTT, Besteffort (non-guaranteed delivery) for TCP-based DDS, and Reliable for UDP-based DDS. The Reliable setting refers to the state in which the ACK return function that is originally performed over TCP is added. In addition, experiments will be conducted using two Raspberry Pi, with the same Raspberry Pi used as the Broker and Subscriber for MQTT. In order to clearly show the difference in performance between the two, the Publisher side will be given 1% and 10% discards for verification, and the data will be observed using Wireshark. The protocol stacks of MQTT and DDS are shown in Fig. 8, and the configuration diagram for performance evaluation is shown in Fig. 9.

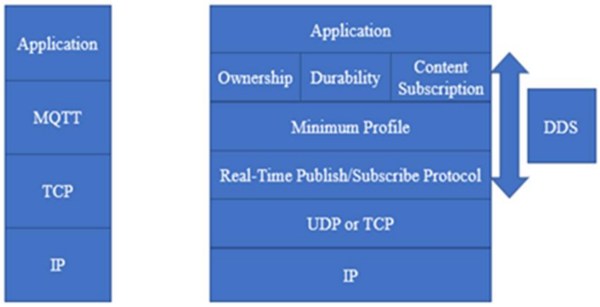


Fig.8 Protocol Stack

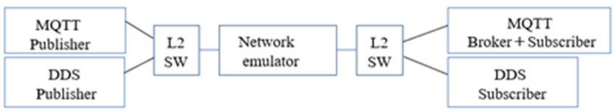
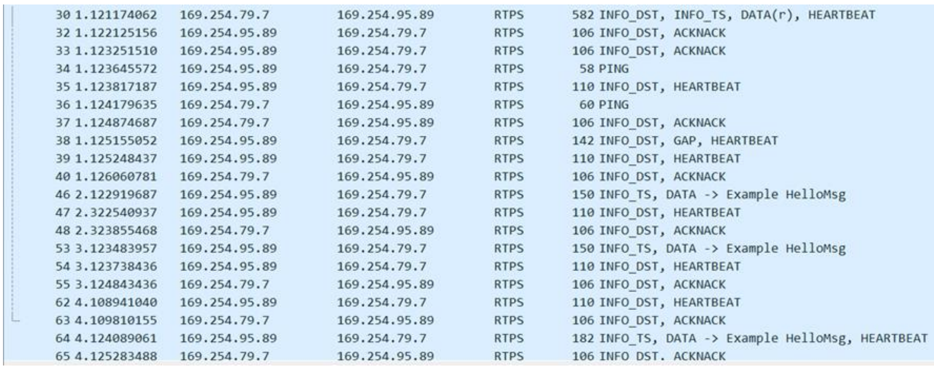


Fig.9 Configuration diagram

6.2. Verification Results

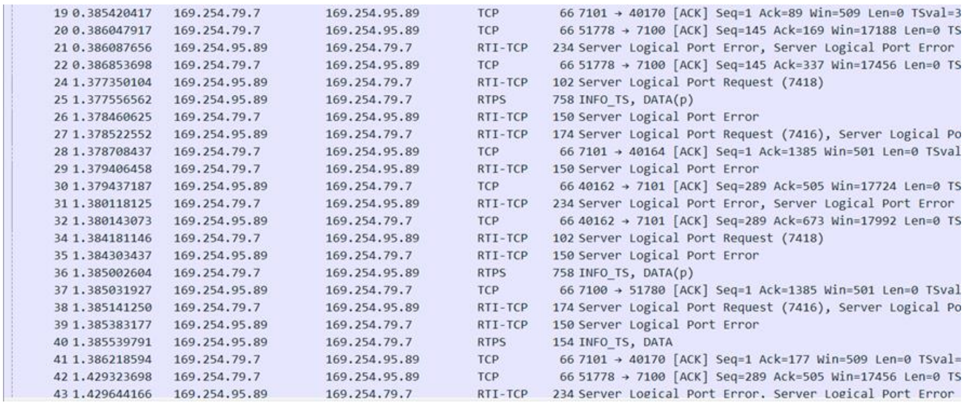
6.2.1. Measurement results in Wireshark

As a representative example of Wireshark results measuring message transmission, Figs. 10 through 12 show the results of MQTT and DDS measurements at 1% waste. In these results, RTPS refers to DDS. These data were taken from Wireshark, which was used for the observation. The traffic for MQTT is the smallest for message transmission, while the traffic for DDS is almost the same, but the traffic for ACK is higher for UDP than for TCP, resulting in a difference in the traffic for message transmission. The difference in the traffic for message transmission is due to the higher ACK traffic for UDP compared to TCP.



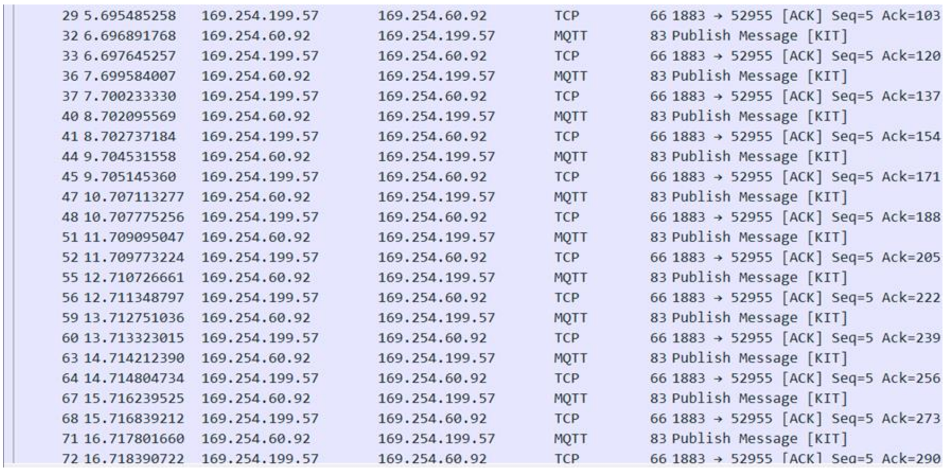
30	1.121174062	169.254.79.7	169.254.95.89	RTPS	582	INFO_DST, INFO_TS, DATA(r), HEARTBEAT
32	1.122125156	169.254.95.89	169.254.79.7	RTPS	106	INFO_DST, ACKNACK
33	1.123251510	169.254.95.89	169.254.79.7	RTPS	106	INFO_DST, ACKNACK
34	1.123645572	169.254.95.89	169.254.79.7	RTPS	58	PING
35	1.123817187	169.254.95.89	169.254.79.7	RTPS	110	INFO_DST, HEARTBEAT
36	1.124179635	169.254.79.7	169.254.95.89	RTPS	60	PING
37	1.124874687	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK
38	1.125155052	169.254.95.89	169.254.79.7	RTPS	142	INFO_DST, GAP, HEARTBEAT
39	1.125248437	169.254.95.89	169.254.79.7	RTPS	110	INFO_DST, HEARTBEAT
40	1.126060781	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK
46	2.122919687	169.254.95.89	169.254.79.7	RTPS	150	INFO_TS, DATA -> Example HelloMsg
47	2.322540937	169.254.95.89	169.254.79.7	RTPS	110	INFO_DST, HEARTBEAT
48	2.323855468	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK
53	3.123483957	169.254.95.89	169.254.79.7	RTPS	150	INFO_TS, DATA -> Example HelloMsg
54	3.123738436	169.254.95.89	169.254.79.7	RTPS	110	INFO_DST, HEARTBEAT
55	3.124843436	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK
62	4.108941040	169.254.95.89	169.254.79.7	RTPS	110	INFO_DST, HEARTBEAT
63	4.109810155	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK
64	4.124089061	169.254.95.89	169.254.79.7	RTPS	182	INFO_TS, DATA -> Example HelloMsg, HEARTBEAT
65	4.125283488	169.254.79.7	169.254.95.89	RTPS	106	INFO_DST, ACKNACK

Fig.10 Packet capture with 1% DDS UDP loss



19	0.385420417	169.254.79.7	169.254.95.89	TCP	66	7101 → 40170 [ACK] Seq=1 Ack=89 Win=509 Len=0 TSval=3
20	0.386047917	169.254.79.7	169.254.95.89	TCP	66	51778 → 7100 [ACK] Seq=145 Ack=169 Win=17188 Len=0 TS
21	0.386087656	169.254.95.89	169.254.79.7	RTI-TCP	234	Server Logical Port Error, Server Logical Port Error
22	0.386853698	169.254.79.7	169.254.95.89	TCP	66	51778 → 7100 [ACK] Seq=145 Ack=337 Win=17456 Len=0 TS
24	1.377350104	169.254.95.89	169.254.79.7	RTI-TCP	102	Server Logical Port Request (7418)
25	1.377556562	169.254.95.89	169.254.79.7	RTPS	758	INFO_TS, DATA(p)
26	1.378460625	169.254.79.7	169.254.95.89	RTI-TCP	150	Server Logical Port Error
27	1.378522552	169.254.95.89	169.254.79.7	RTI-TCP	174	Server Logical Port Request (7416), Server Logical Po
28	1.378708437	169.254.79.7	169.254.95.89	TCP	66	7101 → 40164 [ACK] Seq=1 Ack=1385 Win=501 Len=0 TSval
29	1.379406458	169.254.79.7	169.254.95.89	RTI-TCP	150	Server Logical Port Error
30	1.379437187	169.254.95.89	169.254.79.7	TCP	66	40162 → 7101 [ACK] Seq=289 Ack=505 Win=17724 Len=0 TS
31	1.380118125	169.254.79.7	169.254.95.89	RTI-TCP	234	Server Logical Port Error, Server Logical Port Error
32	1.380143073	169.254.95.89	169.254.79.7	TCP	66	40162 → 7101 [ACK] Seq=289 Ack=673 Win=17992 Len=0 TS
34	1.384181146	169.254.79.7	169.254.95.89	RTI-TCP	102	Server Logical Port Request (7418)
35	1.384303437	169.254.95.89	169.254.79.7	RTI-TCP	150	Server Logical Port Error
36	1.385002604	169.254.79.7	169.254.95.89	RTPS	758	INFO_TS, DATA(p)
37	1.385031927	169.254.95.89	169.254.79.7	TCP	66	7100 → 51780 [ACK] Seq=1 Ack=1385 Win=501 Len=0 TSval
38	1.385141250	169.254.79.7	169.254.95.89	RTI-TCP	174	Server Logical Port Request (7416), Server Logical Po
39	1.385383177	169.254.95.89	169.254.79.7	RTI-TCP	150	Server Logical Port Error
40	1.385539791	169.254.95.89	169.254.79.7	RTPS	154	INFO_TS, DATA
41	1.386218594	169.254.79.7	169.254.95.89	TCP	66	7101 → 40170 [ACK] Seq=1 Ack=177 Win=509 Len=0 TSval=
42	1.429323698	169.254.79.7	169.254.95.89	TCP	66	51778 → 7100 [ACK] Seq=289 Ack=505 Win=17456 Len=0 TS
43	1.429644166	169.254.95.89	169.254.79.7	RTI-TCP	234	Server Logical Port Error, Server Logical Port Error

Fig.11 Packet capture with 1% DDS TCP loss



29	5.695485258	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=103
32	6.696891768	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
33	6.697645257	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=120
36	7.699584007	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
37	7.700233330	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=137
40	8.702095569	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
41	8.702737184	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=154
44	9.704531558	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
45	9.705145360	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=171
47	10.707113277	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
48	10.707775256	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=188
51	11.709095047	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
52	11.709773224	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=205
55	12.710726661	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
56	12.711348797	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=222
59	13.712751036	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
60	13.713323015	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=239
63	14.714212390	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
64	14.714804734	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=256
67	15.716239525	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
68	15.716839212	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=273
71	16.717801660	169.254.60.92	169.254.199.57	MQTT	83	Publish Message [KIT]
72	16.718390722	169.254.199.57	169.254.60.92	TCP	66	1883 → 52955 [ACK] Seq=5 Ack=290

Fig.12 Packet capture with 1% MQTT loss

6.2.2. Experimental results at 1% disposal

The performance results at 1% waste are shown in Fig.13. MQTT showed almost the same amount of traffic for each message sent, and the shape of the graph was stable. The graph shape was also stable. Finally, the DDS traffic for TCP was smaller than the DDS Reliable traffic for sending messages, but the graph shape was disturbed by the regularity of the traffic for sharing information with each other.

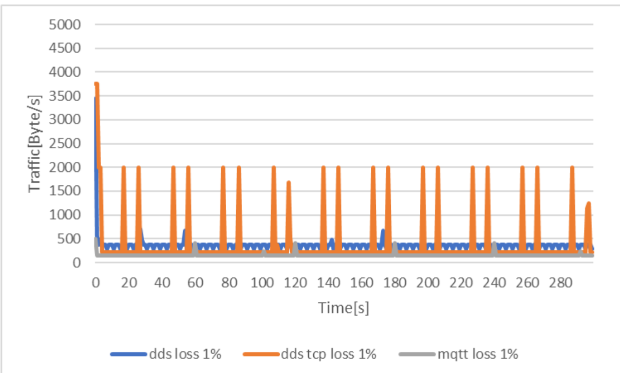


Fig.13 Result in loss 1%

6.2.3. Experimental results at 10% waste

Next, the performance results at 10% disposal are shown in Fig.14. The graphs for both DDS and MQTT are different in terms of their characteristics at disposal, since they both show similar graph shape disorder.

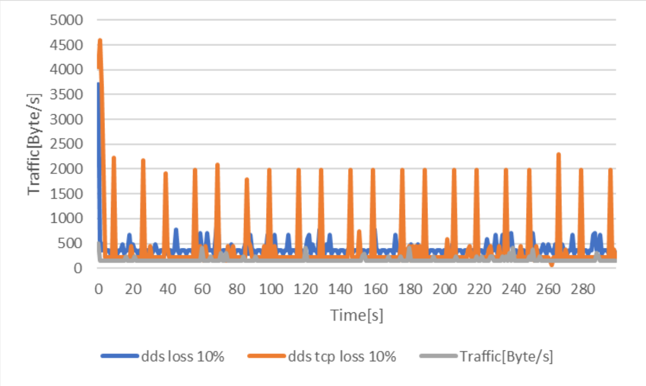


Fig.14 Result in loss 10%

6.2.4. Summary of experimental results

The results of the average traffic per second are shown in Table.4. It can be seen that in all conditions, the traffic at 10% discard is higher than that at 1% discard, and for DDS, TCP has higher traffic than UDP. The comparison of MQTT and DDS shows that MQTT has lower traffic than DDS UDP/TCP. The following results may be due to the fact that TCP has a higher load than UDP and UDP has a smaller header size than TCP.

Table.4 Average traffic per second

Traffic (Byte)	DDS UDP	DDS TCP	MQTT
1%	361.667	374.167	154.293
10%	395.007	401.447	162.593

7. Proposed configuration for IoT DEP

In previous IoT DEPs, MQTT has been applied to connect EPs, and Nodal Points have been realized by connecting distributed MQTT brokers in a virtual ring. However, there is concern that this method increases the delay time on the virtual ring. Therefore, based on these results, a proposal for application to IoT DEP is shown in Fig.15. For endpoint connections, MQTT is applied, prioritizing dissemination and lightweighting. It is also appropriate to apply DDS, which does not require a broker, between Nodal Points to act as a distributed broker for MQTT. Optimization of communication paths can be achieved.

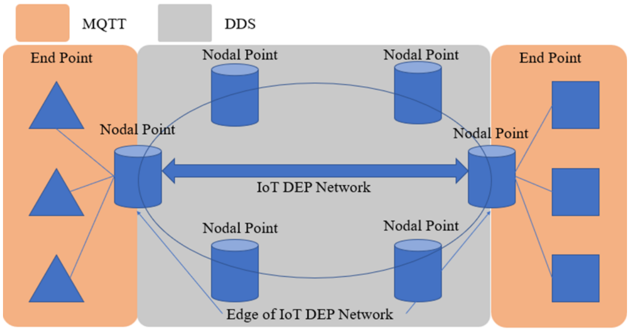


Fig.15 Proposal for IoT DEP

8. Summary

In this paper, we took up MQTT and DDS, which are representative examples of pub/sub communication, and conducted qualitative comparisons and performance evaluations using actual equipment. We also proposed an application method using both MQTT and DDS for the standardization of IoT DEP. In the future, we plan to implement and evaluate the proposed application in Fig. 13 in a limited environment using Raspberry Pi, and work toward practical application.

REFERENCES

[1] A. Shaout, and B. Crispin, Using the MQTT Protocol in Real Time for Synchronizing IoT Device State, The International Arab Journal of Information Technology, Vol. 15, No. 3A (2018)

[2] W. Y. Liang, Y. Yuan, and H. J. Lin, A Performance Study on the Throughput and Latency of Zenoh, MQTT, Kafka, and DDS, Distributed, Parallel, and Cluster Computing (cs.DC), arXiv:2303.09419 [cs.DC]

[3] https://home.jeita.or.jp/cgi-bin/page/detail.cgi?n=1171 (2022-6-19 Access)

[4] M. Weyrich, and C. Ebert, Reference Architectures for the Internet of Things, IEEE Software, Vol. 33, No. 1, pp. 112 – 116 (2016)

[5] T. Yokotani, and K. Kawai, Concepts and requirements of IoT networks using IoT Data Exchange Platform toward International standards, 2019 IEEE Conference on Standards for Communications and Networking (CSCN), pp. 1-6(2019)

[6] T. Yokotani, A proposal of IoT DEP as a Platform for Efficient IoT Service Deployment on Wide Area Networks, IEEE 2019 International Japan – Africa conference on Electronics, Communications and Computers (IEEE JAC-ECC 2019), IEICE Invited session -1 (2019)

[7] T. Yokotani, and K. Kawai, Survey on standardization activities of IoT and proposal of the IoT data exchange platform, IEICE 2020 International Conference on Emerging Technologies for Communications (ICETC 2020), IB3-3 (2020)

[8] T. Yokotani, S. Yamamoto, S. Ohno, K. Sasabayashi, and Koichi Ishibashi, Survey and comparison of Interworking point routing mechanisms for IoT services in wide area ICNs, IEICE 2020 International

Conference on Emerging Technologies for Communications (ICETC 2020), D2-4 (2020)
[9] S. Ohno, K. Terada, T. Yokotani, and K. Ishibashi, Distributed MQTT broker architecture using ring topology and its prototype, IEICE Communication Express, Volume 10 Issue 8, pp.582-586 (2021)
[10] S. Ohno, K. Terada, T. Yokotani, and K. Ishibashi, The control mechanism of distributed MQTT brokers for large volume data transfer and its prototype system, IEICE Communication Express, Volume 11 Issue 3, pp.160-164 (2022)
[11] Y. Noda, S. Ohno, K. Ishibashi, and T. Yokotani, A new routing mechanism based on layer 2 control in MQTT networks with multiple brokers, IEICE Communication Express, Volume 11 Issue 6, pp. 307-312 (2022)
[12] Sebastian Raff. The MQTT Community. Available online: https://github.com/mqtt/mqtt.github.io/wiki (accessed on 10 June 2019)
[13] D. Silva, L. I. Carvalho, J. Soares, and R. C. Sofia, A Performance Analysis of Internet of Things Networking Protocols: Evaluating MQTT, CoAP, OPC UA, Appl. Sci. (2021)
[14] B. H. Çorak, F. Y. Okay, M. Güzel, Ş. Murt and S. Ozdemir, Comparative Analysis of IoT Communication Protocols, 2018 International Symposium on Networks, Computers and Communications (ISNCC), Rome, Italy, pp. 1-6, doi: 10.1109/ISNCC.2018.8530963. (2018)
[15] S. Lee, H. Kim, D. K. Hong and H. Ju, Correlation analysis of MQTT loss and delay according to QoS level, 2013 IEEE International Conference on Information Networking (ICOIN), pp. 714-717 (2013)
[16] https://portals.omg.org/dds/
[17] http://download.prismtech.com/docs/Vortex/html/ospl/DDSTutorial/qos.html
[18] http://portals.omg.org/dds/who-is-using-dds-2/
[19] S. Profanter, A. Tekat, K. Dorofeev, M. Rickert and A. Knoll, OPC UA versus ROS, DDS, and MQTT: Performance Evaluation of Industry 4.0 Protocols, 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, VIC, Australia, pp. 955-962, doi: 10.1109/ICIT.2019.8755050. (2019)
[20] S. Seleznev, and V. Yakovlev, Industrial Application Architecture IoT and protocols AMQP, MQTT, JMS, REST, CoAP, XMPP, DDS, International Journal of Open Information Technologies ISSN: 2307-8162 vol. 7, no.5 (2019)
[21] E. Al-Masri, K. Raj Kalyanam, J. Ratts, J. Kim, S. Singh, T. Vo, and C. Yan, Investigating Messaging Protocols for the Internet of Things (IoT), in IEEE Access, vol. 8, pp. 94880-94911 (2020)
[22] eProsima Fast DDS Documentation, https://fast-dds.docs.eprosima.com/en/latest/index.html
[23] G. Aures, and C. Lübben, DDS vs. MQTT vs. VSL for IoT, Seminar IITM SS 19, Network Architectures and Services (2019)

Study of an implementation method of point-to-multipoint communication for IoT data exchange to reduce traffic on an IoT network

Takahiro Shiohara*, Koichi Ishibashi** , and Tetsuya Yokotani**

* Information and Computer Engineering, Graduate School of Engineering, Kanazawa Institute of Technology, Japan
c6201940@st.kanazawa-it.ac.jp
** College of Engineering, Kanazawa Institute of Technology, Japan
{k_ishibashi, yokotani}@neptune.kanazawa-it.ac.jp

Abstract - Due to the growing interest in the Internet of Things (IoT) in recent years and toward the further development of the IoT, platforms for efficiently exchanging IoT data generated by IoT devices such as sensors and actuators among IoT users are being discussed in various areas. In this platform, the interworking among nodal points, which IoT devices and IoT users connect to, is essential to support wide-area and large-scale IoT systems. In exchanging and sharing IoT data, requirements such as reliability and low latency depend on the IoT service. In addition, it is desirable to share IoT data generated by IoT devices among multiple IoT devices. In address to the above requirements, this paper proposes an implementation method of point-to-multipoint communication to efficiently exchange IoT data among users according to the quality of service required by the IoT data. The proposed method is characterized by the coordination with multicast control at the application level in order to accommodate various quality of services with IoT data while utilizing Data Distribution Service (DDS), which is one of the candidate communication protocols among nodal points. By comparison with the method of using only DDS, we confirmed the proposed method reduces the traffic volumes among nodal points.

Keywords: Internet of Things, DDS, point-to-multipoint communication.

1 INTRODUCTION

In recent years, there has been growing interest in the Internet of Things (IoT) technology, in which all things are connected to a network. With the expansion and development of IoT-based services and businesses, the number of connected IoT devices is increasing year by year, and IoT devices are expected to exceed 29 billion by 2024[1]. Due to the increasing number of IoT devices such as sensors and actuators connected to IoT systems, the interoperability of the network devices, called nodal points that accommodate the IoT devices and IoT user's assets is increasingly being discussed. Therefore, there are many efforts towards developing IoT systems and proposals to efficiently exchange IoT data under wide-are and large-scale IoT networks. For example, there are proposals for monitoring urban transportation systems[2] and research on efficient data collection systems from sensors[3].

In addition, platforms for efficiently sharing data generated by IoT devices (IoT data) among users (IoT users) are also discussed in various places to cope with the growing scale of IoT systems and the increasing number of IoT devices. For example, there is research on providing power-saving IoT services by applying mobile edge computing technology to unmanned aerial vehicles equipped with IoT devices[4].

We are considering a wide-area and large-scale IoT system for efficient exchanging IoT data among users. Our targeted IoT system consists of IoT access networks and an IoT core network. The IoT access network is an access network to accommodate IoT devices and IoT users, and is connected to the IoT core network via a gateway device called a nodal point. In the IoT access network, Message Queuing Telemetry Transport (MQTT)[5] is deployed as a communication protocol for transmitting/receiving IoT data to/from IoT devices and IoT users. MQTT is a candidate communication protocol for IoT systems. The IoT core network is a network for development of a wide-are IoT system, and multiple MQTT broker, which are implemented on nodal points, are interconnected with each other. The interworking among nodal points is deployed by the Data Distribution Service (DDS)[6].

In addition, the targeted IoT system aims to provide IoT services with various service requirements in terms of reliability and low latency. If the reliable mode provided by the DDS is only utilized to achieve interworking among nodal points, overheads of control packet by the DDS increases. On the other hand, interworking based on the best-effort mode by the DDS cannot provide the desired QoS level in the exchange of IoT data with reliability and/or low latency requirements. Therefore, we propose an implementation method which is characterized by the coordination with multicast control at the application level in order to accommodate various quality of services with IoT data while utilizing the DDS. We also show the effectiveness of the proposed method.

The remainder of the paper is organized as follows: Section 2 explains the targeted IoT system in the paper. In section 3, related works related to the paper are presented. And, section 4 proposes the implementation method of point-to-multipoint communication at the application level, section 5 describes the results of evaluation. Finally, section 6 concludes the work.

2 TARGETED IOT SYSTEM

In this section, we explain the targeted IoT system and its challenges. Figure 1 shows the targeted IoT system. IoT devices such as sensors and actuators, and IoT users such as servers are connected to a nodal point via MQTT protocol. Here, the nodal point acts as the MQTT broker. Interworking among multiple MQTT brokers, i.e., nodal points, are coordinated by DDS. That is, nodal points also act as DDS nodes, and implements a software function in order to interwork between the MQTT client and the DDS function.

IoT data generated by an IoT device is published to a nodal point, which is an MQTT broker. The published IoT data is shared among multiple MQTT brokers because it is utilized by multiple IoT users. In other words, IoT data is forwarded by multicast manner from the nodal point, which the IoT device connects to, to multiple nodal points. IoT data generated by the IoT device require various QoS requirements, such as best effort, reliable, and low latency. In addition to IoT data generated by the IoT devices, the nodal point generates novel IoT data based on edge processing of IoT data published by IoT devices.

Such an IoT system has the following issues. Firstly, the amount of traffic forwarded among nodal points increases as the number of connected IoT devices increases. Secondly, due to interference of IoT data with best effort service on the coexistence of IoT data with various QoS requirements, service level for IoT data with reliable and/or low latency requirements degrade. Lastly, since IoT data published from IoT devices is expected to be subscribed in multiple IoT users, the nodal point that receives IoT data from IoT devices is required to forward IoT data with multiple nodal points, and efficient data sharing in point-to-multipoint communication among nodal points must also be considered. And, connections among the nodal points are configured via communication paths with various communication characteristics. That is, it is necessary to consider the existence of nodal points connected via a communication path with large packet loss rates or large latency. In addition, the nodal points accommodate with multiple IoT devices. So, it is necessary to consider that the nodal point receives IoT data with reliable and/or low latency requirements and IoT data with best effort service from multiple IoT devices at the same time.

When DDS is applied in communication among the nodal points, communication services among nodal points depends on DDS functionality. That is, reliability support on DDS is ensured by acknowledgement control from the destination nodes. And, in point-to-multipoint communication on DDS, transmitting of a packet is completed after confirmation of acknowledgements from all destination nodes. Therefore, if there is a nodal point connected via a communication path with large packet loss rate or large latency, point-to-multipoint communication is influenced from the path with large packet loss or large latency. It is necessary to consider a method for ensuring communication methods that does not depend on the DDS functionality and bad communication paths. Based on the above, we propose a multicast control method that considers coordination with DDS by upper-level applications to achieve efficient data communication between DDS functions and IoT data.

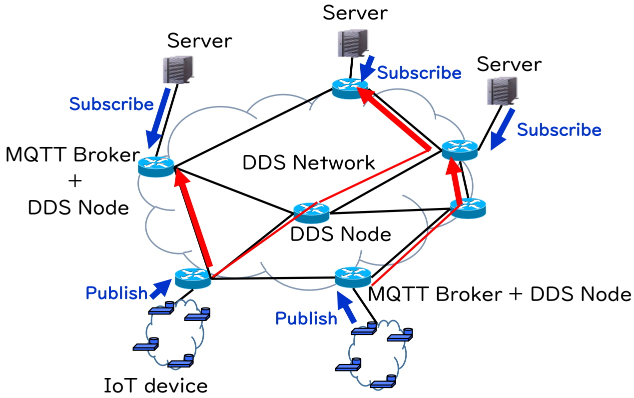


Figure 1: Example of IoT systems.

3 RELATED WORKS

This section describes the research works related to efforts using MQTT and DDS, and multicast control at the application level.

In [7], several MQTT protocols including the open source “Mosquitto” are evaluated in terms of resource consumption and latency, and the results are shown. In [8], the authors proposed a communication scheme for IoT devices and built a platform for evaluating the system. Evaluations were conducted on a data-by-data basis, showing that the system is efficient in distributing data over a network. In [9], performance evaluation and comparison of communication protocols for IoT such as MQTT and DDS are conducted, and its evaluation shows that MQTT significantly reduces round trip time (RTT) for servers and DDS has high performance in protocol implementation. In [10], data transfer using DDS is implemented, showing that it provides low latency and high throughput, and is an effective communication protocol for communication systems such as those in smart cities.

Next, multicast control at the application level is proposed to improve resource consumption, such as throughput, in the target network to address the issues introduced by conventional communication protocols. For example, latency recovery and fault recovery characteristics have been achieved by drastically reducing control traffic in the bandwidth of stagnation[11]. In [12], a protocol for multicasting at the low-bandwidth application layer is proposed to reduce overhead. Simulation evaluations of the protocols implemented in applications show that the proposed protocols can significantly reduce control traffic. In [12], an algorithm is proposed and evaluated to improve end-to-end throughput at the application level. The evaluation results show that the proposed protocol can significantly improve the throughput.

4 PROPOSAL

In this section, we describe a proposed method among DDS nodes to efficiently exchange IoT data in the targeted

IoT system. Firstly, an explanation of DDS is given. Secondly, proposed method is explained.

4.1 DDS

The protocol stack on DDS is shown in Figure 2. DDS is implemented on UDP/IP, and supports various Quality of Services (QoS) for applications that exchange IoT data via DDS. A software compliant with the DDS provides application interfaces for transmission and receive of data, such as a “DataWriter” and a “DataReader”. The DataWriter is the application interface and provides a function to transmit data to other nodes. The DataReader is the application interface and provides a function to receive data from other nodes. In addition, the DataWriter transmits HEARTBEAT packets to support reliable data exchange. That is, The DataWriter transmits a HEARTBEAT to the destination as a packet to confirm the reachability of data with reliable requirements. The DataReader that receives the HEARTBEAT responds with the sequence numbers of the packets it received before receiving the HEARTBEAT. The sender confirms that the packet has reached the destination by receiving the response to the HEARTBEAT (Figure 3).

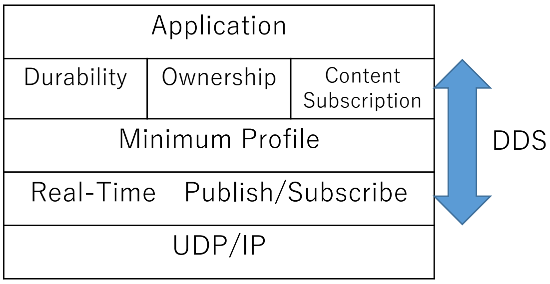


Figure 2: Protocol stack of DDS.

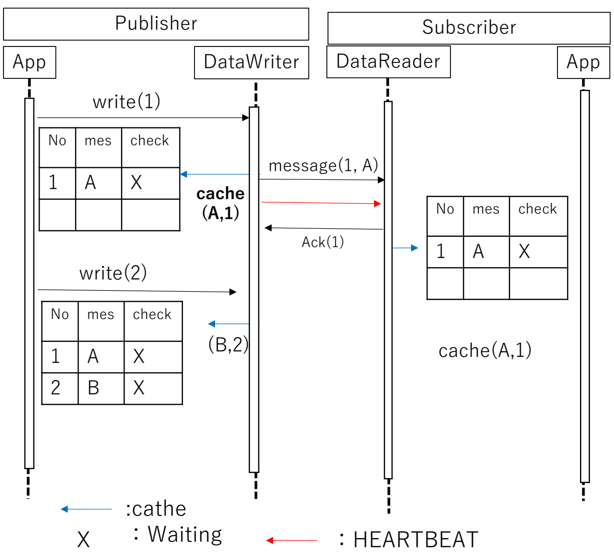


Figure 3: Message sequence of DDS Reliability mode.

4.2 Proposed Method

Figure 4 shows the flowchart on the nodal point of the proposed method when forwarding IoT data from the IoT device to other nodal points. IoT data from the IoT device are received on an application at application level and forwarded to other nodal points via DDS protocol. The application also generates novel IoT data according to an edge process of received IoT data. In the proposed method, IoT data are forwarded under DDS best-effort service. Therefore, IoT data with reliable and/or low latency requirements are provided a required QoS at application level. That is, the application for forwarding IoT data, called forwarding application, support to control for reliable point-to-multipoint communication. The forwarding application firstly decides whether IoT data require to reliable and/or low latency services according to requirements with IoT data. If IoT data requires reliable and/or low latency service, IoT data is transmitted by way of the control process. The control process in the forwarding application achieves reliable data transfer instead of DDS reliability service.

That is, when it transmits/forwards IoT data with reliable and/or low latency requirements, firstly transmitted packets are copied to a buffer for retransmission control. And an entry, that manages reliability to destinated nodes, is created. Lastly, the packets are transmitted by way of the DDS function for data transmission with best effort manner. A node that receives packets responds with the acknowledgement packets if received packets require acknowledgement for reliable communication. The transmitting node updates the entry for managing retransmissions upon receipt of an acknowledgment packet.

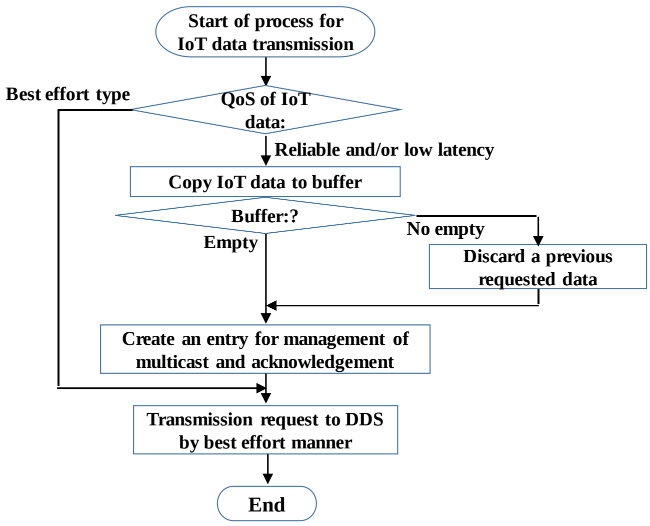


Figure 4: Flowchart on nodal point for forwarding IoT data.

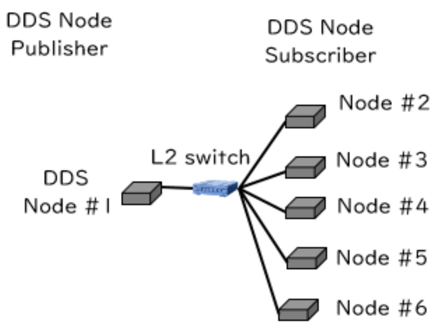


Figure 5: Evaluation systems.

5 EVALUATION

To verify the effectiveness of the proposed method, we constructed an experimental network with multiple DDS nodes which is located on nodal points (Figure 5) and evaluated the traffic volume between DDS nodes, that is traffic volume on the IoT core network. In the experimental network, we evaluated the traffic volume when transferring IoT data via the proposed method and via DDS reliability service, called “using only DDS”. We also compared the traffic volume according to the generation probability of IoT data with reliable and/or low latency requirements. In evaluation, Node#1 transmits IoT data to other DDS nodes at 1 packet/second in 1000 seconds. And the data transmitted and received on Node#1 are counted. Ratio of IoT data with reliable and/or low latency requirements varied from 10% to 0.5 %. In addition, the flow of messages in point-to-multipoint communication with QoS that provides reliability service in DDS was verified.

Figure 6 shows the traffic volume when IoT data is forwarded by using the proposed method and by DDS reliability service, called “using only DDS”. In Figure 6, the number of receiving DDS nodes varies from 1 to 5, and ratio of number of IoT data with reliable and/or low latency requirements is 10%. When “Using only DDS” is applied, the overall traffic volume increased because the traffic of acknowledgment control for confirmation to guarantee the reliability is required for forwarding all IoT data. In the proposed method, the traffic volume is suppressed because acknowledgement is performed at the application level only for IoT data with reliable and/or low latency requirements.

Figure 7 shows the traffic volume depending on the ratio of IoT data with reliable and/or low latency requirements. The traffic volume of the proposed method increases the increase in acknowledgement messages, as the ratio of IoT data with reliable and/or low latency requirements in the sending node decreases. Because in the proposal method, IoT with reliable and/or low latency requirements data require only acknowledgement message. On the other hand, when “using only DDS” is applied, the traffic volume at the sending node remains constant regardless of the percentage of IoT data with reliable and/or low latency requirements because reliability service in DDS always checks for transmission acknowledgement. Therefore, using on DDS traffic volumes do not change regardless ratio of IoT data with reliable and/or low latency.

In order to confirm the results of Figure 6 and Figure 7, we explain behavior of message sequence we clarify to behavior of message sequences on point-to-multipoint communication under DDS reliability mode in this part. Figure 8 shows the captured message sequences of point-to-multipoint communication from Node#l (IP address: 192.168.1.1) to Node#2 (IP address: 192.168.1.1) and Node#3 (IP address: 192.168.1.1). It is noted that queue size is set 1 of queuing delay because we assume real-time application. In Figure 8, due to the loss of acknowledgements from Node#2, the message requesting an acknowledgement from Node#1, described as “HEARTBEAT” in Figure 8, continues to be retransmitted. In addition, the results in Figure 8, in DDS reliable node in Figure 9. In this case, since data queue size for retransmission is set to 1, due to the packet loss of acknowledgement from Node 2, the data queue is filled with data to be multicast to Node 2 and Node 3. And the queueing of the next data is blocked. That is, in communications that require an acknowledgement to provide reliability, the delay or loss of response from the destination nodes connected via the unstable communication paths affects the transmission of subsequent packets. This affects QoS support for data with low latency requirements when data with low latency requirements conflicts with data with reliable requirements.

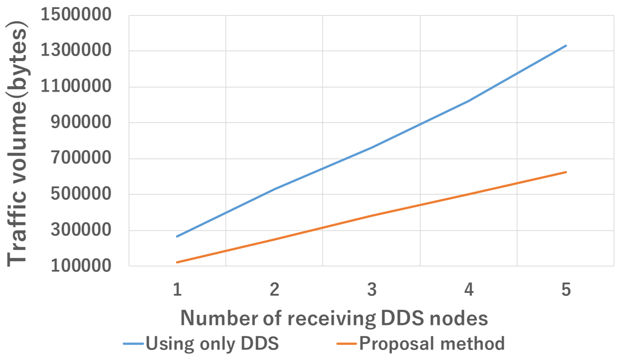


Figure 6: Comparison of traffic on IoT core network.

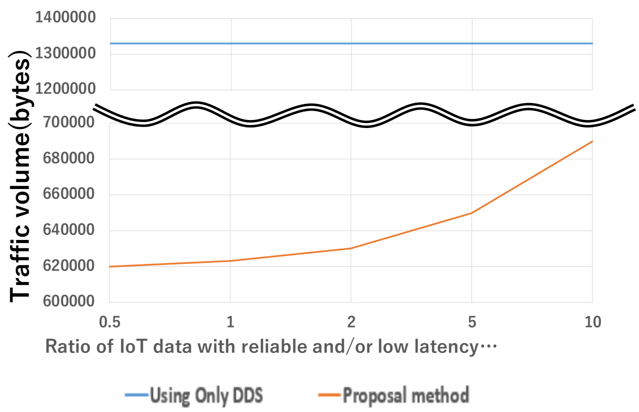


Figure 7: Traffic volume according to ratio of IoT data with reliability and low latency.

192.168.1.1	192.168.1.2	RTPS	49 INFO_TS, DATA
192.168.1.1	192.168.1.3	RTPS	49 INFO_TS, DATA
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.1	192.168.1.3	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.3	192.168.1.1	RTPS	50 INFO_DST, ACKNACK
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT
192.168.1.1	192.168.1.2	RTPS	49,49 INFO_DST, HEARTBEAT

Figure 8: Packet sequence on DDS reliability mode.

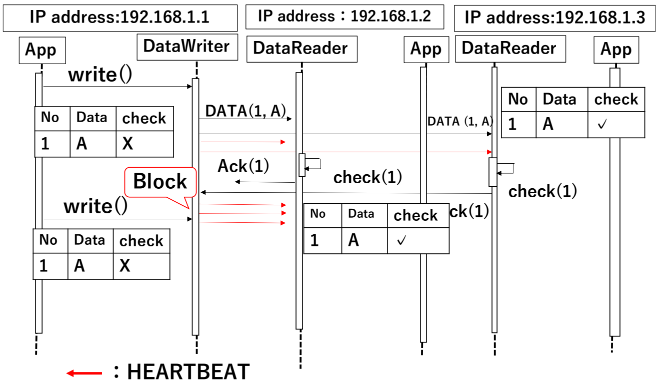


Figure 9: Acknowledgment retransmission sequence in point-to-multipoint communication.

6 CONCLUSION

In this paper, we propose the implementation method for a multicast control coordination scheme in conjunction with DDS as an efficient method of transferring IoT data in IoT systems. We also compared with the data transfer capability of the proposed method and the data transfer using only the DDS function. As results of the evaluation, we confirmed that the proposed method is effective in reducing the traffic volume compared to the communication method using only the DDS function. In addition, we verified the operation in point-to-multipoint in DDS reliability mode. We confirmed that in DDS reliability mode, retransmission for acknowledgement affects transmission of subsequent packets.

In an actual IoT system, various communication characteristics such as packet loss and latency on the communication paths between nodal points are assumed. Therefore, it is necessary to further evaluate the proposed method by considering the communication characteristics of each path in the target IoT system, such as packet loss rate and latency.

REFERENCES

[1] Ministry of Internal Affairs and Communications,”2022 Summary of the State-of-the-Art Report on Information and Telecommunications (in Japanese), “https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r04/html/nf3r1000.html,Accessed, May 2023.

[2] P, Arman Syah, and H, Leslie Hendric Spits Warnars, "Intelligent traffic monitoring system (ITMS) for smart city based on IoT monitoring," 2018 Indonesian Association for Pattern Recognition International Conference. IEEE, pp.161-165, September 2018.

[3] A.P.Plageras, K.E.Stergiou, C.Wang, and H.Gupta, "Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings," Future Generation Computer Systems Vol.82, pp.349-357, May 2018.

[4] M.N.Hossein, M.Bagaa, and T.Taleb. "UAV-based IoT platform: A crowd surveillance use case," IEEE Communications Magazine, vol.55, no.2, pp.128-134, November 2017.

[5] Biswajeeban Mishra and Attila Kertesz, “The Use of MQTT in M2M and IoT Systems: A Survey,” IEEE Access, vol. 8, pp.201071-201086, November 2020.

[6] P. Bellavista, A.Corradi, L.Foschini, and A.Pernafini, " Data Distribution Service (DDS): A performance comparison of OpenSplice and RTI implementations," 2013 IEEE symposium on computers and communications, IEEE, Vol.2, pp.377-383, July 2013.

[7] S.Profanter, A.Tekat, K.Dorofeev, M.Rickert, A.Knoll, "OPC UA versus ROS, DDS, and MQTT: performance evaluation of industry 4.0 protocols," IEEE International Conference on Industrial Technology, pp.955-962, February 2019.

[8] B.Melvin, K.Erkin, P.Marc-oliver, and C.Georg. "Open-source mqtt evaluation," 2021 IEEE 18th Annual Consumer Communications & Networking Conference. IEEE, pp.1-4, January 2021.

[9] L.Xiangtao, H.Ning, Z.Peng and Z.Yu,"The method of Internet of Things access and network communication based on MQTT." Computer Communications Vol.153, pp.169-176, March 2020.

[10] A.Tanushree, N.Payam, M.R.Barzegarn, and V.Luigi. "Multi-level time-sensitive networking (TSN) using the data distribution services (DDS) for synchronized three-phase measurement data transfer," IEEE Access Vol. 7, pp.131407- 131417, September 2019.

[11] L.Jie, Y.Zhou, and H.Chen. "Age of information for multicast transmission with fixed and random deadlines in IoT systems," IEEE Internet of Things Journal, vol.7, no.9, pp.8178-8191, March 2020.

[12] B.Suman, B.Bhattacharjee, and C.Kommareddy. "Scalable application layer multicast," Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications, August 2002.

[13] Z.Ying, B.Li,B, and J.Guo. "Multicast with network coding in application-layer overlay networks," IEEE Journal on Selected Areas in Communications, vol.22, no.1, pp.107-120, January 2004.

Design of Low-rate DoS Attack Detection in Robust WRED

Daisuke Sato†, Hiroshi Inamura*, Shigemi Ishida*, Yoshitaka Nakamura‡,

†Graduate School of Systems Information Science, Future University Hakodate, Japan
‡Faculty of Engineering, Kyoto Tachibana University, Japan
* School of Systems Information Science, Future University Hakodate, Japan

Abstract - Low-rate DoS attacks degrade TCP transmission performance by exploiting a vulnerability in TCP's RTO mechanism. Since the average attack traffic is small, it is difficult to detect it with conventional DoS attack method. One of the problems in existing mechanisms are their low resistance to miss-detections. Based on the characteristics of low-rate DoS attacks, we designed a Robust WRED algorithm as a mitigation mechanism with improved resistance to detection errors.

Keywords: DoS, Low-rate DoS, WRED, Network Security

1 Introduction

Denial of Service (DoS) attacks and Distributed DoS (DDoS) attacks are one of the major threats in the network security field. DDoS attacks are more large-scale attacks than DoS attacks and are a bigger threat. Examples of large-scale DDoS attacks include the world's largest bps (bits per second) type DDoS attack [1] targeting GitHub in 2018, which recorded a maximum of 1.35Tbps; One example is the pps (packets per second) type DDoS attack [2], which was the largest in history, targeting a major European bank. In both cases, service denial was successful by generating a large amount of attack traffic, but it was restored in about 10 minutes by Akamai's response[1][2]. This indicates that DDoS attacks have a large amount of attack traffic, so capturing and detecting the characteristics is easy. Low-rate DoS (LDoS: Low-rate DoS, also known as Low-rate Shrew DoS) attacks[3] can block TCP flow with a low average attack traffic volume[3][4]. Since LDoS attacks have a low average attack traffic volume, detection methods based on traffic volume for DoS attacks and DDoS attacks are difficult to detect[5]. Therefore, research on detection methods of LDoS attacks is required.

RED (Random Early Detection) and WRED (Weighted RED), which are representative algorithms of AQM (Active Queue Management), are widely used on the Internet[6]. However, existing research indicates that RED is vulnerable to LDoS attacks[4][7]. RRED (Robust RED) has been proposed as a method to eliminate the vulnerability of RED to LDoS attacks[8]. RRED protects network resources from LDoS attacks and stabilizes flow by installing detectors and detecting and discarding LDoS attack traffic before queuing by RED. However, the detection conditions for LDoS attacks in RRED's detectors are high in false positives, and there is a possibility that normal flow other than LDoS attack traffic will also be suppressed[8].

While there are efforts to improve the detection accuracy of LDoS attack detection methods, we believe that it is necessary to propose a defense mechanism that assumes that detection results include unconfident results. The degree to which a given traffic is the optimal pulse waveform shape for LDoS attacks can be used to identify and mitigate attack traffic. We propose Robust WRED (RWRED), a multi-class RED that can identify and mitigate attacks using WRED. We conducted simulation experiments and confirmed that the proposed Robust WRED successfully divided bandwidth based on the degree of LDoS conformance.

The remainder of this paper is organized as follows. Section 2 describes the basics of LDoS attack and RED algorithm as well as related work. Section 5 presents the design of our Robust WRED (RWRED), followed by simulation evaluations in Section 6. Finally, Section 7 concludes the paper.

2 LDoS Attack and RED

This section describes the following related techniques necessary for the discussion: Low-rate Shrew DoS attacks intentionally cause continuous TCP retransmission timeouts. This study utilizes RED and its derivative algorithm, WRED.

2.1 TCP Retransmission Time Out

In TCP communication, a retransmission timer is started each time a packet is sent. If no response is received from the receiver within *minRTO* of the minimum RTO (Retransmission Time Out), which is the maximum waiting time of the retransmission timer, the packet is considered to be lost and the packet is retransmitted. The initial value of RTO is determined by the formula (1).

$$\begin{aligned} RTO = \quad \max \quad & \{minRTO, SRTT \\ + \quad & \max(G, RTTAVR \times 4)\} \end{aligned} \tag{1}$$

$$minRTO > SRTT + \max(G, RTTAVR \times 4) \tag{2}$$

where *minRTO* is the minimum value of *RTO*, *SRTT* is the smoothed round trip time (RTT), *G* is the operating system-dependent clock granularity, and *RTTAVR* is the mean deviation of *RTT*. *RTO* is recommended by the IETF to be set to a minimum value of 1s[9]. In many cases, the initial value of RTO is set to minRTO as in the formula (3) because the formula (1) holds in many cases. The initial value of RTO is generally set to minRTO, as in the formula (3).

$$RTO_1 = minRTO \tag{3}$$

$$RTO_i = RTO_{i-1} \times 2 \tag{4}$$

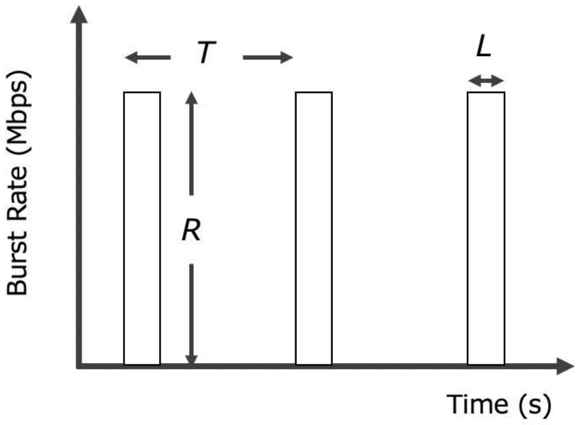


Figure 1: Attack Parameter

In TCP, RFC 6298[10] defines the specification that when it is judged that the same packet has been dropped consecutively, the RTO is retransmitted with the RTO increased by a factor of two. If no response is received from the receiver within *RTO* of a packet sent *RTO* times consecutively, the RTO_i of the packet is set by the expression (4). However, RFC 6298 indicates that the maximum value of *RTO* shall be 60s or longer. If the packet is successfully transmitted, *RTO* is reset to its initial value *minRTO*. This is called Karn's algorithm and is implemented as a retransmission control algorithm in most TCP implementations[11].

2.2 Low-rate Shrew DoS Attack

The LDoS attack was demonstrated in 2003 by Kuzmanovic and Nightly[3]. the LDoS attack consists of three parameters $\langle R, L, T \rangle$, as shown in Figure1. where R is the burst rate, L is the burst length, and T is the burst interval. LDoS attacks are stealthy DoS attacks because they achieve DoS attacks with low average traffic volume and are difficult to distinguish from the general traffic. LDoS attacks are attacks against transport layer protocols that exploit the retransmission timeout mechanism of the TCP retransmission control algorithm. TCP detects a packet loss if no ACK is returned from the destination after waiting *RTO*. If packet loss occurs continuously, *RTO* is set by the formula (4) and increases periodically.

We explain the specific attack method by assuming that the bandwidth and buffer size of the bottleneck link in the network over which the target TCP is communicating is C and B, respectively. First, a target TCP packet is lost by sending enough attack traffic to satisfy C and B. The target TCP packet is continuously lost by sending the attack traffic again at the time when the packet retransmits. The LDoS attack can be considered as a packet that sends attack traffic with a parameter R of C, L long enough to fill B or R long enough to fill C, L long enough to fill B, and L long enough to fill C. If the length of the traffic parameter R is set to *RTT* of the target TCP or T is set to *minRTO* of the target TCP, the TCP flow will be more suppressed[12].

2.3 RED Algorithm

The RED algorithm is a buffer management technique[13] proposed as a congestion avoidance strategy for TCP networks. The main design goal of the RED algorithm is to provide congestion avoidance to the network by monitoring and controlling the average queue size. The RED algorithm provides early congestion detection by monitoring the average queue size, notifying hosts by marking packets to control the average queue size and dropping packets early. When a sender is notified of early congestion, it can reduce the amount of traffic, thereby reducing network performance degradation.

Initial congestion detected by the RED algorithm is notified to the host by either dropping packets or marking packets. Packets are divided into three categories based on the average queue size, which is divided into two thresholds, the maximum threshold, and the minimum threshold, and each category is processed differently. If the average queue size is less than the minimum threshold, the packet is enqueued without packet marking or packet discarding. When the average queue size is greater than the minimum threshold and less than the maximum threshold, packet marking and discarding are performed based on the packet marking probability, which increases proportionally to the average queue size. If the average queue size is greater than the maximum threshold, packet marking or packet discarding is performed for all packets.

2.4 WRED

The WRED (Weighted RED) algorithm can run the RED algorithm separately for each traffic class. [14][15]. The WRED algorithm integrates the functionality of the RED algorithm with the Precedence functionality to achieve preferential traffic processing for packets with high Precedence[6]. Here, the WRED algorithm can set minimum and maximum thresholds based on Precedence, and the thresholds are generally set higher for higher priorities[13]. The WRED algorithm refers to the precedence when discarding packets and can manipulate the discard probability using thresholds by precedence. Since the maximum and minimum thresholds can be set for each precedence, packets with higher precedence can be enqueued preferentially.

3 Related Study

LDoS attacks are difficult to detect using volume-based detection methods for DoS and DDoS attacks because LDoS attacks succeed with low average attack traffic by generating attack pulses in RTO cycles. Therefore, various approaches have been studied for the detection of LDoS attacks, and some of them are based on the attack parameters such as burst length L and burst interval T to detect the attacking traffic. [16]. The method detects LDoS attack traffic as 1-second bursts whose burst length L is greater than or equal to the RTT of other flows connected to the same server and whose burst interval T is *minRTO*. This detection method is a basic detection method for detecting pulse-wave attack traffic, which is a characteristic of LDoS attacks. This LDoS attack detection method is referred to as the attack parameter-based LDoS attack detection method and the conventional LDoS attack detection method. The conventional LDoS attack detection method assumes only optimized attack pulses, and it is believed that manipulating the attack parameters of LDoS attacks and changing them from the optimal parameter values can avoid detection.

As one of the defense methods against the LDoS attack proposed by Kuzmanovic et al.[3], we investigate attack mitigation by detection and throttling at routers. RED-PD (RED with Preferential Dropping)[17] or RED is applied to routers to verify the identifiability of attack flows for LDoS attacks by comparing the target TCP throughput. RED-PD detects LDoS attack traffic by setting a target bandwidth, but its performance is shown to be insufficient as a detection method. PD could be improved by applying strong bandwidth limiting only to flows with high potential for LDoS attack traffic, instead of bandwidth limiting all flows fairly.

There is an existing detection method[16] that uses burst length (L) and burst interval (T) as detection criteria. This detection method measures the length of burst transmission time and the time interval between burst transmission in the flow to be detected and detects attacks based on the similarity between the optimal values of burst length (L) and burst interval (T). Flows with high similarity are judged as attack flows, and those with low similarity are judged as normal flows. The optimal values of burst length (L) and burst interval (T) are the same as the recommended values of L for 2RTT-3RTT and 1-second burst interval (T) for other flows connected to the same server. Thus, the optimal value of the attack parameter is used for detection.

However, existing detection methods may falsely detect attack flows as normal flows. Since the optimal values of burst length (L) and burst interval (T) are used as the detection criteria, LDoS attacks that manipulate burst length (L) may be falsely detected as normal flows. For example, consider the case of an attack in a network where the bandwidth of the bottleneck link is C. If the optimal attack parameters for detection are $R = C$, $L = 200$ms, and $T = 1000$ms, and the attack parameters set for the attack flow are $R = C$, $L = 100$ms, and $T = 1000$ms, the attack If the attack parameters are set to $R = C$, $L = 100$ms, and $T = 1000$ms, the attack flow is mistakenly judged as a normal flow and false detection occurs.

4 Research Goal and Proposed Method

The objective of this study is to improve the immunity to miss-detection of existing suppression mechanisms. The existing suppression mechanism, Robust RED[8], has a low tolerance to miss-detection, and when a miss-detection occurs, it may protect attacking flows and deter normal flows. The reason for the low immunity to positive in existing suppression mechanisms is that they deter attacks by processing only two types of flows: 100 % attacking flows and 100 % normal flows.

We propose a method of suppression based on the *degree* of LDoS attack flow, instead of classifying it in two ways: attack flow or normal flow. In addition to discarding and queuing, we also provide multiple partial discards. Specifically, we construct an attack detection and mitigation mechanism based on the RED algorithm for LDoS attacks. Based on the existing Robust RED algorithm, our method is multi-classed by replacing the queuing mechanism from RED to WRED and mitigates LDoS attacks, and protects TCP flows by prioritizing traffic according to the probability of the traffic being LDoS attack traffic as identified by the attack detection part. The proposed method is called Robust WRED. The proposed method is called Robust WRED (RWRED).

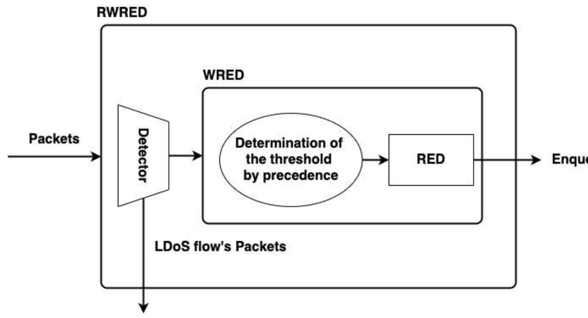


Figure 2: RWRED system

5 Design of Robust WRED

5.1 Overview of RWRED

The RWRED is designed based on the RRED[8] and has a Detector block in front of the WRED block as shown in Figure 2. The Detector block identifies incoming packets by the flow and calculates the probability of LDoS attack traffic for each flow. The Detector block identifies each input packet and calculates the probability that the packet is LDoS attack traffic for each flow. The packets are identified by source IP address, destination IP address, source port number, destination port number, and flow protocol, which are obtained from the header of the input packet. Depending on the calculated probability of the packet being LDoS attack traffic, the Detector block either drops the packet or marks it as a Precedence (IP Precedence). Packets with a high probability of being LDoS attack traffic are queued inferiorly, and packets with a low probability of being LDoS attack traffic are queued preferentially. It behaves as if bandwidth is reserved for the flow with a low probability of being LDoS attack traffic.

Since the RWRED mechanism can prioritize LDoS attack traffic according to its probability of being LDoS attack traffic, it is thought to be able to detect and deter LDoS attack traffic while protecting normal traffic.

5.2 LDoS Conformance Index for Attack Detection

We define the LDoS Conformance Index (LCI), which indicates the degree to which a given flow is a malicious LDoS/LDDoS traffic, as the basis for Robust WRED priority control. We apply an existing detection method[16], to define LDoS Conformance Index for detection based on attack parameters. Since existing detection methods use the optimal values of burst length (L) and burst interval(T) as criteria for detection, we define LDoS Conformance by the similarity between burst length (L) and burst interval(T). From the above, we define

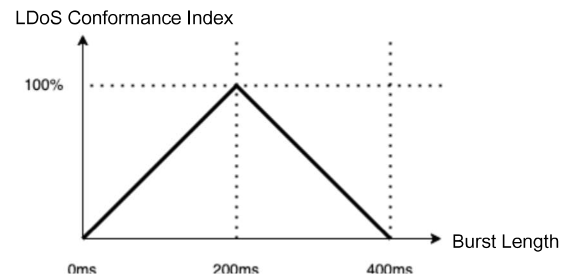


Figure 3: LCI for Burst Length(L)

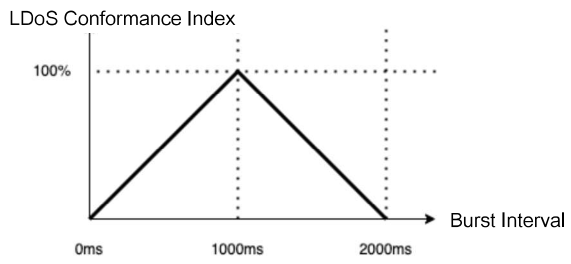


Figure 4: LCI for Burst Interval(T)

the Overall LDoS Conformance Index as the product of the LDoS Conformance Index of burst length (L) and the LDoS Conformance Index of burst interval (T).

Since the optimal attack parameters for detection are L = 200ms and T = 1000ms, burst length (L) takes values in the range of 0ms-400ms based on 200ms. The definition of the LDoS Conformance Index for burst length (L) is shown in Figure 3. For example, if the length of the burst transmission in the flow to be detected is 0 ms and 400 ms, the LCI of the burst length (L) is 0.0 percent. When the length of burst transmission in the flow to be detected is 100ms and 300ms, LCI of the burst length (L) is 50.0%.

LCI for burst interval (T) is similar to that of the burst length (L), taking values in the range of 0ms-2000ms with 1000ms as the reference. Figure 4 shows the definition of LCI for burst interval (T). Thus, we define the LDoS Conformance Index for attack parameter-based detection as the product of the LCI for burst length (L) and the LCI for burst interval (T). To reproduce the conventional detection method, 50% LCI is used as the decision threshold.

5.3 Link Share and RED Threshold derived from LCI

Set the threshold value to be given to the WRED queue control from the ideal bandwidth share in order to split the flow according to the LDoS Conformance Index.

In order to suppress LDoS attack traffic and allow normal traffic to pass, the appropriate bandwidth to be given to the flow in question using the LDoS conformance is considered to be $LinkBandwidth \propto (1 - LCI_i)$. To express the per-flow share of the link bandwidth, we define the per-flow Normalized Flow Share (NFS), which is normalized with respect

Table 1: Ideal Link Share and LCI

	LCI	ideal Link Share
TCP Sender 1	0.00(0%)	100[Mbps]×0.250=25.0[Mbps]
TCP Sender 2	0.10(10%)	100[Mbps]×0.225=22.5[Mbps]
TCP Sender 3	0.20(20%)	100[Mbps]×0.200=20.0[Mbps]
TCP Sender 4	0.30(30%)	100[Mbps]×0.175=17.5[Mbps]
TCP Sender 5	0.40(40%)	100[Mbps]×0.150=15.0[Mbps]

to the link bandwidth, as follows

$$NFS_i = \frac{(1 - LCI_i)}{\sum_{f \in AllFlowintheLink}(1 - LCI_f)} \quad (5)$$

$$FS_i = (\text{Link-bandwidth}) \times NFS_i \quad (6)$$

Ideally, link bandwidth should be divided according to the LDoS Conformance Index for each flow. Equation (5) is a formula for calculating the ideal bandwidth partition for each flow, and equation (6) is a formula for calculating the ideal acquisition bandwidth for each flow. Table reftab:ideal shows the relationship between the ideal acquired bandwidth and LDoS conformance calculated using the formula (6) . As described above, the ideal link bandwidth share according to LDoS Conformance Index is defined by (5).

$$Max\_thresh_i = (\text{Upper limit of threshhold}) \times NFS_i \quad (7)$$

To achieve bandwidth partitioning according to LDoS conformance, the threshold is optimized. The ideal bandwidth allocation is given by the formula (5), and the queue control parameter Max_thresh_i for each flow given to WRED is determined using the formula (7).

5.4 Handling Non-Adaptive Protocols

In order to treat non-adaptive protocol flow fairly with adaptive protocol flow, the minimum and maximum thresholds of the RED algorithm are set to the same value and a process such as Tail Drop is performed. Non-adaptive flows are trasmitted by UDP protocols, while adaptive flows are transmitted by TCP in this paper. There is a difference between TCP and UDP in whether or not the amount of data sent is reduced when packets are dropped. This difference may cause non-adaptive flow to occupy undue bandwidth. The proposed mechanism solves this problem by limiting the queue lengths available for non-adaptive flow for providing extra resources for adaptive flow. For the non-adaptive flow, we apply a Tail Drop-like process by setting the minimum and maximum thresholds to the thresholds calculated by the minimum threshold, thus, the suppression is stronger than for the non-adaptive flow.

6 Evaluation Plan of RWRED performance

6.1 Overview of Experiments

The purpose of the evaluation simulation is to clarify that the proposed mechanism, Robust WRED, is more tolerant to miss-detection than the existing mechanism, Robust RED.

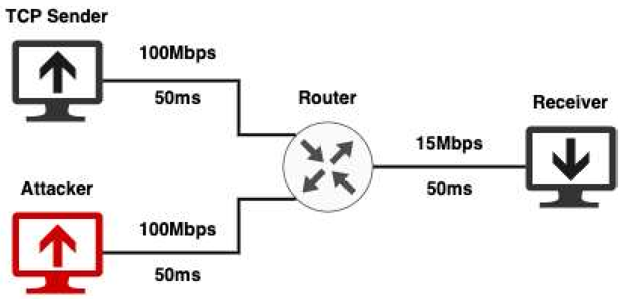


Figure 5: Simulation Scenario for Evaluation 1

We evaluate the resistance to miss-detection of the proposed mechanism and the existing mechanism.

Existing detection mechanisms are likely to cause detection errors. Based on the simulation results of the validation in Section 6.2, the existing detection has a high likelihood of overlook LDoS attacks with burst length (L) set outside of the optimal value, as normal flows, and a high likelihood of falsely detecting TCP flows with RTT set close to the optimal burst length (L) as attacking flows. (L) is more likely to be miss-detected as an attack flow than a TCP flow whose RTT is set close to the optimal value of burst length (L).

6.2 Experiment Setting

Figure 5 shows the evaluation simulation environment used in Evaluation 1. The evaluation system was constructed using ns-3[18], a discrete event network simulator. The network consists of four nodes: TCP Sender, which sends TCP packets, Attacker, which is the LDoS attack node, Router, which is the intermediate node, and Receiver, which is the receiving node. The bandwidth of the links between the TCP Sender and Router and between the Attacker and Router is 100 Mbps, the propagation delay is 50 ms, and the bandwidth of the link between the Router and Receiver is 15 Mbps. The bandwidth and propagation delay of the links between the Router and Receiver are set to 15 Mbps and 50 ms, respectively. The bottleneck link of the evaluation network for the proposed mechanism in this time domain detection is the link between the Router and Receiver, and the bottleneck queue is the output queue of the Router. The bottleneck queue is queue-controlled by RED, and other queues are queue-controlled by Tail Drop, with the minimum threshold of RED set to 225 and the maximum threshold set to 450, and queue control is set on a per-packet basis. TCP packet size is set to 1400Byte.

6.3 Miss-detection in Attack Parameter based Detector

Possible miss-detections in detection based on attack parameters are illustrated below.

Example.1 A false negative occurs when the configuration values deviate from the optimal attack parameters in the LDoS attack model. For example, consider the case of an attack on a network where the bandwidth of the bottleneck link is C. If the optimal attack parameters for detection are $R = C$, $L = 200$ ms, and $T = 1000$ ms, and if the attack parameters set for attacking flows are $R = C$, $L = 100$ ms, and $T = 1000$ ms, the attack If the attack parameters are set to $R = C$, $L = 100$ms, and $T = 1000$ms, the attack flow is mistakenly judged as a normal flow and miss-detection occurs.

Example.2 A false positives occurs when the length of time of bursts of normal flow and the time interval between bursts of normal flow are similar to the optimal attack parameters in the LDoS attack model. False positives occur when the RTT of normal flow is similar to the optimal attack parameters in the LDoS attack model. For example, if the normal flow is TCP and the RTT is set to 200 ms, false positives occur because the length of the burst transmission time is similar to the attack flow $L = 200$ ms in terms of waiting for ACKs. If the normal flow is assumed to be a TCP flow and the RTT is set to 1000 ms, false positives occur because the time interval between burst transmission is similar to $T = 1000$ ms.

7 Conclusion

In this paper, we proposed Robust WRED as a mitigation mechanism with improved resistance to detection errors in attack parameter based detection.The proposed mechanism introduces the per-flow LDoS Conformance Index and divides the link bandwidth according to the LDoS Conformance Index for the flow. The proposed mechanism will be effective suppression and mitigation mechanisms for LDoS/LDDoS attacks, which are difficult to detect.

REFERENCES

[1] Sam Kottler. February 28th ddos incident report. available from ⟨https://github.blog/2018-03-01-ddos-incident-report/⟩, 2018.

[2] Tom Emmons. Flargest ever recorded packet per second-based ddos attack mitigated by akamai. available from ⟨https://blogs.akamai.com/2020/06/largest-ever-recorded-packet-per-secondbased-ddos-attack-mitigated-by-akamai.html⟩, 2020.

[3] Aleksandar Kuzmanovic and Edward W. Knightly. Low-rate tcp-targeted denial of service attacks: The shrew vs. the mice and elephants. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, number 12 in SIGCOMM ’03, pages 75–86, 2003.

[4] Aleksandar Kuzmanovic and Edward W Knightly. Low-rate tcp-targeted denial of service attacks and counter strategies. *IEEE/acm transactions on networking*, 14(4):683–696, 2006.

[5] Wu Zhijun, Li Wenjing, Liu Liang, and Yue Meng. Low-rate dos attacks, detection, defense, and challenges: A survey. *IEEE Access*, 8:43920–43943, 2020.

[6] Inc. Cisco. Qos: Congestion avoidance configuration guide, 2017.

[7] Mina Guirguis, Azer Bestavros, and Ibrahim Matta. Exploiting the transients of adaptation for roq attacks on in-

ternet resources. In *Proceedings of the 12th IEEE International Conference on Network Protocols, 2004. ICNP 2004.*, pages 184–195. IEEE, 2004.

[8] Changwang Zhang, Jianping Yin, Zhiping Cai, and Weifeng Chen. Rred: robust red algorithm to counter low-rate denial-of-service attacks. *IEEE Communications Letters*, 14(5):489–491, 2010.

[9] IETF. Transmission control protocol. available from⟨https://tools.ietf.org/html/rfc793⟩, 9 1981.

[10] IETF. Computing tcp's retransmission time. available from ⟨https://tools.ietf.org/html/rfc6298⟩.

[11] Tanenbaum Andrew S. and Wetherall David J. Computer networks fifth edition, 2011.

[12] Jingtang Luo, Xiaolong Yang, Jin Wang, Jie Xu, Jian Sun, and Keping Long. On a mathematical model for low-rate shrew ddos. *IEEE Transactions on Information Forensics and Security*, 9(7):1069–1083, 2014.

[13] Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, 1993.

[14] Mark Wurtzler. Analysis and simulation of weighted random early detection (wred) queues. *Diss., University of Kansas*, 2002.

[15] R. Makkar, I. Lambadaris, J.H. Salim, N. Seddigh, B. Nandy, and J. Babiarz. Empirical study of buffer management scheme for diffserv assured forwarding phb. In *Proceedings Ninth International Conference on Computer Communications and Networks (Cat.No.00EX440)*, pages 632–637, 2000.

[16] A. Shevtekar, Karunakar Anantharam, and N. Ansari. Low rate tcp denial-of-service attack detection at edge routers. *IEEE Communications Letters*, 9(4):363–365, 2005.

[17] R. Mahajan, S. Floyd, and D. Wetherall. Controlling high-bandwidth flows at the congested router. In *Proceedings Ninth International Conference on Network Protocols. ICNP 2001*, pages 192–201, 2001.

[18] nsnam.org. ns-3 — a discrete-event network simulator for internet systems. available from⟨https://www.nsnam.org/⟩.

A Two-Dimensional-Trust-Based Recommendation Method for Job Placement Assistance

Yoshinobu Kawabe and Tetsuhisa Oda

Aichi Institute of Technology
Yachigusa 1247, Yakusa-cho, Toyota, Aichi 470-0392, Japan
E-mail : { kawabe, oda }@aitech.ac.jp

Abstract - Based on university students' employment preferences, we develop a recommender system for employment opportunities that match those preferences. The typical concerns of university students are job type and work location. In our approach, students can answer desired and undesired items for these criteria as text in an open-ended response format. After analyzing the text, we calculate a trust value, representing an agreement between the student's preferences and the company profile. Since students describe desirable and undesirable items independently, college students' preferences may be inconsistent to the company profile description, leading to difficulty coping with the inconsistency. Our trust values are a pair of trust (the degree to which desirable items are satisfied) and distrust (the degree to which undesirable items are included). Hence, we can directly deal with this kind of inconsistency. In this study, by applying the result from the fuzzy logic, we determine a trust value without inconsistency. This paper also discusses how to determine trust and distrust levels. The trust level is determined by the degree to which the company profile text matches the student's expectations and is assigned a value between 0 and 1. On the other hand, the distrust level is determined by whether or not the student's undesired items appear in the company profile and is assigned a value of 0 or 1. We discuss validating these criteria by comparing them with a case study for deciding a trust value to rescue request tweets posted in a disaster.

Keywords: On-Line Trust, 2D Trust Model, Fuzzy Logic, Job Placement Assistance

1 Introduction

This paper introduces a fuzzy-logic-based method for recommending companies that match the job preferences of university students based on their job preferences. University students' main interests regarding companies they are employed by are the type of job and workplace location. In this study, we asked university students to answer "what they want" and "what they do not want" regarding these conditions in the form of free descriptions, and based on the results, we calculated a two-dimensional trust value of [1][2]. The two-dimensional trust values are a pair of a trust level (the degree to which desired items are satisfied) and a distrust level (the degree to which unwanted items are included) and correspond to the observation points of Oda's FCR method [3][4].

Since students tend to describe their preferences and non-preferences independently, the requirements of college students regarding corporate profiles may be inconsistent. We computed the net goodness of fit (called a one-dimensional trust value) of students' preferences to the company profile using the integrated value of the FCR method.

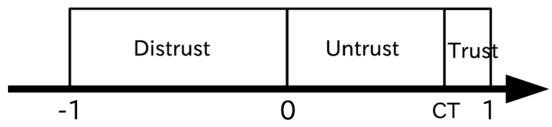


Figure 1: Marsh's Trust Model

In this study, the assignment of the degree of trust and the degree of distrust was conducted using the same approach as in our previous study, the assignment of a two-dimensional trust value to rescue request tweets posted during a flood disaster [5]. Specifically, trustworthiness is determined by the degree of compatibility between the text of the company profile and the student's expectations and is a value between 0 and 1. On the other hand, the distrust level is assigned based on whether or not the student's "undesired items" appear in the company profile and has a value of 0 or 1. Although it is possible to determine the distrust level based on "how many unwanted items appear" using the same criterion as the trust level, we did not do so in this study. To validate the criterion of this paper, we discuss how to determine the distrust level properly.

2 Preliminaries: Two-Dimensional Trust Representations

Marsh and Dibben introduced trust values, which range from -1 to 1, and classified trust notions into *trust*, *distrust*, *untrust*, and *mistrust* [6]. We extended their trust values for two dimensions [1][2] to address inconsistent trust evaluations. This section explains some basics of (mainly two-dimensional) trust computations.

2.1 Conventional Trust Values

A conventional trust value is a real number in $[-1, 1)$. Readers interested in the details of calculating trust values can find them here [6], but in this paper we directly handle the calculated trust values. Marsh and Dibben introduced the following four notions of trust (see also, Fig. 1):

- *Trust*: The notion represents a case where a trust value is positive and exceeds a predefined value called a cooperation threshold. In this case, a trustee should be

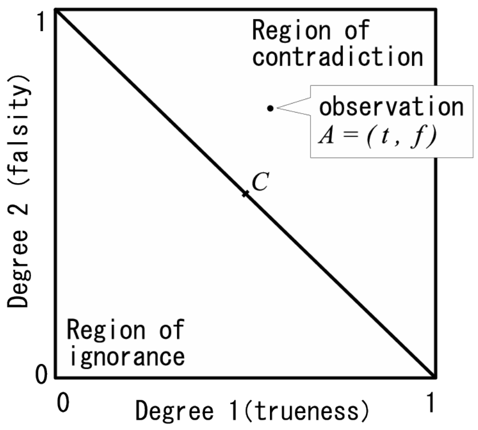


Figure 2: Two-dimensional HLS

trusted, and the trust value is regarded as a measure of how much an agent believes the trustee.

- *Distrust*: Here the trust value is negative, and an agent believes that a trustee will actively work against her in a given situation.
- *Untrust*: Although the trust value is positive, it is not high enough to produce cooperation. An agent cannot determine if a trustee is actually trustworthy.
- *Mistrust*: Initial trust has been betrayed. More precicely, mistrust represents a situation either a former trust was destroyed or a former distrust was healed. The mistrust notion is a time-related trust property.

For these properties, see studies by Primiero [7] (on distrust and mistrust) and [8] (on trust and distrust).

2.2 Fuzzy-Logic-Based Two-Dimensional Trust Values

Marsh and Dibben classified trust notions in a one-dimensional setting, i.e., trust and distrust are at both extremities. However, Lewicki et al. [9] suggested that trust and distrust are entirely different dimensions from a psychological viewpoint. Trust is a property closely related to human impressions, and a technique for impression formation based on mathematical psychology should be applied to trust values.

Oda [3][4] developed a Fuzzy-set Concurrent Rating (FCR) method with fuzzy logic that enables us to measure and analyze human impressions. The FCR method allows two or more dimensions for representing a truth value, and our two-dimensional trust representation is an application of the FCR method for trust and distrust notions.

2.2.1 FCR Method

The FCR method employs the Hyper Logic Space model (HLS) as a logic space for multiple-dimensional multiple-valued logic. Figure 2 shows a two-dimensional space based on $true$ and $false$. For any $t, f \in [0, 1]$, pair (t, f) is called an observation. t and f are independent; we do not assume such conditions as $t+f = 1$. We call $\{(t, f) \mid t, f \in [0, 1] \wedge t+f > 1\}$ the region of contradiction. $\{(t, f) \mid t, f \in [0, 1] \wedge t+f < 1\}$ is called the region of ignorance, or the region of irrelevance. Finally, $\{(t, f) \mid t, f \in [0, 1] \wedge t + f = 1\}$ is the consistent region.

Given observation (t, f), we need to calculate an actual truth value, which is called an integration value. Integration values can be calculated in several ways, and we employ the reverse-item averaging method, where integration value I_2 is defined with $I_2(t, f) = \dfrac{t + (1 - f)}{2}$. The integration value is the average of the degree of the positive elements and the complementary degree of the negative elements.

Another important value in the FCR method is the degree of contradiction [3][4] or the contradiction-irrelevance degree. In the field of personality psychology, some situations are allowed, including “I like it, but I don’t like it” or “I don’t care for it at all.” The degree of such confusion/irrelevance is formulated with the degree of contradiction. For observation (t, f), degree of contradiction $C(t, f)$ should satisfy $C(t, f) = 1$ for complete confusion, $C(t, f) = -1$ for complete ignorance, and $C(t, f) = 0$ for a consistent situation. $C(t, f) = t + f - 1$ is usually employed where $C(t, f)$ represents the distance between (t, f) and the consistent region.

2.2.2 Two-Dimensional Trust Model

We employ the degrees of trust $Trust$ and distrust $DisTrust$ defined with $Trust = DisTrust = \{ v \mid 0 \le v \le 1\}$ and define a two-dimensional trust value as an element of $Trust \times DisTrust$. Observation $(1, 0) \in Trust \times DisTrust$ has a high degree of trust and a low degree of distrust and represents a case where a trustee is completely trusted; this observation corresponds to trust value 1 of [6]. Observation $(0, 1)$ represents a case of complete distrust and corresponds to trust value -1. Observation $(0.5, 0.5)$, which falls exactly between $(1, 0)$ and $(0, 1)$, corresponds to 0 in conventional trust values.

To define such trust notions as trust, distrust, and untrust in our two-dimensional trust model, we employ the following transformation:

$$\left[\begin{pmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{pmatrix} \left\{ \begin{pmatrix} t \\ d \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} + \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}\right] \times \frac{1}{\frac{\sqrt{2}}{2}}$$
$$= \begin{pmatrix} t - d \\ t + d - 1 \end{pmatrix} = \begin{pmatrix} i \\ c \end{pmatrix}.$$

First element $i = t - d$ can be calculated with the reverse-item averaging method. Actually, the value of i is calculated by normalizing $I_2(t, d)$ to be a value in region $[-1, 1]$; note that the range of integration value $I_2(t, d)$ was originally $[0, 1]$.

The value of i was regarded as a conventional trust value given by Marsh and Dibben. From the definition of $i = t - d$, a net trust value is calculated by subtracting the degree of trust from the degree of distrust, which matches our intuition. In Figure 2, the consistent region is the line between $(1, 0)$ and $(0, 1)$ and corresponds to the set of conventional trust values. Observation (t, d) in the consistent region satisfies $t + d = 1$ and is regarded as an assumption on the trust and distrust degrees. The theory of conventional trust values implicitly introduces this assumption.

Preferred items regarding work location:
I prefer to work and live in Nagoya or Tokyo. It is also OK to work in my home town Hamamatsu.

NG items regarding work location:
I'm not particularly eager to go to outlying areas such as Hokkaido, Okinawa, and Shikoku.

Preferred occupations:
I am considering IT sales, planning, other sales jobs, financial institutions, and recruiting services.

NG Occupations:
I have no NG occupations.

Figure 3: Description of One Student's Wishes

Trust notions are defined with the value of i. Let CT be a cooperation threshold. If we have $i = t - d \geq CT$, then it is a case of trust; if i is negative then it is case of distrust; if we have $0 \leq i < CT$, then it is a case of untrust. Note that for the case of distrust, condition $i < 0$ is equivalent to $t < d$; i.e., a trustee is distrusted if the degree of distrust exceeds the degree of trust.

3 Calculation of Two-Dimensional Trust Values Based on Matching of Employment Preferences And Job Offers

In this section, we deal with the issue of helping university students find jobs by applying two-dimensional trust values. Specifically, we evaluate the degree to which a description in job openings matches the student's preferences as a two-dimensional trust value.

This study focuses on students in the Department of Information Science at Aichi Institute of Technology. About 80% of the students commute from the Tokai region of Japan. The general trend is that they are strongly oriented toward the IT industry and hope to live in their hometown (i.e., they wish to work in the Tokai region). The number of job openings is about 15,000, mainly for engineering students.

3.1 Preprocessing: Extracting Workplace Locations And Job Types

We first obtained information from the students on their "preferences" and "things to avoid" about work location and job type. This information was obtained through open-ended questions; we show an example of the result in Figure 3.

We use the morphological analyzer MeCab [10] to extract words related to place names and occupations. Then, we conduct a keyword search to find a match between the company profile text and student preferences.

To find a better match, after extracting proper nouns, such as the region's name, we add information on neighboring areas, considering the characteristics of the target student's preferences; especially, note that students like to get a job on local orientation toward the Tokai region. For example, if the student wishes to work in "Nagoya," the neighboring districts "Aichi" and "Owari" are also used as keywords for the search.

- XXX Corporation
  - **Tokyo**, Sendai, **Nagoya**, Osaka, Hakata, and 40 other locations. Miyagi, **Tokyo**, Nagano, **Aichi** (**Nagoya**), Osaka, and Fukuoka
  - Software, **Information** Processing, and **Information Services**. Development, sales, and maintenance of office equipment (**information** processing equipment) and systems. Listed on the first section of the Tokyo Stock Exchange. Flextime system available (core hours: 11:00-15:00). Telework system available. Clients include government agencies, **financial** institutions, etc.
  - **Sales jobs**, system engineers, customer engineers, technical staff, development, and research staff. A driving license is required for some positions.
  - (The rest of this job offer contains information on capitalization, annual sales, etc.)

Figure 4: Example of Job Opening

Similarly, in Figure 3, the student has listed Hokkaido, Okinawa, and Shikoku as their least preferred work regions. These are not exact names of regions but rather broad regional names. Most of the place names in the job openings (typically, location of branch offices, factories, etc.) are city names. Therefore, for the three words "Hokkaido, Okinawa, and Shikoku," we will also add more detailed place names within the regions for matching, such as "Ehime, Asahikawa, Kagawa, Takamatsu, Kochi, Sapporo, Muroran, Matsuyama, Tokushima, Naha, Hakodate, and Kitami."

In finding a better match on job categories, because target students major in information technology, some search words are added before searching. For example, when "IT" appears in the description of desired items, the term "information" is also used.

On the other hand, to conduct a job matching, we created a dictionary with job information based on the information in the job posting, as shown in Figure 4; such a posting contains information such as the location of the head office, the place of work, and the text of the company profile. In the example in Figure 4, we underline the occurrence of words that match the student's desired information described above.

3.2 Calculating Two-Dimensional Trust Values

In this study, we calculated the degree to which words from the student's preferences appeared in the job postings for each work location and job type. For example, in the above case, the evaluation values are as follows[1]:

- The matching degree of location preference is 0.6; actually, 3 out of 5 words "Aichi, Tokyo, Owari, Hamamatsu, and Nagoya" as matching candidates match the

[1] Although we give an example in English, we actually search for Japanese keywords.

job information; and

- The match level for job type is 0.8; 4 out of 5 possible matching words "information, sales, finance, recruiting, and service" match the job information.

In this study, we used the goodness of fit as trust values concerning the work location or the type of job. The trust level is the degree to which the text of the company profile matches the positive aspects desired by the student. The value is a real number between 0 and 1.

On the other hand, the distrust level was determined by whether or not the words that the students did not want appeared in the company profile. This decision is based on the fact that students tend to avoid firms whose profile contains items the students wish to avoid. This results in a distrust level of either 0 or 1.

In Figure 3, the student listed Hokkaido, Okinawa, and Shikoku as the least preferred work locations. For the case of this student, the level of distrust regarding the work location is determined by whether or not the name of one of the following places or regions appears in the work location field of the job opening: Ehime, Asahikawa, Okinawa, Kagawa, Takamatsu, Kochi, Sapporo, Shikoku, Muroran, Matsuyama, Tokushima, Naha, Hakodate, Hokkaido, or Kitami. For the firm shown in Figure 4, no such place name appears in the job opening. Thus, this student's distrust level concerning this firm is 0.[2] Therefore, the two-dimensional trust value about the workplace location of this firm is $(0.6, 0)$.

In the above matching, we calculated the trust and distrust levels independently. Therefore, there may exist inconsistencies in the trust and distrust levels. For example, for the two-dimensional trust value $(0.6, 0)$ above, we can calculate the degree of the discrepancy from a consistent trust value using Oda's fuzzy logical approach; the degree is $|(0.6 + 0) - 1| = | - 0.4| = 0.4$. Also, we can calculate the observation point without inconsistency with Oda's reverse-item averaging method, and the result is $(0.8, 0.2)$; note that the first element of the pair is the integration value.

One can also obtain Marsh's one-dimensional trust value from the two-dimensional trust value, considering the observation point $(0, 1)$ as Marsh's score -1 and $(1, 0)$ as score 1, respectively. We can see the one-dimensional trust level as the "net trust value" obtained by subtracting the distrust level from the trust level. From the viewpoint of fuzzy logic, a one-dimensional trust value is a trust level without the inconsistency of evaluation, corresponding to the integration value[3].

In the above example, the distrust level is 0, so the first element of the pair $(0.6, 0)$ and the corresponding one-dimensional trust value 0.6 coincide. If the job offer includes a negative statement regarding the location of the job, the trust value would be 1. In this case, the two-dimensional trust value is $(0.6, 1)$, and the Marsh trust value is negative -0.4.

[2] Note that the job offer form also states "40 other locations," which may include specific locations the students do not wish to work at, although it is not specified. Only text matching was used in this study, and semantic analysis was not conducted.

[3] There is a difference in that the integration value ranges from 0 to 1, while the one-dimensional trust value ranges from -1 to 1.

- XXX Corporation $(0.8, 0.6, 451)$
  - Tokyo, Sendai, Nagoya, Osaka, Hakata, and 40 other locations. Miyagi, Tokyo, Nagano, Aichi (Nagoya), Osaka, and Fukuoka
  - Software, Information Processing, and Information Services. Development, sales, and maintenance of office equipment (information processing equipment) and systems. Listed on the first section of the Tokyo Stock Exchange. Flextime system available (core hours: 11:00-15:00). Telework system available. Clients include government agencies, financial institutions, etc.
  - Sales jobs, system engineers, customer engineers, technical staff, development, and research staff. A driving license is required for some positions.
- YYY Corporation $(0.6, 0.8, 2000)$
  - Tokyo Hamamatsu, Nagoya, and Osaka. Chiba, Tokyo, Kanagawa, Shizuoka, Aichi, Kyoto, Osaka, and Hyogo.
  - System design, construction, operation, maintenance, various software development, etc. We support IT infrastructure in a wide range of industries from finance, manufacturing, telecommunications, and services to space development. We provide solutions in various phases from system development, operation, and maintenance to infrastructure construction.
  - System Engineer.
- ZZZ Corporation Nagoya Branch $(0.6, 0.8, 1219)$
  - ...

Figure 5: Output Result (Partly)

Similarly, for the job type, we defined the distrust level based on whether or not the job type the student does not want appears in the job offer text. Since the student's statement in the previous section states that "I have no NG occupations," the value of the distrust level is 0. Therefore, the two-dimensional trust value for the job category is $(0.8, 0)$.

3.3 Presenting A List of Recommendatable Companies

Applying the above technique, we give trust values for work locations and job types for each job posting's entry. Then, we compute one-dimensional trust values to find a match between student and company profiles. Finally, a list of 20 job openings is presented to the student. Figure 5 shows an example of the output; actually, they are in Japanese. The list of the output is sorted in the order of:

1. one-dimensional trust value for job type,
2. one-dimensional trust value for work location, and
3. the number of employees of the company.

These elements are shown as a triple placed on the right-hand side of the company name.

In this study, we implemented the recommender system on Linux (Ubuntu 22). We used a set of job data in a CSV format with approximately 15,000 entries, extracted keywords, and converted them into a form Common Lisp could handle. The conversion was conducted with a shell script, using the morphological analyzer MeCab for natural language processing. For collecting students' preferences, we used Moodle. Students can describe their preferences in a free-style text format, and we extracted keywords such as work location or occupations from the text, and the resulting keywords were given to a Lisp program. Finally, the resulting Lisp program can calculate trust values and match the student's wishes and company profiles. We used Steel Bank Common Lisp (SBCL 2.1.11) in this study.

Our recommender system has presented some job openings that may not appropriately reflect the student's preferences; for example, some job offers contain work locations in foreign countries. However, the recommender system can present a sufficiently large number of job offers, and on the whole, the system gives a list of companies that match the student's preferences.

4 Discussion: How to Determine Distrust Levels

In our previous work [2][5], we determined the trust level part of a two-dimensional trust value by accumulating positive factors. This study also collects favorable aspects of a firm's job openings for the trust level part of a two-dimensional trust value; that is, we increase the trust level when keywords that match the student's expectations appear in the company profile text.

On the other hand, we can see at least two types of cases in determining the distrust level. This section discusses how to determine the distrust level.

4.1 Case 1: Distrust Level as Cumulative Degree of Negative Evaluation

The first approach is to determine the distrust level based on the cumulative degree of unfavorable evaluations, similar to the method used to determine the credibility level. The following is an example in [2].

Example 1 *In three countries, an opinion poll was conducted about the approval ratings of each country's governments. We used the following items to answer this question: "Do you trust your government?"*

1. *I have no idea;*
2. *Yes, I do;*
3. *No, I do not;*
4. *Sometimes yes, sometimes no.*

The number of answers for each item is $a_1^c, \dots, a_4^c$ *for country c; also, we have* $s^c = a_1^c + a_2^c + a_3^c + a_4^c$. *In this example, we calculate the degrees of trust* t_c *and distrust* d_c *of the government with* $t_c = \dfrac{a_2^c + a_4^c}{s^c}$ *and* $d_c = \dfrac{a_3^c + a_4^c}{s^c}$. *A survey was conducted with 100 residents each in the countries of* X, Y, *and* Z, *and the following are the results:*

$$\begin{aligned}(a_1^X, a_2^X, a_3^X, a_4^X) &= (10, 20, 30, 40),\\ (a_1^Y, a_2^Y, a_3^Y, a_4^Y) &= (50, 30, 10, 10), \textit{ and}\\ (a_1^Z, a_2^Z, a_3^Z, a_4^Z) &= (20, 25, 5, 50).\end{aligned}$$

For each country, we can give the degrees of trust t_c *and distrust* d_c *as follows:*

$$(t_X, d_X) = (0.6, 0.7), (t_Y, d_Y) = (0.4, 0.2) \textit{ and } (t_Z, d_Z) = (0.75, 0.55).$$

For each country we can also calculate the one-dimensional trust value i_c *and the degree of inconsistency* c_c*:*

$$(i_X, c_X) = (-0.1, 0.3), (i_Y, c_Y) = (0.2, -0.4) \textit{ and } (i_Z, c_Z) = (0.2, 0.3)^4 .$$

Suppose we define the two-dimensional trust values as in the approach of this example. In that case, the value of the distrust level gradually increases as the number of responses on items *3.* and *4.* in the questionnaire increases. In this sense, the distrust level is cumulative and takes various values between 0 and 1. We can see that the distrust level reaches extremities (i.e., values around 0 or 1) when either "almost all respondents answered with item *1.* or item *2.*, and the distrust level becomes around 0" or "almost all respondents answered with item *3.* or item *4.*, and the distrust level gets higher and becomes around 1." This consideration suggests that, in general, this method of determining the distrust level is unlikely to result in extreme values.

4.2 Case 2: Distrust Level in Unusual Circumstances

Another way to determine the distrust level is to rigidly set it to 0 or 1. In our previous study [5], we used such a distrust level in determining the authenticity of rescue request tweets in the event of flooding. This case concerns the determination of the distrust level in unusual circumstances.

During the torrential rainstorm in western Japan in 2018, many rescue requests were posted on SNS, and detailed address information on the posters appeared in some of those messages. Personal addresses are typically private information, and people regularly do not post them unnecessarily on SNS. Posting a private address indicates that the poster was really in imminent danger. Hence, the credibility of such a message as a rescue request becomes high.

[4] In this footnote, we provide a more detailed analysis for the pairs of (i_c, c_c). For country X, there is some degree of distrust of the government, and citizens in country X are somewhat confused since the degree of contradiction is positive. For country Y, the degree of trust exceeds the degree of distrust, but the degree of contradiction is negative, which suggests that the people have little interest in their government. For countries Y and Z, although their integration values are the same, the degree of contradiction is positive for country Z. Note that we can compare countries Y and Z, even though the conventional trust model cannot since the degree of contradiction is not addressed.

- I am the person who called for rescue after the heavy rain in Arii, Mabi-cho, Kurashiki City. I have just been rescued safely. Thank you very much for your concern. I will delete this tweet to avoid confusion. Thank you for your cooperation. #Kurashiki #Mabi-cho
- #Rescue completed #Kurashiki-shi, Mabi-cho, Arii XXX. We have received a report that they have been rescued safely. Sorry for your concern. Thank you so much.

Figure 6: Non-Rescue Requests Containing (Nearly) Complete Address Information

However, in an emergency, when determining whether a message is a genuine rescue request, it is natural that the distrust level reaches maximum if there is even one suspicious element. Note that this approach to determining the distrust level differs from the cumulative approach used in the previous section.

Figure 6 shows some postings in the above disaster. In these messages, words such as "thank you" or "sorry for your concern" appear disproportionate to a rescue request. Therefore, the distrust level is set to 1, resulting in a low one-dimensional trust value. Such phrases typically appear in messages that report the completion of the rescue, citing past actual rescue request tweets.

For the application to job search assistance shown in this paper, we also set the distrust level to 0 or 1, considering students tend to avoid firms with "undesired items" because it may affect their later life. Summarizing, we can see the following:

- For the case of judging rescue requests in the event of flooding, we should employ a high distrust level to exclude information that would interfere with rescue because human lives are at stake; and
- For the case of job search assistance, students want to use a high distrust level since they wish to exclude any job information containing undesired conditions.

From this observation, we use the distrust levels at extremities (i.e., near 0 or 1) to vehemently reject information that is not appropriate for related parties, as in the case of the two examples listed in this section. In other words, this type of distrust determination method should be used when the impact of the failure is significant for the parties concerned. A detailed examination of the validity of this approach is a future issue.

5 Conclusion

We developed a recommender system for students' job-finding assistance. After collecting information on job openings and student preferences, we calculated the two-dimensional trust values that represent the fitness of matching between the student and the company. However, the two-dimensional trust value may be inconsistent since the trust and distrust parts are given independently. This study obtained a corresponding one-dimensional trust value for a consistent evaluation; specifically, we got the one-dimensional trust value by calculating the integration value of Oda's FCR method. Then, we employed the resulting trust value in the recommendation process. In this study, we also discussed determining the distrust levels.

This paper uses a coarse criterion that we only use, either 0 or 1, to determine the distrust level, especially for emergency trusts. Introducing a finer step, as future work, may be possible. It is also necessary to provide criteria for which type of distrust we should employ in which situations, which is also an interesting issue.

REFERENCES

[1] K. Ohkubo, T. Oda, Y. Koizumi, T. Ohki, M. Nishigaki, T. Hasegawa, and Y. Kawabe, "Trust representation under confusion and ignorance," in *Proceedings of International Workshop on Informatics (IWIN 2018)*, pp. 191–198, 2018.

[2] Y. Kawabe, Y. Koizumi, T. Ohki, M. Nishigaki, T. Hasegawa, and T. Oda, "On trust confusional, trust ignorant, and trust transitions," in *Trust Management XIII - 13th IFIP WG 11.11 International Conference, IFIPTM 2019, Copenhagen, Denmark, July 17-19, 2019, Proceedings*, pp. 178–195, 2019.

[3] T. Oda, "Measurement technique for ergonomics, section 3: Psychological measurements and analyses (3) measurements and analyses by kansei evaluation," *The Japanese Journal of Ergonomics*, vol. 51, no. 5, pp. 293–303, 2015. In Japanese.

[4] J. Deng, T. Oda, and M. Umano, "Fuzzy logical operations in the two-dimensional hyper logic space concerning the fuzzy-set concurrent rating method," *Journal of Japan Association for Management Systems*, vol. 17, no. 2, pp. 33–42, 2001.

[5] Y. Kawabe, Y. Koizumi, T. Ohki, M. Nishigaki, and T. Hasegawa, "Finding trustable rescue requests with Zip numbers," in *The 37th Fuzzy System Symposium (FSS 2021)*, pp. 601–606, 2021. In Japanese.

[6] S. Marsh and M. R. Dibben, "Trust, untrust, distrust and mistrust – an exploration of the dark(er) side," in *Proceedings of the Third International Conference on Trust Management*, iTrust'05, (Berlin, Heidelberg), pp. 17–33, Springer-Verlag, 2005.

[7] G. Primiero, "A calculus for distrust and mistrust," in *Trust Management X* (S. M. Habib, J. Vassileva, S. Mauw, and M. Mühlhäuser, eds.), (Cham), pp. 183–190, Springer International Publishing, 2016.

[8] G. Primiero, F. Raimondi, M. Bottone, and J. Tagliabue, "Trust and distrust in contradictory information transmission," *Applied Network Science*, vol. 2, p. 12, 2017.

[9] R. J. Lewicki, D. J. B. McAllister, and R. J. Bies, "Trust and distrust: New relationships and realities," *Academy of Management Review*, vol. 23, pp. 438–458, 1998.

[10] "MeCab: Yet another part-of-speech and morphological analyzer." `https://taku910.github.io/mecab/`.

Session 4:
Sensing and Analysis
(Chair: Takuya Yoshihiro)

Analysis and Sharing of Cooking Actions Using Wearable Sensors

Ayato Kumazawa†, Fuma Kato‡, Katsuhiko Kaji‡, Nobuhide Takashima*,
Katsuhiro Naito‡, Naoya Chujo‡, and Tadanori Mizuno‡

†Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
‡Department of Information Science, Aichi Institute of Technology
* Mitsubishi Electric Engineering Co., Ltd.
{b22706bb, k20035kk, kaji}@aitech.ac.jp

Abstract - There are many different ways to share cooking recipes and pervasive in the world. Now that the Internet has become widespread and information sharing has become easier, recipes can be browsing on the Internet. Mechanisms already exist for sharing dish quality and taste through photos and text. On the other hand, there is no mechanism to compare cooking ability and efficiency with past selves and others. This research aims to increase the enjoyment of cooking by sharing and comparing recipes and cooking actions. The approach is to use wearable sensors to sense the cooking process and extract features for comparison. The contribution of this research is twofold. The first is a proposal for a framework for a system that allows compare cooking ability and efficiency with past selves and others. The second is the actual sensing of cooking behavior and providing concrete examples of analysis and comparison with respect to transitions in behavior and place.

Keywords: Wearable Sensing, Analysis Cooking Actions, Location Estimation, Behavior Modification

1 Introduction

There are many different ways to share cooking recipes and pervasive in the world. Now that the Internet has become widespread and information sharing has become easier, recipes can be browsing on the Internet. There are several ways to share recipes, such as posting them on your own blog or using a recipe sharing service. Recipe sharing services include Cookpad[1] and Kurashiru[2]. There is also a function to report what you have created using a recipe. The Tsukurepo in Cookpad and the Taberepo in Kurashiru fall into this category.

Mechanisms already exist for sharing dish quality and taste through photos and text. On the other hand, there is no mechanism to compare cooking ability and efficiency with past selves and others. We thought that such a system would help could motivation in cooking by realize the growth of their own abilities or compare own cooking skills with others. If multiple cooking data are collected, it is possible to compared how different your cooking process compared to the recipe author or how much better you are than you were in the past. For example, when comparing yourself now and past, you can see that your cooking has improved, and it is thought that motivation to more become better at cooking will increase. Also, if you compare yourself with a professional, you may be able to find areas for improvement and get closer to becoming a better cooking.

This research aims to increase the enjoyment of cooking by sharing and comparing recipes and cooking actions. The approach is to use wearable sensors to sense the cooking process and extract features for comparison.

The contribution of this research is twofold. The first is a proposal for a framework for a system that allows compare cooking ability and efficiency with past selves and others. The second is the actual sensing of cooking behavior and providing concrete examples of analysis and comparison with respect to transitions in behavior and place.

2 Related work

This chapter describes related research, divided into behavior estimation, cooking, and location estimation methods. Each category describes its relationship to this research.

There are different types of research to behavior estimation. For example, there are estimations of walking behavior, medical behavior, and cooking tasks. And various sensors are used in action estimation. For example, some research has used acceleration and angular velocity to estimate behavior [1]. Research that also uses other sensors includes research that adds BLE to estimate activities in the home [2], and research that adds biometric and environmental sensors to estimate behavior [3]. This research utilizes acceleration and angular velocity, which are available in most wearable sensors, to estimate the cooking process and analyze the motion.

Various methods of supporting cooking have been researched for cooking recipes and videos. For example, there's research on estimating cooking behavior from video [4]. Other, attaching sensors to cutting boards and knives to provide feedback on the cutting process [5], and organizing the process when preparing multiple cooking items at the same time to improve cooking efficiency [6]. These research studies provide support for the cooking itself. This research aims to share and compare data during the cooking process, and after cooking, to show the growth and areas for improvement, and to increase the motivation for cooking.

There are various methods for location estimation. One of these methods is PDR, which uses sensors such as acceleration and angular velocity. There is research on location estimation by PDR with smartphones and smartwatches [7]. In this research, we also wanted to use PDR to estimate the location of wearable sensors. However, in PDR, errors ac-

[1] https://cookpad.com/
[2] https://www.kurashiru.com/

cumulate due to sensor noise, which reduces the accuracy of position estimation. There are many methods to compensate for this error. Examples include FootSLAM [8] and ActionSLAM [9], which adapt the SLAM algorithm. FootSLAM is a method that corrects for errors by taking advantage of places where straight lines are walked, such as hallways. ActionSLAM is a method that corrects for errors by using sitting, standing locations. This research utilizes the error correction method of ActionSLAM to estimate location based on the relationship between the cooking process and location.

3 Analysis and Sharing of Cooking Actions Using Wearable Sensors

The Figure 1 shows the overall picture of this research. In each section, the overall picture is divided into four items: Sensing of cooking behavior, Sharing of sensing data, Analysis of cooking ability and efficiency, and Visualize history and comparisons with others.

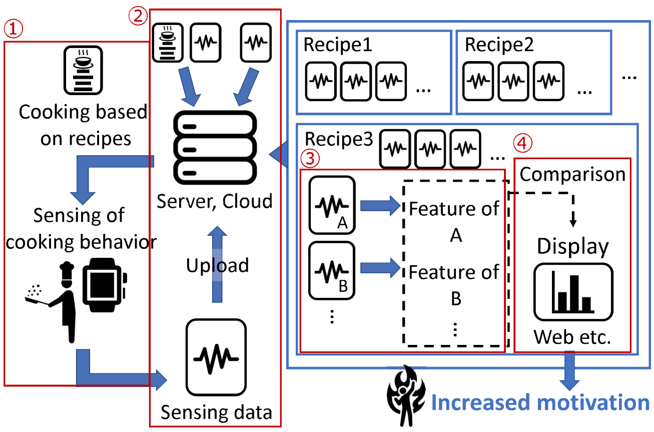


Figure 1: Overall picture of research

3.1 Sensing of cooking behavior

In part 1 of the Figure 1, wearable sensors are used to convert the cooking motion into data. This research a smartphone and a smartwatch are used as wearable sensors to collect position data with the smartphone and arm movements during cooking with the smartwatch.

One reason for the need for location estimation is the relationship between the cooking process and place. For example, the cutting process takes place where the cutting board is located, and the baking process takes place where the stove is located. In other words, it is necessary to use the relationship between the cooking process and the place to estimate the cooking process according to the location.

This system is intended for the general public who are cooking or just starting out. Currently, smartwatches are not widely used, and it is required to use this system need to be provisioned. If the system becomes widely used in the future, it will be possible to use it without additional devices.

3.2 Sharing of sensing data

In part 2 of the Figure 1, the data obtained by sensing the cooking operation is shared with the cloud and servers. The recipe creator shares the data of the actual cooking process together with the recipe when sharing a recipe. The recipe using users share their work data when post the results of cooking according to the recipe.

The creator cooks the recipe according to the recipe with sensing after the recipe is completed. And when posting a recipe, the sensing data post together. The using users perform sensing when cooking with using posted recipes. And when reporting the results of the creation, the sensing data is posted.

3.3 Analysis of cooking ability and efficiency

In part 3 of the Figure 1, the shared sensing data is analyzed and features are extracted. Features include the number of times the ingredients are cut and the number of times the refrigerator is opened.

The analysis contents includes cooking process estimation and feature extraction. Process estimation uses arm motion and position data during the cooking operation to estimate the current process. Feature extraction analyzes the cooking action data and extracts features. For example, the number of times and pace of cutting in the cutting process and the number and speed of pan shaking in the baking process are extracted. And place transfers, the number of times the refrigerator was opened and the time spent washing dishes are extracted.

3.4 Visualize history and comparisons with others

In part 4 of the Figure 1, the features are compared and the history and the skill or unfamiliarity of the cook with others, as well as areas of growth and improvement, are displayed. Users can see how well they are doing and what they need to improve from an overall perspective from the comparison results, which increases the motivation needed to continue cooking in the future. Comparisons can be made with professionals, people at the same level, friends, or past selves. In comparison with professionals, we can see how much difference there is between us and them, in comparison with others at the same level, we can see what they do better than others. And in comparison with friends, we can see comparisons that are only possible between friends. The comparison of cooking operations is explanation into comparison of actions at fixed points and comparison of places.

Comparison of the operation at fixed points is described using the cutting process as an example. Features include the number of cuts, the degree of force used to cut, and the pace of cutting. If the number of cuts is greater than that of others, it is likely that they are finer cuts, conversely, if it is less, it is likely that they are coarser cuts. If the force used in cutting is stronger than that of others, it is possible that the user is using force unnecessarily. If the pace of cutting is faster than others, the cook is considered to be a good cook, if slower to be inexperienced. In this way, multiple features can be compared to indicate skillful or inexperienced. The differences in skillful or inexperienced also reveal areas for growth and improvement.

Next, an example of comparison of places. The number of visits to the refrigerator and the timing of these visits would indicate differences in the unnecessary opening and closing of the refrigerator and the return trips to and from the refrigerator. If we look at the time spent at the cutting board as preparation time for baked cooking, we can see the difference in how much time it took. Thus, the transitions between places and the time spent in represented to show point for skillful or inexperienced.

4 Analysis and comparison of actions at fixed points

The analysis and comparison of actions at fixed points was conducted using the cutting process as an example to see what characteristics could be observed.

4.1 Survey contents

A sensing series of actions, including the cutting process, were performed to collect the data to be analyzed. The sensing participants consisted of three students who do not usually cook (Participant A-C), a student who usually cooks (Participant D), and a people who cook every day (Participant E).

As a sensing method, a smartwatch (TicWatch E3) was attached to the arm of the person holding the knife, and acceleration including gravity was collected using a home-made application. Participants were asked to slice a cucumber into round slices. The thickness of the cut is not specified. The cooking process begins with rinsing the cucumbers, cutting them into round slices, and transferring them to a plate. The cooking was also performed while taking video for later analysis.

4.2 Analysis of data

An example of the resulting acceleration norm is shown in Figure 2. This data is after norm, a moving average filter was used to remove sensor noise. Labels for the cooking process are guessed from the video taken and input manually. Synchronization with the video is done by having the user first swing his or her arms, taking advantage of the large change in acceleration that occurs when the arms are swung. Let the section in which the cucumbers are washed be Wash, the section in which they are sliced into rounds be Cut, and the section in which they are transferred to a plate be Plating of dish. Features are extracted from this data. This time, the number of cuts, cutting pace, and acceleration during cutting are calculated in the Cut.

The first step is to estimate the number of cuts. The video and data show a significant change in acceleration at the timing of the cut. Therefore, this time, peak detection is used to estimate the number of times it was cut. The threshold is set at $13.5m/s^2$, and the portion of the maximum value between when the threshold is exceeded and when it is below

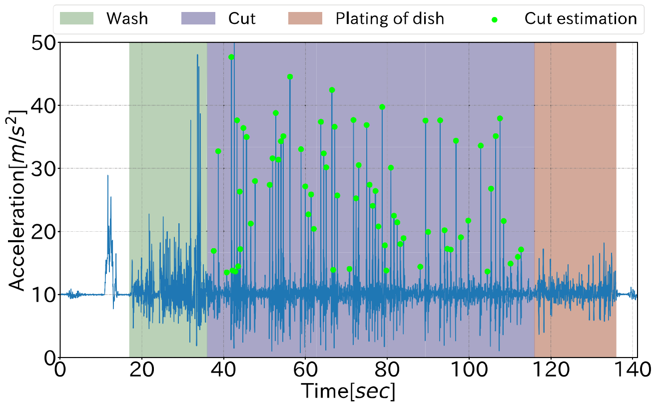


Figure 2: Sensor data obtained (Participant A)

the threshold is estimated to have been cut once. And, no new estimation is performed for 0.2 seconds after the threshold value is exceeded. The result is displayed as a Cut estimation in Figure 2. The acceleration in the area estimated to have been cut from the results is collected, and the average and maximum values are extracted as features.

Next, the average of the cutting pace is calculated. The cutting pace uses the number of seconds between each cut. The interval longer than 3 seconds was considered to be an interruption in the temporary cutting motion, such as picking up a cucumber attached to a knife, and was not used in the calculation of the average.

The 20 second data of the cut section of all participants after the above processing is shown in Figure 3, and the features in all the extracted data are shown in Table 1. Figure 3 shows 20 seconds of data in order to make the pace of cutting visually clear. The actual and estimated results for the number of cuts, the average pace of cutting, the average and standard deviation of the acceleration at the time of cutting, and the average of each are placed in Table 1. The error in the estimation of the number of cuts varied from participant to participant. Therefore, the estimation threshold needs to be improved in the future.

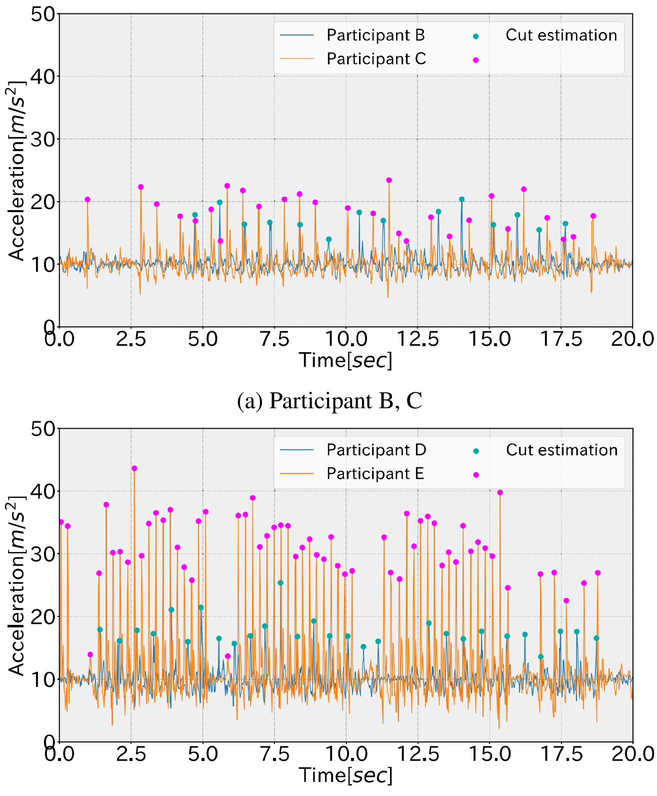
4.3 Comparison of data

The features obtained from the analysis were compared to determine the differences among participants. The graphs of average pace and acceleration are shown in Figure 4 and Figure 5, respectively.

First, a comparison is made between participants A-C who do not usually cook. The number of times the participant B made a cut was less than that of the others. In fact, participant B's cucumbers were shorter than the others. If the thickness after the cut is approximately the same based on the number of cuts, the length of the cut object can be estimated, and if the length of the cut object is the same, the average thickness when cut can be estimated. The acceleration of participant A is larger than that of the others. The mean is seen as the average speed of slashing down, and the standard deviation as the stability of the slashing down speed. From this, it can be inferred that participant A has more cutting power than the others. The standard deviation for participant A is larger than

Table 1: Features by participant

Participant	Actual of times $[Times]$	Estimation results $[Times]$	Mean pace $[sec]$	Mean acceleration $[m/s^2]$	Standard deviation of acceleration
A	60	70	1.18	26.8	9.45
B	43	42	1.01	17.3	2.14
C	76	87	1.23	17.2	3.02
D	106	106	1.21	18.1	2.87
E	159	164	0.52	32.6	5.57
Overall mean	88.8	93.8	1.03	22.4	4.61



(a) Participant B, C

(b) Participant D, E

Figure 3: Estimation results for 20 seconds of the Cut portion of the participant

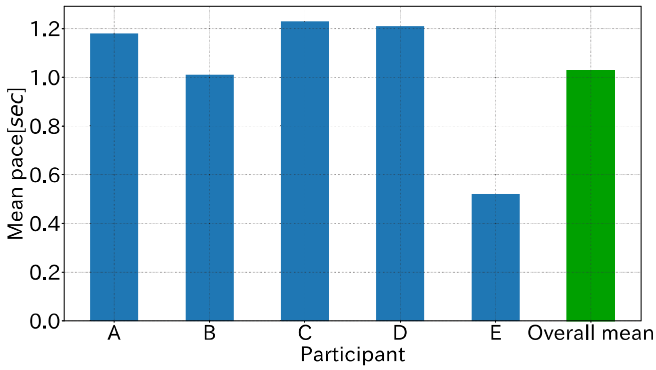


Figure 4: Graph of mean pace

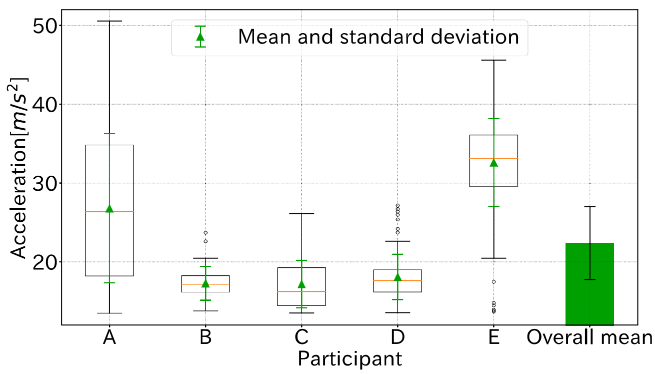


Figure 5: Graph of acceleration

that of the others. From this, it can be inferred that participant A is not cutting with a steady force.

Next, a comparison is made between participants A-C, who do not usually cook, and participants D and E, who do cook. The number of times the respondents who cook on a regular basis cut more often than those who do not cook. In fact, the thickness of the cucumbers cut by the person who usually cooks them was almost uniform. In other words, if the length of the object to be cut is the same, a good cook will cut more frequently. A large difference was observed between participant E and the other participants in terms of the pace at which they cut and the acceleration at which they cut. The pace of the participant E was more than twice as fast as that of the others participants. And, the acceleration was larger than that of participants B to D. This is thought to be because the pace of cutting is faster and therefore the speed of cutting down is also faster.

The comparison was made among a small number of people, but in the future it will be necessary to make comparisons among a very large number of people. In such a case, a method to indicate where you are in relation to the average of all participants is possible. For example, for the present data, the standard deviation of the acceleration of participant A is larger than the overall average, and the cutting pace of participant E is about twice as fast as the overall average.

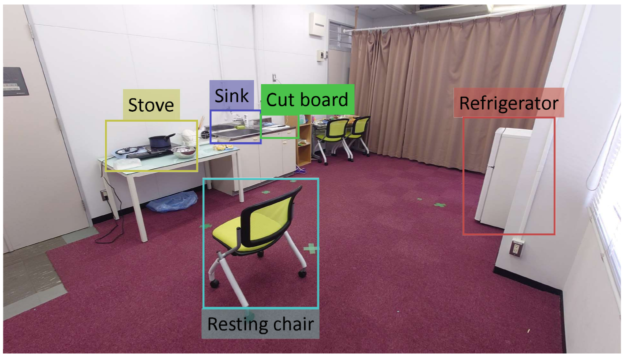
In this way, features such as the number of times and the degree of force can be obtained just by cutting, and the results of feature comparison can be expected to increase the motivation of the participants. For example, the comparison results show that participant A is less stable than the others, so he tries to do better because he is less stable than the others. The fact that participant E cuts faster than the others gives us confidence about the pace at which he cuts. The baking and mixing process are also expected to increase motivation, as are the cutting process.

5 Analysis and comparison of transfers between fixed points

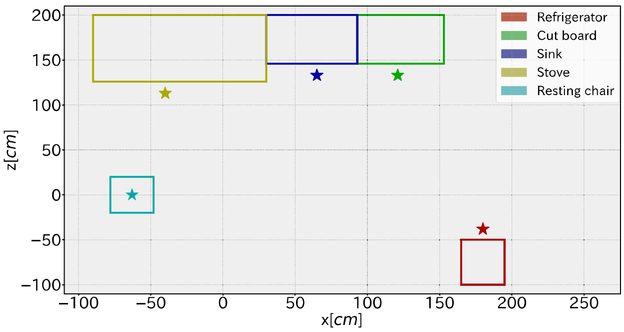
Next, the analysis and comparison of transfers between fixed points was conducted using the hanburg steak cooking process as an example to determine what characteristics were actually observed.

5.1 Survey contents

Sensing of hanburg steak cooking was performed to collect data to be analyzed. Four participants were students, of whom data were collected twice for participants C and D. In ascending order of participant number, the its are those who usually cook. The cooking environment is arranged as shown in Figure 6. Figure 6b is the result of surveying the actual stop position of the cooking environment, which will be explained later.



(a) Actual cooking environment



(b) Standing of on the graph

Figure 6: Cooking Environment

The following section describes the cooking utensils and ingredients used and the cooking process in the cooking environment. Cooking utensils include a cutting board, knife, bowls, bowls for beating eggs, chopsticks, frying pan and lid, and a fry pan. Materials include ground meat, onions, eggs, milk, breadcrumbs, and seasonings. Leave as meat, onions, eggs, and milk in the refrigerator, and the breadcrumbs and seasonings under a cutting board. In the hanburg steak cooking process, chop onion, add minced meat, beaten egg, milk, bread crumbs, and seasonings. Then, remove air from the pan, shape it and bake it over medium heat for about 10 minutes. The materials are prepared in such a way that there is a surplus, and the excess materials are returned to their original place. It is also a condition to wash the cooking utensils used in the cooking process, except for the pan, lid and spatula.

The hanburg steak cooking process involves a waiting period during the cooking process called baking time, in the meantime, differences can be seen between participants, such as washing the utensils used. Even in the waiting time, differences between skilled and inexperienced operators can be seen, and can be used for comparison. For this data collection, a refrigerator, a cutting board, a sink, a stove, and a resting chair were prepared as places. The process includes cutting, mixing, and baking, and requires at least a cutting board and a stove as a place. A refrigerator for storing materials and a sink for washing dishes are also needed. A resting chair is also provided as a place to wait during baking time. The stopping position of fixed points (the star marks position in Figure 6b) was determined in advance, and the participants were asked to work there.

As a sensing method, a mocopi[3] (Figure 7) was attached and the movement history was collected. mocopi is a device that can motion capture by attaching inertial sensors to six locations on the head, waist, arms, and legs. The main purpose of this research is to share and compare the cooking behavior, so we will use mocopi, which is currently easy to use and can be expected to provide a certain degree of accuracy in position estimation. In the future, movement histories will be collected using wearable sensors that can be more easily prepared. The movement history is obtained from the motion capture results and used for analysis.

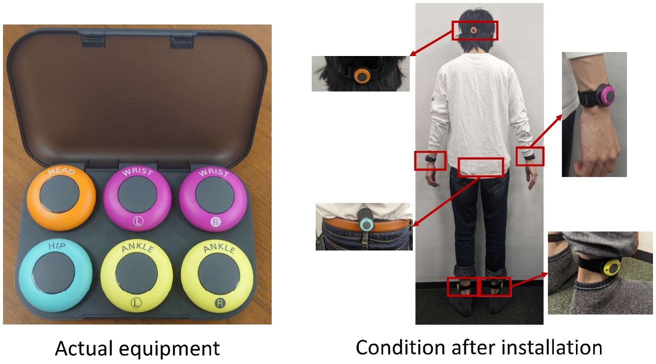


Figure 7: Details of mocopi

5.2 Movement estimation using wearable sensors

Movement estimation was performed using wearable sensors, and the rough accuracy was investigated. Mocopi was used as a wearable sensor. Estimation is performed using the position of the waist in the motion capture results. A portion of the actual trajectory obtained is shown in Figure 8. The upper triangle represents the start point and the lower triangle represents the end point. From this point, the accumulated error due to sensor noise is corrected.

Use a place to sit or stand as a method of compensation. In this case, a place to sit or stand is prepared in advance, and

[3]https://www.sony.jp/mocopi/

the refrigerator / resting chair place is used to compensate for this. These two places are distinguished from other places by the use of the position of the hips, because they involve a slouching motion. A threshold value is set for the waist position and the current location, and correction is applied so that the robot will reach the set stopping position after a certain period of time has elapsed after the threshold value is exceeded. The difference between the current location that exceeds the threshold and the set stop position is determined, and a correction value is calculated and reflected in subsequent data as appropriate. If a new threshold value is exceeded, a new correction value is obtained and updated. In this case, if the position of the waist is less than 80 cm, and based on the placement of the stopping position and the time spent in that place, the current location is $x < 25$ and the user has been there for at least 5 seconds, then the user is in a rest chair, and if $25 < x$ and $z < 50$ and the user has been there for at least 1 second, then the user is in a refrigerator.

The actual movement trajectory results with correction are shown in Figure 9 and the correct data for place estimation over time and after correction are shown in Figure 9. The location estimation with the corrected data is based on the nearest stop point from the current location as the place where it. In the case of participant A, after the correction, the about roughly estimated. However, in the case of the others, the results are not as good as those of participant A. Taking the first data of first time for participant D (Figure 11) as an example, there is a weak to estimate places that are close in distance, such as the sink and the cutting board. To improve the accuracy of the estimation, measures such as having them sit down periodically to increase the frequency of correction and the number of correction points can be considered.

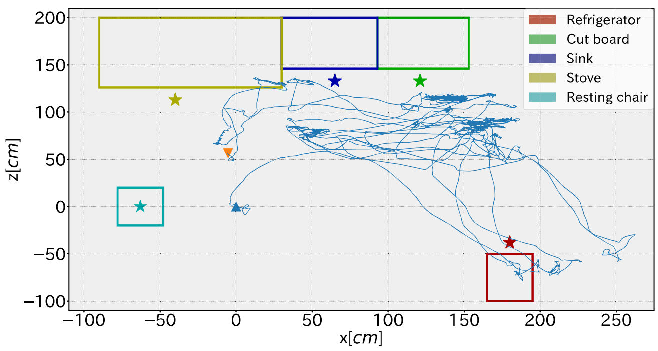


Figure 8: Example of obtained movement trajectory data (Participant A)

5.3 Analysis and comparison of data

The data were analyzed and compared using the correct answer data to determine what differences existed between participants. Originally, we would like to analyze and compare the data from the estimate data. However, as described in subsection 5.2, the estimation accuracy is not high, so we will use the correct data that we have prepared to check the estimation accuracy, which is accurate data. The correct answer data for each participant is shown in Figure 12.

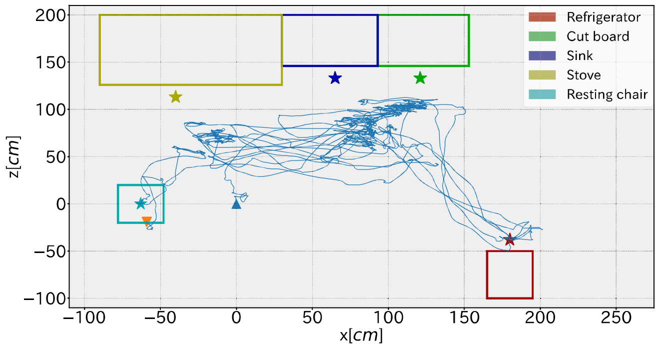


Figure 9: Movement trajectory correction result (Participant A)

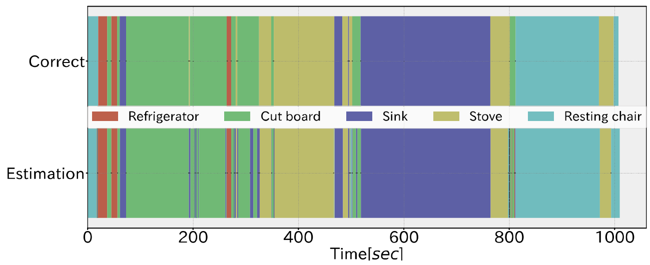
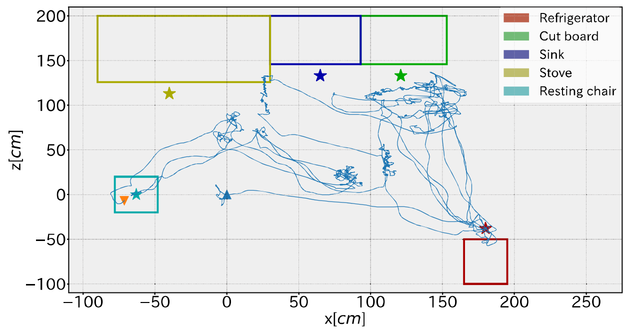


Figure 10: Location estimation result (Participant A)

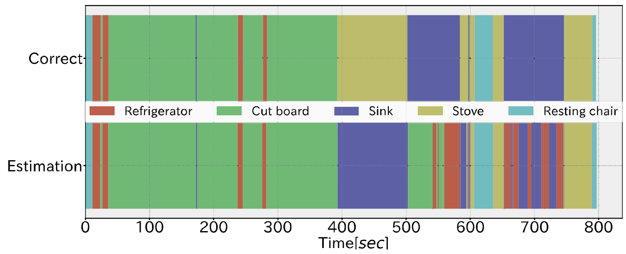
First, the time spent in fixed points is analyzed and compared. The time spent in fixed points by each participant and the overall mean and standard deviation are shown in Figure 13. No comparison was made for the time spent in the refrigerator and the break chair, since the cumulative time is not considered to indicate skillful or inexperienced. Compare the time spent at the sink. The majority of the time spent at the sink is spent washing dishes, so the comparison is based on how long it takes to finish washing dishes. Looking at the Figure 13, it can be seen that second time for participant D finished the fastest and participant B finished the slowest. From here, it is possible to analyze whether the washing dishes process is faster or slower than the overall average when compared to the overall average.

Users who see the results are happy if the overall washing dishes speed is fast, and think about efficient washing dishes methods if it is slow. As for the cutting board and stove, it is difficult to compare them from the time spent because each is part of the baking preparation time and baking time. Therefore, the baking preparation time and baking time are analyzed and compared in terms of cooking time.

Next, the cooking times are analyzed and compared. The cooking time by participant and the overall mean and standard deviation are shown in Figure 14. The total cooking time is the time from the start of data collection to the end, the baking time is the time from the time the stove was on for more than 30 seconds to the end, and the preparation time for subtract baking time from total cooking time. The total cooking time will not be analyzed because the movements performed in between are important. Look at the baking time, recommended time is 10 minutes, while the baking time for participant D is shorter for both times, and the second time is out of the standard deviation of the overall mean. In fact, the second



(a) Movement trajectory correction result



(b) Location estimation result

Figure 11: Examples of the many misjudgments that have occurred (Participant D: 1st)

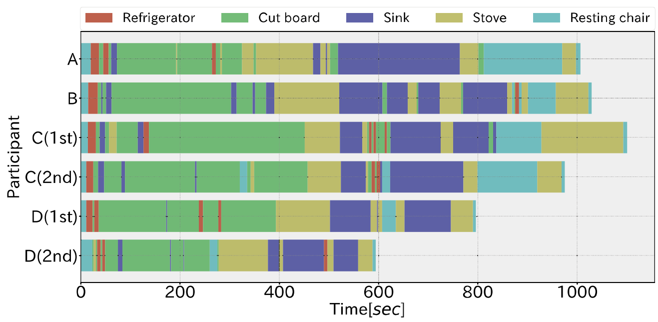


Figure 12: Place of stay by time for each participant

hanburg steak was not cooked all the way through. From this point on, if the behavior deviates from the standard deviation of the overall average, it can be analyzed as abnormal behavior, such as undercooked or overcooked.

The user who sees this result will think that if the baking time is not long enough from an overall perspective, he will bake a little longer, and if it is overcooked, he will finish baking a little faster. The preparation time for baking was shorter for second time for participant D and longer for second time for participant C than for the others. From here, it is possible to analyze whether the baking preparation is faster or slower than the overall average. The user who sees this result is happy if the preparation is fast from an overall perspective, and thinks about how to prepare efficiently if it is slow. And when compared to their past selves, they both have reduced their cooking time .

In addition, transitions in place are analyzed and compared. The location transition was calculated as the number of times a person moved from the current place to the next. A to B and B to A are seek as separate number of times. The method is to assume that a person moves from the current place to the next place when it has been in the current place for more than 1 second. An example of the actual data obtained is shown in Figure 15. As an example, we will analyze and compare the number of visits to the refrigerator. The number of visits to the refrigerator by each participant and the overall mean and standard deviation are shown in Figure 15. It should come to the refrigerator at least twice for loading and unloading the food. However, a large number of foodstuffs may mean that the foodstuffs are taken in and out as needed and used foodstuffs are immediately put away, or that the foodstuffs are divided into multiple portions because they cannot all be carried. Therefore, judgments of skillful or inexperienced will vary from user to user and will not be analyzed. From this, the user can see how much more or less open the door is from the overall viewpoint, and if it is less, the user can see that there is no waste in movement, and if it is more, the user tries to reduce the amount of movement.

Thus, comparisons of time spent in place, transition, and cooking time revealed skillful or inexperienced points in comparison with others, while individual comparisons showed growth. From this point, users can confirm areas for improvement and realize their growth, which is expected to increase their motivation for cooking in the future.

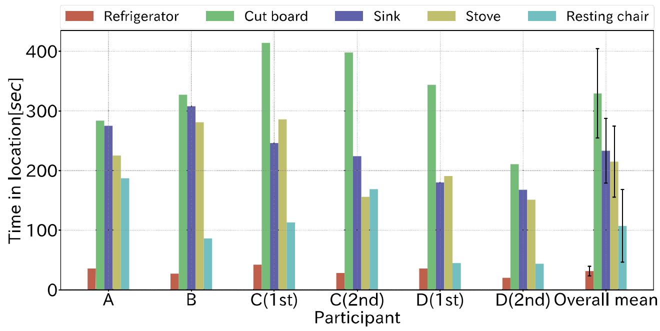


Figure 13: Cumulative time spent by location

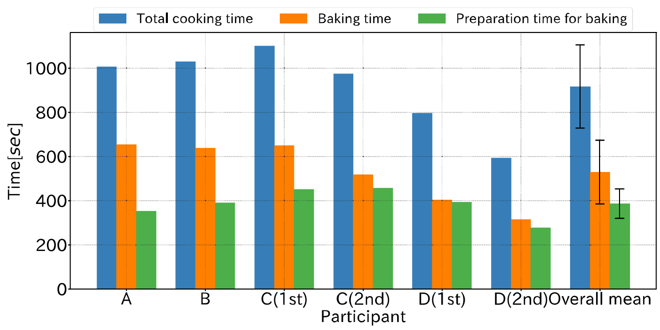


Figure 14: Cooking time

6 Conclusion

This research aims to increase the enjoyment of cooking by sharing and comparing recipes and cooking tasks. The approach is to use wearable sensors to sense the cooking process and extract features for comparison. The skill of a cook is expressed not by taste or appearance but by the actions taken

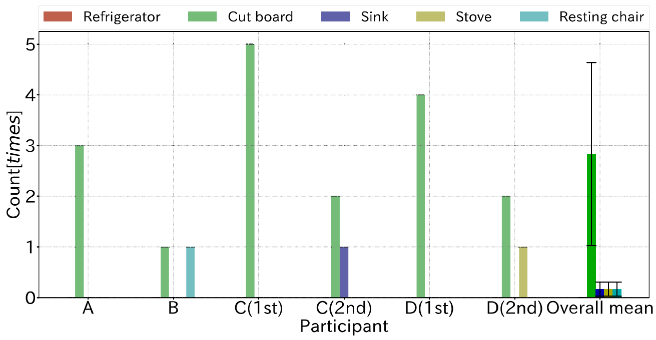


Figure 15: Number of times each moved from the refrigerator

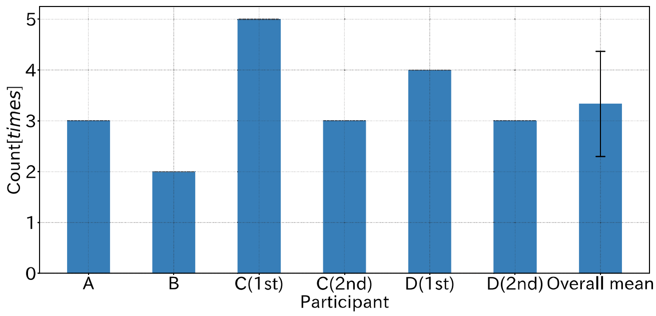


Figure 16: Number of moved to the refrigerator

during the cooking process. Analysis and comparison of the cooking behavior data collected as a contribution to this research shows an increase in user motivation.

The overall picture of this research can be divided into four categories: sensing of cooking behavior, sharing of sensing data, analysis of cooking ability and efficiency, and visualize history and comparisons with others. Wearable sensors are used to collect data on the cooking process and analyze it to extract features. The extracted features are then shared, compared, and the results are displayed.

The analysis and comparison at fixed points was conducted using the cutting process as an example. The various features and comparisons obtained as a result of this process will reveal the handiness and unfamiliarity of the process, which will be replicated in other processes. In the analysis and comparison of transfers between fixed points, hanburg steaker cooking was used as an example for analysis and comparison. Movement estimation is still in a state of inaccuracy. Therefore, the analysis and comparison from the correct answer data revealed some dexterity and unfamiliarity.

Future work is needed to analyze and compare the results at fixed points, in addition to the cutting motion. For example, there are various processes such as baking, mixing, etc. Each process is investigated. For the extraction of features, we will examine what kind of features can and should be extracted from each process. For example, in the case of the baking process of an omelet, the speed of rolling, the number of times and others may be considered. For comparisons, the extracted features are used to determine what kind of differences in cooking can be seen. For example, in the case of the baking process of an egg omelet, there is a difference in the speed of rolling.

The analysis and comparison of transfers between fixed points needs to be validated with more accurate estimation methods. In this study, the analysis and comparison were conducted using the correct answers, but in the future, the analysis and comparison will be conducted using the estimate results. I could also analyze the unusual behavior during the survey and do it with recipes that are not good to change the order. In addition, a complementary mechanism for location estimation and action recognition is realized with reference to ActionSLAM2. Specifically, increased accuracy of behavior recognition by narrowing down that a person would perform if it were in this place, and make a correction for the location that if you are doing this action, you would be in this place.

A system such as a research overview needs to be created, shared, and displayed. For sharing, it is necessary to consider what kind of data structure is to be shared. The content of the labeling must be expected to increase the motivation for cooking. Currently, we are thinking of creating a system that wraps around the recipe site. For example, a bar graph could be used to show your position in relation to the total number of cuts. Other possibilities include using a radar chart to represent one's features. It is necessary to verify whether such a display method actually increases motivation and to consider other methods.

REFERENCES

[1] Jiang, Z.: Human Action Recognition Method Based on Wearable Inertial Sensor, Proceedings of the 2021 6th International Conference on Cloud Computing and Internet of Things, CCIOT '21, p. 73–76 (2021).

[2] De, D., Bharti, P., Das, S. K. and Chellappan, S.: Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare, IEEE Internet Computing, Vol. 19, No. 5, pp. 26–35 (2015).

[3] Demrozi, F., Pravadelli, G., Bihorac, A. and Rashidi, P.: Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey, IEEE Access, Vol. 8, pp. 210816–210836 (2020).

[4] Okumura, T., Urabe, S., Inoue, K. and Yoshioka, M.: Cooking Activities Recognition in Egocentric Videos Using Hand Shape Feature with Openpose, in Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, CEA/MADiMa '18, p. 42–45 (2018).

[5] Masuda, H., Kato, K., Ikematsu, K., Takegawa, Y. and Hirata, K.: Tablet Cutting Board: Tablet-Based Knife-Control Support System for Cookery Beginners, in Proceedings of the Augmented Humans International Conference 2023, AHs '23, p. 287–293 (2023).

[6] Nakabe, J., Mizumoto, T., Suwa, H. and Yasumoto, K.: Optimal cooking procedure presentation system for multiple recipes and investigating its effect, Algorithms, Vol. 14, No. 2, p. 67 (2021).

[7] Wakaizumi, T. and Togawa, N.: A PDR Method Combining Smartphone and Smartwatch based on Multi-Scenario Map Matching, 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), pp. 308–309 (2021).

[8] Angermann, M. and Robertson, P.: FootSLAM: Pedestrian simultaneous localization and mapping without ex-

teroceptive sensors—Hitchhiking on human perception and cognition, Proceedings of the IEEE, Vol. 100, No. Special Centennial Issue, pp. 1840–1848 (2012).

[9] Hardegger, M., Tröster, G. and Roggen, D.: Improved actionSLAM for long-term indoor tracking with wearable motion sensors, Proceedings of the 2013 International Symposium on Wearable Computers, pp. 1–8 (2013).

Design and Implementation of a Method for Estimating Bicycle Air Pressure Decrease based on Vibration Sensing of Bicycles using Smartphone

Rui Yamaguchi†and Katsuhiko Kaji†

†Graduate School of Business Administration and Computer Science, Aichi Institute of Technology

{b23724bb, kaji}@aitech.ac.jp

Abstract - Some people experience accidents or near misses while riding because they did not conduct bicycle inspections before riding. There is a need to promote riders to conduct bicycle inspections to reduce the number of people who experience this. In this study, we focus on air pressure inspection and propose a method for estimating bicycle air pressure decrease based on vibration sensing using smartphone. The acceleration in the direction perpendicular to the ground is acquired with a smartphone, and the standard deviation and the amplitude spectrum of each frequency are used as features to estimate the air pressure decrease using random forests and other methods.

Keywords: smartphone, sensor signal processing, bicycle, air pressure

1 Introduction

There is a need to promote bicycle inspections because only a few people who ride a bicycle do so every time. According to a survey conducted by au Insurance, 86.9% of bicycle riders do not conduct any or few necessary inspections before riding(whether the brakes work, whether the tires are inflated sufficiently, the position, color, and angle of reflective materials, whether they are dirty, whether the handlebars and saddle are not wobbly, whether the chain is loose or rusty, whether the bell and buzzer ring, whether the lights come on)[1]. There is a need to promote riders to conduct bicycle inspections before riding to reduce the number of malfunctions and accidents or near misses experience.

In this study, we focus on tire air pressure inspection among the inspections that should be conducted before riding, and the purpose of this study is to realize air pressure inspections without introduction / operation costs and time consuming. Existing inspection methods need to use exclusive devices such as air gauges to confirm air pressure. Even with an exclusive device, there is a need to remove the cap from the valve of the tire in order to measure it. This method is time consuming.

In this study, a method for estimating air pressure decrease using smartphone is proposed. There is need for a sensor to estimate air pressure decrease. Smartphones are equipped with many sensors. There is no need to purchase new sensors because smartphones are widely used nowadays. In addition, they can quickly confirm the results of air pressure estimation because they are often carrying their smartphone with them at all times. For these reasons, we considered that air pressure inspections could be promoted using smartphone without introduction / operation costs and time consuming.

The flow of promoting air pressure inspection using smartphone is shown in the Figure 1. The rider inflates the tire to the maximum proper pressure indicated on the side of the tire. Next, the rider rides bicycle while in possession of a smartphone. The smartphone collects feature values that occur at the best timing for tire inflation using equipped sensors. The smartphone estimates whether the air pressure is the best timing for inflation. The best timing for inflation is when the air pressure is about to decrease below the minimum proper pressure. The minimum proper pressure is the minimum pressure that can make the tire perform effectively. The smartphone promotes the rider to inflate when it estimates the best timing for inflation. In this flow, There is no need to confirm the air pressure frequently or to use exclusive devices to measure it. Therefore,we consider that air pressure inspection can be conducted without installation/operation costs and time consuming.

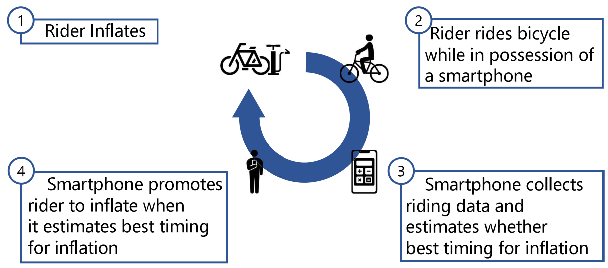


Figure 1: The flow of promoting air pressure inspection using a smartphone

2 Related work

There are products that help to confirm air pressure such as air pressure sensors from Kashimura[2] and Tyrewiz from Quarq[3]. The air pressure information is wirelessly transmitted to a monitor in the automobile. The air pressure is displayed on a monitor in real time.

Tyrewiz from Quarq[3] attaches small barometric sensors to two bicycle tires. The exclusive application is installed on a smartphone. The air pressure information is wirelessly transmitted to a exclusive application. The air pressure is displayed on a exclusive application in real time. The rider can always confirm the air pressure. Therefore, there is no time consuming to confirm the air pressure. However, exclusive sensors are expensive and there is a need to manage batteries. Therefore, installation / operation costs are high. In this study, Smartphone is used as a sensor. There is no introduction and operation cost with this method.

There is a product that helps to inflate such as Smart Air Inflator from KuKiire[4]. When the inflator is attached to the tire, it automatically inflates the tire. The air pressure can be automatically inflated to the preferred pressure at the time of inflation. Therefore, there is no time consuming process to conduct air inflation. However, there are dangerous riding conditions caused by the judgement of the rider. In this study, the smart phone estimates the best timing for inflation. Dangerous riding conditions are prevented for this method.

There is product that does not require air pressure such as Air Free Concept from Brigestone[5]. A tire that does not require air is called a punctureless tire. When regular tires are replaced with punctureless tires, there is no need for air pressure inspection without time consuming. However, punctureless tires cost more than regular tires. In addition, the ride quality is different between regular and punctureless tires, and the rider feels uncomfortable. In this study, bicycles used in everyday life will be used. The ride quality is not affected for this method.

There are studies on air pressure such as TPMS. TPMS is a air pressure monitoring system using a barometric sensor. TPMS are broadly classified into direct and indirect methods.

There is a direct method study that proposes a technology to improve the performance of barometric sensors[6]. A barometric sensor is directly attached to the tire and measures the air pressure. Therefore, air pressure can be measured with high accuracy. However, barometric sensors are expensive to produce.

There is a indirect method study that uses a single barometric sensor to estimate the air pressure of four tires[7]. One barometric sensor measures the air pressure of one tire. The air pressure of one tire is used to estimate the air pressure of the other three tires. If the air pressure of four tires is to be measured directly, there are need to use four barometric sensors. However, only one barometric sensor is used to estimate air pressure. Therefore, air pressure estimation can be conducted at low cost. However, a dedicated sensor is used even if only a little. Therefore, it is costly to a certain extent.

In a previous study, we examined air pressure decrease estimation from bicycle riding speed[8]. The force required to pedal decreases as air pressure decreases. Therefore, we considered that the riding speed would decrease as well. An experiment was conducted to estimate air pressure decrease using riding speed. From the results, features that can be estimated for air pressure decrease did not occur. As the reason for this, features that can be changed by the rider's discretion were used. For example, even hen the pedals felt heavy, the rider adjusted the force and drove at the same speed as when there was enough air pressure. Therefore, there is a need to use features that occur with air pressure decreases that the rider cannot change.

3 Method for estimating air pressure decrease based on vibration sensing

This chapter describes the proper pressure required for air pressure decrease estimation. Then, the flow of air pressure decrease estimation using vibration is described.

3.1 Proper pressure

There is a proper pressure for effective tire performance for bicycles. The maximum value of the proper pressure is referred to as the maximum proper pressure. The minimum value of the proper pressure is referred to as the minimum proper pressure. If the pressure is within the proper range, it indicates for effective tire performance. If the rider continues to ride outside of the proper pressure, the tire will suddenly burst, and the tube inside the tire will deteriorate. Therefore, air pressure inspection is important to prevent tire burst and deterioration. There may be a maximum proper pressure indicated on the side of the bicycle tire as shown in the figure 2. Also, there may be a maximum and minimum proper pressure indicated on the side of the bicycle tire. For city bicycles, only the maximum proper pressure is often indicated. For road bicycles, the maximum and minimum proper pressure are often both indicated.

Figure 2: Maximum proper pressure indicated on the side of the tire

When the air pressure is about to decrease below the minimum proper pressure, this is the best timing to inflate. If the air pressure is more than the minimum proper pressure, there is no problem to ride. If the air pressure is less than the minimum proper pressure, it is dangerous to ride. Therefore, there is a need to promote inflation to the rider when the air pressure is about to decrease below the minimum proper pressure.

3.2 Flow of air pressure decrease estimation using vibration

Bicycle vibration was used as a method to estimate the best timing for inflation. Bicycle vibration changes depending on air pressure. A schematic diagram of vibration changes associated with changes in air pressure is shown in the figure 3. If the air pressure is high, vibration is not absorbed very well. If the air pressure decreases to a certain degree, vibration is absorbed by the increased cushioning. If air pressure decreases significantly, vibration is not absorbed by the decreased cushioning. However, the rider can clearly aware that the condition is dangerous. Therefore, we conduct the estimation to the range where the rider does not aware that the condition is dangerous.

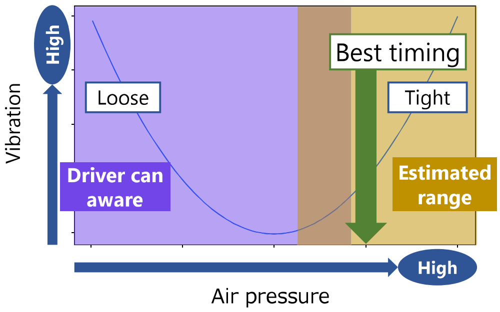


Figure 3: Schematic diagram of vibration changes associated with changes in air pressure

The rider enters the minimum proper pressure for the bicycle into the smartphone. The minimum proper pressure is necessary to acquire the best timing for inflation. The rider attaches the smartphone as shown in the figure and rides on a paved road. The smartphone is attached to the bicycle so that it is perpendicular to the ground using a smartphone holder. Vibration changes depending on the ground type ridden on. Therefore, the estimation is conducted on paved roads that are generally used. Acceleration in the y-axis direction during riding can be used as the vibration generated by the contact between the ground and the tires when the smartphone is attached as shown in the figure 4. However, the gravity acceleration is also included in the acceleration in the y-axis direction. Therefore, the acceleration in the y-axis direction without the gravity acceleration, is used.

Features of air pressure decrease are extracted from the acceleration data acquired during riding when the rider arrives at the destination. The standard deviation of acceleration in the y-axis direction and the amplitude spectrum for each frequency are used as the feature values. The acceleration standard deviation in the y-axis direction is considered to indicate the strength of the bicycle vibration. Also, the amplitude spectrum of each frequency in the y-axis direction is considered to be a

Figure 4: Smartphone attached and acceleration axis

distribution of frequencies of the vibration occurred in the riding. Air pressure decreases as tire cushioning increases within the estimated range. Therefore, it is considered that the vibration of the bicycle is decreased and the acceleration standard deviation is decreased. In addition, it is considered that the vibration at higher frequencies is absorbed and the higher frequency amplitude spectrum is decreased.

Preliminary experiment was conducted to confirm whether the standard deviation of acceleration in the y-axis and the amplitude spectrum of each frequency can be used as features for air pressure decrease estimation. The experiment was conducted with nine male undergraduate or graduate students. A smartphone was attached to the bicycle as shown in the figure 4. Subjects were asked to ride bicycle at 50 kPa increments from 150 kPa to 300 kPa for pressure and the acceleration in the y-axis direction was acquired. The acceleration in the y-axis direction was acquired as shown in the figure 5. The standard deviation was acquired from the acquired acceleration in the y-axis direction without the gravity acceleration. The acceleration standard deviations of the nine persons were averaged, and the variation of the values for each air pressure is shown in the figure 6. The standard deviation of acceleration in the y-axis direction decreases as the air pressure decreases, and a trend similar to that within the estimated range shown in the figure 3 occurs. FFT (Fast Fourier Transform) was conducted on the acceleration in the y-axis direction in order to analyze the frequency components in the y-axis direction. The amplitude spectrum for each frequency in the y-axis direction is shown in the figure 7. A trend that the amplitude spectrum with higher frequencies decreases as air pressure decreases occurred. Therefore, the standard deviation of acceleration in the y-axis direction and the amplitude spectrum for each frequency were used to estimate the air pressure decrease.

Machine learning is conducted using the features to estimate whether the air pressure is about to decrease below the minimum proper pressure. The data set format of the features is shown in the table 1. Explanatory variable is the standard deviation of acceleration in the y-axis direction and the amplitude spectrum for each frequency. The objective variable is air pressure. Amplitude spectrum for each frequency uses the average of

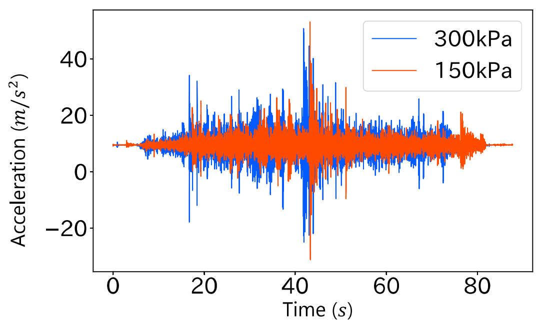


Figure 5: Acceleration in y-axis direction for each air pressure

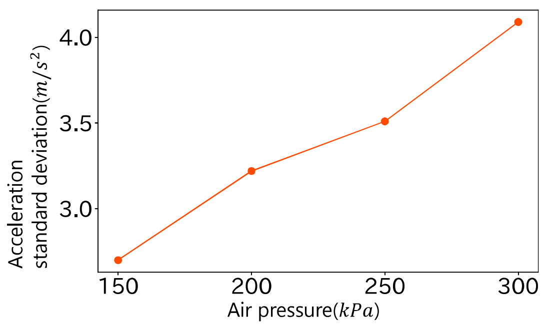


Figure 6: Variation of mean acceleration standard deviation in y-axis direction for each air pressure

0.5 Hz intervals of the frequency. Areas where no significant change occurred when air pressure decreased were not used as features. From the figure 7, little change in frequency was occurred after 40Hz. Therefore, the amplitude spectrum in the range of 0.0Hz to 40.0Hz was used. The reason for this narrowing of the range is that a larger amount of data would be burdensome to process on a smartphone. A model trained on this data set is used to classify whether the air pressure is below the minimum proper pressure. There are methods of direct classification using SVM and other methods. Another method is to conduct a regression to estimate air pressure as a numerical value using a random forest or other methods and then classify the air pressure.

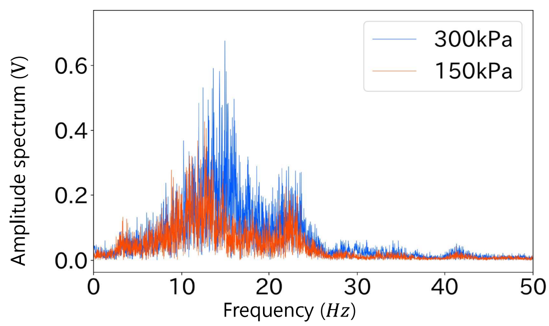


Figure 7: Amplitude spectrum for each frequency generated by bicycle riding

Table 1: Dataset of air pressure decrease features

Explanatory variable		Objective variable
Acceleration standard deviation in y-axis direction (m/s^2)	Amplitude spectrum in y-axis direction (V)	Air pressure of tire (kPa)

4 Conclusion

In this study, we focus on air pressure inspection and propose a method for estimating bicycle air pressure decrease based on vibration sensing using smartphone. The acceleration in the direction perpendicular to the ground is acquired with a smartphone, and the standard deviation and the amplitude spectrum of each frequency are used as features to estimate the air pressure decrease using random forests and other methods.

In the future, there is a need to conduct an evaluation experiment whether the algorithm used in this study can estimate the air pressure decrease. And there is a need to implement air pressure decrease estimation as a smartphone application.

REFERENCES

[1] au Insurance. survey on bicycle inspection and equipment. `https://www.au-sonpo.co.jp/corporate/news/detail-287.html`. (reference 2022-05-14).

[2] Kashimura. air pressure sensor. `https://www.kashimura.com/goods/car/tpms/kd220.html`. (reference 2022-05-14).

[3] Quarq. TyreWiz. `https://www.sram.com/en/quarq/series/tyrewiz`. (reference 2022-05-14).

[4] KUKiiRE. Smart Air Inflator. `https://kukiire.com/`. (reference 2023-01-19).

[5] Brigestone. Air Free Concept. `https://www.bridgestone.co.jp/technology_innovation/air_free_concept/`. (reference 2023-06-06).

[6] Bian T, Yulong Z, Zhuangde J, Ling Z, Nansheng L, Yuanhao L, Chao M. Fabrication and Structural Design of Micro Pressure Sensors for air pressure Measurement Systems (TPMS), *sensors*, Vol. 9, No. 3, pp. 1382-1393, 2009.

[7] Simone F, Luca O, Tommaso C, Alessandro P, Sergio M.S. h-TPMS : a hybrid air pressure monitoring system for road vehicles, *Mechatronics*, Vol. 74, 2021.

[8] Rui Y, Katsuhiko K. Basic Study of a Method for Estimating Bicycle Air Pressure Decrease Using a Sensor in a Smartphone, *Multimedia, Distributed, Cooperative, and Mobile Symposium2022*, pp. 158-162, 2022.

Evaluation of Human Recognition Method by 2D-LiDARs

Haruki Mochizuki† Takuya Watanabe* Yuya Sawano*
Yoshiaki Terashima**, Ryozo Kiyohara†

†Kanagawa Institute of Technplogy, Japan
*Graduate School of Kanagawa Institute of Technology, Japan
**Soka University, Japan

Abstract -Automated vehicle technologies which includes human detection technologies are developed by many manufacturers. Moreover, these technologies are adopted to many automated robots or carts. On the other hand, once or twice a year, an elderly person wanders into a university campus. Then human detection technologies Therefore, it is important to find human with robot carts by utilizing human recognition technology. In many cases, stationary human recognition is realized by 3D LiDAR (3 Dimensional Light Detection and Ranging) device. However it requires many computer resources. In this paper, we propose a new human recognition method by 2D-LiDARs and evaluate our method..

Keywords: LiDAR, Human Detection, Autonomous Cart, Machine Learning. PointCloud.

1 INTRODUCTION

In developed countries, the birthrate is declining and the population is aging. As a result, labor shortage has become a social problem. On the other hand, expectations for the security industry are rising and demand is increasing.

However, in the current security industry, there are physical problems, etc., and there is a shortage of personnel, and the introduction of robots that can replace humans is essential.

Since robots can operate 24 hours a day, they can be used efficiently on campuses such as universities as delivery robots and guidance robots during the day, cleaning robots in the early morning, and security robots at night.

Since people may accidentally enter the university campus, nighttime security is important. Elderly people in particular are not malicious and need to be protected by security guards. Such cases occur once a year or less, and personnel expenses should not be spent on security for such cases. Therefore, it is important for robots to detect suspicious persons. Technology for robots to judge a person and technology to judge that the person is a suspicious person are important.

If a suspicious person is someone other than a university faculty member or student, the suspicious person can be judged by whether or not the robot behaves like showing an ID when the robot approaches. Therefore, a method for robots to detect the presence of people is important.

Expensive robots are equipped with many sensors such as LiDAR (Light Detection and Ranging), cameras, and millimeter-wave radar, and technologies which has been established for autonomous vehicles for human recognition.

However, it is not realistic to have enough such expensive sensors in a robot used in a campus because of its cost. Moreover, since these technologies for autonomous driving are aimed at collision prevention with humans, it is important to reliably recognize humans rather than erroneously recognizing non-humans as humans. Furthermore, since these recognition technologies are executed while driving at a constant speed and there are various moving objects in the surroundings, it is important to detect obstacles and predict their movements.

On the other hand, On the other hand, these technologies for security have the same purpose of preventing collisions, but the accuracy of distinguishing between people and objects is important. It is also possible to take a method of sensing while repeatedly stopping the vehicle and approaching it if there is a possibility of a person to further improve the recognition accuracy.

In many cases, robots move using SLAM (Simultaneous Localization and Mapping) technology with 3D LiDAR. It runs while estimating its own position using SLAM. In SLAM, which processes 3D point cloud information, the computational complexity is very large. Therefore, the computational complexity of processing other than autonomous driving should be minimized.

Therefore, we think that it should be clarified whether one or more 2DLiDARs can identify people with the same accuracy as 3DLiDAR. The number of point clouds that can be acquired by 3DLiDAR and 2DLiDAR is clearly different as shown in Fig.1 and Fig.2.

2 PROPOSED METHOD

First, the recognition system processes the difference with the pre-trained background in SLAM and recognizes the unusual. Next, we judge whether the extracted data are human or not. For that purpose, we need to learn humans in advance. Humans learnimg from the front as shown in Fig. 3, and Fig. 4, but recognition system do not know which angle the data can be acquired. Therefore, it is necessary to confirm whether it is possible to determine whether a person is human even with data from various angles. In this paper, we evaluated data from various angles.

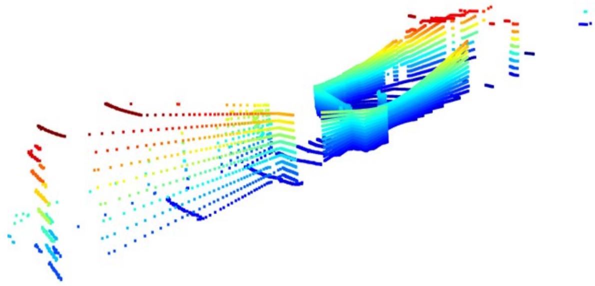


Fig.1 Example of pointcloud from 3D-LiDAR

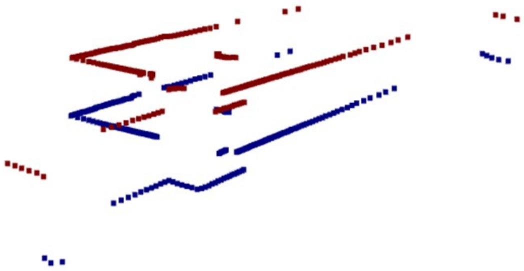


Fig.2 Example pf pointcloud from 2D-LiDARs

Fig.3 Learning of human

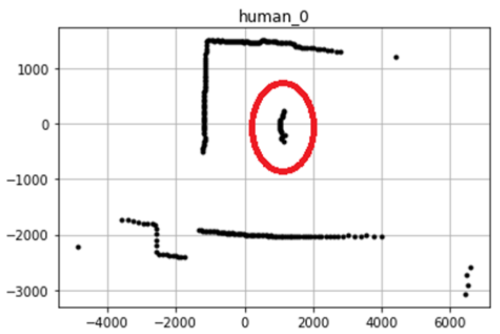


Fig.4 pointcloud data of Fig. 3

REFERENCES

[1] Velodyne LiDAR https://velodynelidar.com/products/puck/ （2023/6/3）

Estimating the best time to view cherry blossoms using Natural Language Processing tasks

Ryoga Sato[*], Yusuke Takamori[**], Kenji Terada[***], Masaki Endo[***], Tsuyoshi Tanaka[****], Shigeyoshi Ohno[***], and Hiroshi Ishikawa[*****]

[*] Electronic Information Course, Polytechnic University, Japan
b20311@uitec.ac.jp
[**] Division of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, Japan
y-takamori@jaist.ac.jp
[***] Information and Communication Engineering, Polytechnic University, Japan
{k-terada, endou, ohno}@uitec.ac.jp
[****] Digital Transformation of Skill, Polytechnic University, Japan
t-tanaka@uitec.ac.jp
[*****] Graduate School of Systems Design, Tokyo Metropolitan University, Japan
ishikawa-hiroshi@tmu.ac.jp

Abstract – With the rapidly expanding use of social networking services (SNSs), information is exchanged in real time. Taking advantage of the immediacy of this information, we have conducted research to extract seasonal events such as the best time to view cherry blossoms. Earlier studies have proposed methods using co-occurring words post information to increase the number of posts targeted for estimation of the best time to view cherry blossoms. Those studies have shown improvement in the accuracy of estimating the best time to view cherry blossoms. However, earlier methods of co-occurring word extraction also treat posts including negative words such as "not," "don't," and "nothing" as estimating the best time to view cherry blossoms, which might result in the estimation of a different time from that announced by the Japan Meteorological Agency. In this study, we focus on the inclusion of posts that do not indicate the blooming of cherry blossoms and propose a method to improve the accuracy of estimating the best time to view cherry blossoms using text classification based on natural language processing in addition to the conventional method of using co-occurring words to increase the number of data. Using this method, the effectiveness of natural language processing was verified and compared to that of conventional methods for estimating the best time to view cherry blossoms.

Keywords: Cherry blossoms; NLP; SNS; text mining; tourism

1 INTRODUCTION

With the spread of the information society, people are accessing information related to a daily basis and using the internet more and more. Furthermore, with the popularization of SNSs, an increasing number of tourists are using SNSs to gather tourist information and refer to tourist guidebooks and reviews of tourist attractions. Then, that tourist information posted by users is shared among SNS users. Tourists often benefit by this shared information. Many people use X [1], is a leading social networking platform, for sharing tourist information, posting and browse information about events, hobbies, etc. Users can also add location information to their posted messages, which called "posts." Such posts with location information, called "geotagged posts"[2], are used to share current events and locations. Because the information obtained from geotagged posts reflects the actual real-world situation, they are expected to play an important role as a tool for tourists to obtain local tourism information in real time. Therefore, we use information related to geo-tagged posts on X to estimate the best time to view cherry blossoms. Endo et al. [3] propose a method for estimating the best time to view cherry blossoms using simple moving averages. This method collects geotagged posts that include specific keywords and then estimates the best time to view seasonal organisms by obtaining a simple moving average of the number of posts per day. Takahashi et al. [4] propose a method using weighted moving averages to improve the estimation accuracy of the method described by Endo et al. This method can improve the results of estimation of the best time to view cherry blossoms in each area while maintaining lower costs than methods using simple moving averages. Takamori et al. [5] investigated the frequency of words used with the word "cherry blossom" from the content of posts, then defined the words for which skewness and kurtosis coincide with the word "cherry blossom" as cherry blossom co-occurrence words. They proposed a method of using these co-occurrence words to increase the volume of post data for estimating the best time to view cherry blossoms. This method provides more accurate estimation of the times to view cherry blossoms than conventional estimation of the time to view cherry blossoms using only the word "cherry blossom". However, when co-occurrence terms to increase the number of data are used, even posts that do not indicate the blooming of cherry blossoms are treated as targets for estimating the best time to see cherry blossoms. Therefore, even posts that have nothing to do with the blooming of cherry blossoms are treated as data for estimating the best time to see cherry blossoms. As described herein, we propose a method for estimating the best time to view cherry blossoms by applying text classification using natural language processing to information posted on X. Natural language processing is a technology or method that allows computers to understand and process sentences in the natural language spoken by humans. One type of natural language processing technology is Bidirectional Encoder Representations from Transformers (BERT) [6], a natural language processing model developed by Google Research [7] that can understand context. Compared to conventional natural language processing models such as statistical language models such as n-gram language models and vector space models such as LSA, BERT can understand not only the meanings of words but also their context. Therefore, BERT can analyze natural language with higher accuracy, and can classify and extract information. This paper is organized as follows: Section 2 describes research related to this paper; Section 3 presents an explanation of the proposed method; Section 4 presents the experimentally obtained results and their evaluation; Section 5 explains the findings obtained from this study.

2 RELATED RESEARCH

Sato et al. [8] propose a method to estimate the best time to view cherry blossoms for a certain period in the future using time series prediction for geo-tagged posts.

Meira et al. [9] combine machine learning and natural language processing methods to predict whether a user's review is positive or negative. They combine supervised and unsupervised methods to develop a model to infer whether a particular tourist attraction is liked or disliked.

Mathias et al. [10] propose a new method of personalized tour recommendation for museum visits by optimizing visitor preferences and by automatically extracting the importance of artworks from museum information. Following the proposed tour clearly has been shown to improve visitor satisfaction.

As described above, several papers have described estimation of cherry blossom viewing times using SNS and extraction of tourist information by combining natural language processing. In this paper, we demonstrate the effectiveness of estimating the best time to view cherry blossoms by combining natural language processing and the moving average method.

3 PROBREM AND METHOD

Since tourists make plans according to seasonal events, such as cherry blossom viewing, and tourism workers organize their events accordingly, there is a need for tools to estimate seasonal events. Therefore, a X-based method for estimating the best time to view cherry blossoms is one way to solve this problem. On the other hand, when we looked at the content of posts used to estimate the best time to see the cherry blossoms, we found a scattering of posts unrelated to the blooming of cherry trees, such as the names of places and foods, such as "Sakurazaka" and "Sakura mochi. Therefore, we thought that removing posts that had nothing to do with the blooming of cherry trees would improve the accuracy of estimating the best time to see cherry blossoms. We also chose natural language processing as the method for classifying posts. In this paper, we propose a method for estimating the timing of cherry blossom viewing by using natural language processing to classify text posted on X.

This section describes the proposed methodology. First, we explain the main flow of the proposal. For this method, experiments were conducted in Tokyo and Kyoto based on descriptions in a report by Takamori et al. First, we select words that co-occur with cherry blossoms from posts which include the word "cherry blossom" in their text, which were used in an earlier study. Next, the selected co-occurring words are used to collect posts which contain one of those words in the text. Next, we use the selected co-occurring words to collect posts that contain one of those words in the body of the text. After the collected posts are used to fine-tune BERT, then text classification is done using the BERT model created for the collected posts. Finally, the posts collected and the posts with text classification are used to infer the best time to view cherry blossoms during the period from March 1 to April 30 in 2021 and 2022.

3.1 Collecting posts for co-occurrence selection

We collect posts that include the kanji, katakana, and hiragana of the word "cherry blossom" in the text. The period extended from February 2015, when we started collecting posts, to May 2022. We collected posts for the two areas to estimate the best times to view cherry blossoms.

3.2 Morphological analysis

Morphological analysis of the posts we collected in 3.1 was performed using MeCab [11] to break down the post body into words. MeCab, an open-source tool for morphological analysis of Japanese, breaks down sentences into words and parts of speech. This study applied the following preprocessing steps for morphological analysis. First, URLs and symbols were eliminated as a cleaning process to remove noise in the text. Next, the words were normalized to facilitate morphological analysis. Additionally, we converted full-width alphanumeric characters to half-width alphanumeric characters, half-width kana to full-width kana, and all alphabetic characters to lowercase. All numerals were replaced with "0". After such preprocessing, the post text was decomposed word-by-word using MeCab. Furthermore, using MeCab, a library for morphological analysis, and IPAdic, a dictionary for morphological analysis, we extracted words which corresponded to the three parts of speech "noun," " adjectival noun," and "verb" which could be adopted as collocations. Then we counted their frequency of occurrence. An adjectival noun, which is one part of speech in the large category by UniDic [12], is the part of speech which represents the stem part of an adjectival verb.

3.3 Co-occurrence selection

We only performed co-occurrence judgments for the top 1% of words we collected in section 3.2 in terms of frequency. The method used for determining co-occurrence was to calculate the skewness kurtosis of the frequency distribution of each word and to identify those words which met certain criteria as co-occurring words. Skewness kurtosis is a statistic that expresses how much the distribution is skewed from a normal distribution and the degree to which it is sharp. It was judged to be an appropriate indicator for identifying the co-occurring words which have the same peak as the keyword in the annual change of posts. The criteria for judgment are equations (1) and (2) shown below. The formulas are expressed as Ss, where S and K respectively denote the skewness kurtosis of the frequency distribution of each word, and where Ss is the skewness obtained from posts that include "cherry blossom" in the text.

$$(Ss - 1) \leq S \leq (Ss + 1) \quad (1)$$

$$K > 2 \quad (2)$$

Skewness kurtosis was calculated by extracting posts in which words were included in the text from January 1, 2018 to December 31, 2018. Skewness Ske kurtosis Kur is obtained from equations (3) and (4) below, where n is the number of days in the period and $x_i(i{:}1, 2, \ldots, n)$ is the value for each day in the period. From the time period used for this study, n is 365. Also, $\bar{x}$ denotes the mean value for the specified period; s stands for the sample standard deviation.

$$Ske = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^3 \quad (3)$$

$$Kur = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (4)$$

For this study, we set the period to include all months during January−December because it was necessary to check the trend of posts showing the same trend as the cherry blossoms throughout the year. In addition, the posts were collected on a server managed by our manuscript team, but there were periods during which collection was stopped because of power outages, etc. The calculated skewness kurtosis values would be skewed if such periods were included in the calculation. Therefore, to include as little as possible of the period during which we were not able to collect data, we adopted the period of January 1, 2018 − December 31, 2018 as the period used to calculate the skewness kurtosis, which satisfies these two conditions.

3.4 Fine-Tuning of BERT

In the conventional method, if a post includes a word to be searched for, it is treated as a target for estimating the best time to view cherry blossoms, even if the post does not indicate the blooming of cherry trees. Examples of posts that do not indicate cherry blossoming include names of places, stores, and people. The purpose of this method is to increase the number of data using co-occurring words, and in addition, to improve the accuracy of estimating the timing of cherry blossom viewing by eliminating posts unrelated to the blooming of cherry trees through text classification by BERT.

For text classification by BERT, fine-tuning was performed using a pre-training model from Tohoku University [13]. For this method, we collected posts in Tokyo that contained either kanji, katakana, hiragana word "sakura" or one of the co-occurring words among the post information accumulated to date. During the period of 13 days from March 23 through April 4, 2015, we collected 9704 posts used as training data for natural language processing. This is true because the difference of 6 days between the bloom date and the full bloom date was taken on the back side as well because the bloom date in Tokyo in 2015 was March 23 and because the full bloom date was March 29. The following preprocessing was performed for creating the training data used for fine-tuning. First, line feed codes in the text data were removed and alphabetic characters were converted to lowercase. Next, URLs were removed to eliminate noise in the text. Emoji characters were removed using demoji [14]. Half-width and full-width symbols were then removed and the Japanese text was normalized with neologdn[15]. Also, as a process for numbers, numbers separated by commas were replaced with 0. All numbers were replaced with 0.

As a criterion for fine-tuning, we classified posts as posts indicating that cherry blossoms are in bloom (LABEL_1) and posts that are unrelated to cherry blossoms (LABEL_0). Specific post classifications are shown in Table 1.

3.5 Text Classification using BERT

We collected posts including the word "cherry blossom" in kanji, katakana, and hiragana in the text and any of the co-occurring words in the text for the period of February 1, 2015 − February 28, 2022 for the two areas. As a pre-processing step for text classification by BERT, the same cleaning process that was performed for the training data in 3.4 was performed for post data. Then, the BERT model tuned in 3.4 was used to perform text classification using the respective post data.

3.6 Estimating the best time to view

The target posts for estimating the best time to view cherry blossoms were posts containing "cherry blossom" in kanji, katakana, and hiragana, posts containing any of the

Table 1: Examples of labeling criteria for post contents

Label	Post contents
LABEL_1	お花見シーズンsoon come
	今週末にはお花見散歩出来そうですね
LABEL_0	今日は花粉日和っぽいいかにも
	大学に来たら改修工事真っ最中
	桜餅食べてる
	桜坂に来てみた

co-occurring words, and the results of text classification using BERT for posts of these two types. The period for estimating the best time to view the cherry blossoms was from March 1 through April 30 in 2021 and 2022. Estimation of the viewing period was done using the same simple moving average method as used for the method described earlier. Since the number of posts tends to increase on Saturdays and Sundays, a moving average is obtained. An *n*-day simple moving average is obtained using equation (5), where x represents the number of posts per day.

$$Avg\ n = \frac{1}{n}\sum_{i=0}^{n-1} x(k-i) \tag{5}$$

Periods satisfying equations (6) and (7) below simultaneously for three or more consecutive days were inferred as estimating the best time to view cherry blossoms. The period with more posts than the one-year average was identified using equation (6). The period with a sharp increase in the number of posts, as derived using equation (7), is inferred as the best time to view cherry blossoms. In equations (6) and (7), x_i represents the number of posts on day *i*; Avg *n* represents the moving average over *n* days.

$$x_i > \text{Avg } 365 \tag{6}$$

$$\text{Avg } 10 < \text{Avg } 20 \tag{7}$$

When only posts containing the word "cherry blossom" in kanji katakana hiragana in the text were retrieved, the number of posts per day was roughly 30, but the value of the time-series prediction using posts with co-occurring words in the text was about 400 on average, resulting in a large error margin. Therefore, to avoid loss of estimation accuracy, we used moving averages that were larger than the 5- and 7-day moving averages used in previous papers.

4 EXPERIMENT RESULTS

This section presents results of co-occurrence word determination and fine-tuning of the text classification model conducted in this study, as well as a result of estimating the best view cherry blossoms. The loss rate achieved using the created text classification model was 16.7%, requiring 41 epochs. The results of the co-occurrence word judgments and the estimation of the best time to view are presented below. Tables 2 and 3 present the results we collected for each of the two Tokyo and Kyoto co-occurrence word datasets obtained from the co-occurrence word judgments. As shown in Table 2, Tokyo includes the name of Ueno Park, a famous cherry blossom viewing spot, and petal-related words related to cherry blossoms. For Kyoto, as shown in Table 3, the names "Gion" and "Uji" are included, which are also famous for their cherry blossoms. This finding suggests that the proposed method for determining co-occurring words can determine words that are related.

Table 2: List of Co-occurring Words in Tokyo

cherry blossom	sakura	uenopark	さく	ソメイヨシ
ヒルズ	ミッドタウン	七	五反田	年
付着	義	初	咲い	咲き
国際	大学	大崎	始め	川沿い
恩賜	日和	日野	最後	本髪
来年	毎年	江東	皆	花
花びら	花園	見物	見頃	調布
開花	靖国			

Table 3: List of Co-occurring Words in Kyoto

cherry blossom	割	咲き	宇治	平野
年	散っ	淀川	疏水	祇園
神社	蹴上	開花		

Next, Figures 1 and 2 present results obtained from estimation of the best time to view cherry blossoms for Tokyo and Kyoto based on post data from March 1, 2021 through April 30, 2021. Figures 3 to 4 present results obtained for estimating the best time to view for Tokyo and Kyoto based on post data of March 1, 2022 through April 30, 2022. Arrows in the figure indicate the period from the date of bloom to the date of full bloom as announced by the Japan Meteorological Agency. They were used as the correct period for evaluating estimation results. In each figure, the number of posts with co-occurring words in the text and the number of posts with text classification by BERT are shown as line graphs, with the estimated time period and times of observations shown as a fill color. However, since the objective is to see changes in the number of posts, the vertical axis values of posts for the conventional and proposed methods are different. Figure 4 portrays the results obtained for Kyoto, where the proposed method estimates the best time to view cherry blossoms about two weeks longer than the correct period. Presumably, this result was obtained because the correct prediction period was from the date of bloom to the date of full bloom, and because it takes about 10 days to 2 weeks from the date of full bloom to the end of the period when the cherry blossoms are at their best after they have fallen. Presumably, the same reason applies for judging the period after the correct period as the viewing period in Figures 1−3. In Figures 1 and 2, the existence of a viewing period before the correct period might be attributable to the earlier blooming of cherry blossoms in 2021 because of abnormal weather conditions compared to a normal year. Figure 3 shows a wider period of time estimated as the viewing period than the correct period, presumably because the text classification model classifies cherry blossoms as blooming, irrespective of the degree of blooming. Therefore, posts indicating that cherry trees are about to bloom or that they are in the process of falling are also used as data. This phenomenon is believed to be the same as that shown in Figures 1, 2, and 4.

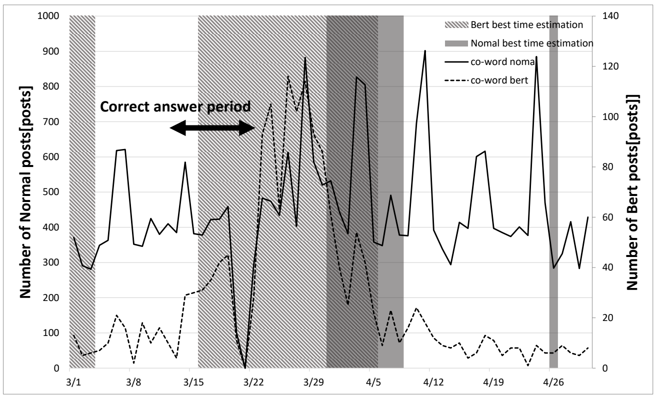


Figure 1: Estimated results in Tokyo in 2021.

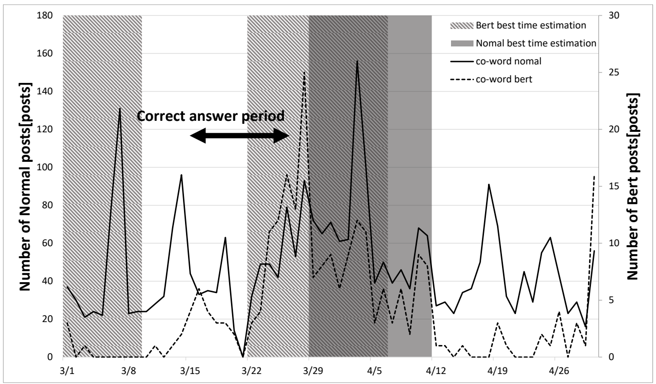


Figure 2: Estimated results in Kyoto in 2021.

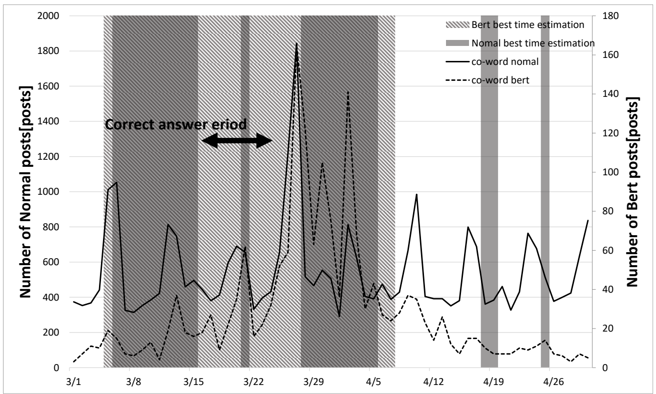


Figure 3: Estimated results in Tokyo in 2022.

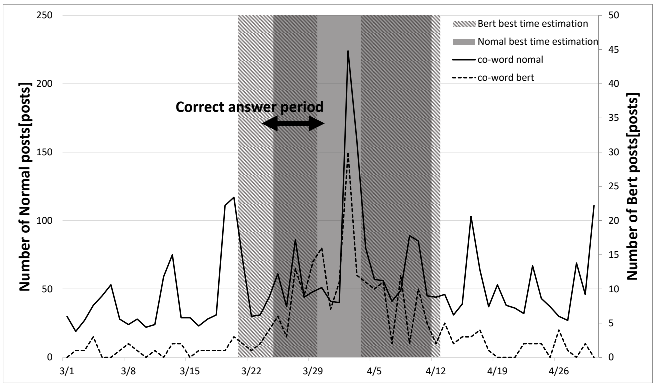


Figure 4: Estimated results in Kyoto in 2022.

We evaluated the results of estimating the best time to view cherry blossoms using conventional methods and the results of classifying posts using text classification by BERT to estimate the best times to view cherry blossoms. Table 4 presents each method and the recall rate compliance rate, an index to evaluate the accuracy of each method, for Tokyo in 2021. Table 5 presents the recall rate compliance rates of the respective methods and the index used to evaluate its accuracy in Tokyo in 2022. Table 6 presents the recall rate compliance rates of the respective methods and the index used to evaluate its accuracy in Kyoto in 2021. Table 7 presents the recall rate compliance rates of each method and its accuracy in Kyoto in 2022. The period from bloom to full bloom announced by the Japan Meteorological Agency is the correct period. The recall rate is an indicator reflecting how well the estimated results fall within that period. The proportion of the estimated results which are included in the correct period is calculated as the compliance rate. In Tables 4−7, the Earlier method refers to results of estimating the best time to view cherry blossoms without text classification by BERT. The proposed method refers to results obtained by estimating the best time to view cherry blossoms based on the classification of posts using text classification by BERT. In addition, "or" in each table indicates that a post is considered to be in the season if it is judged to be in season either by the results of posts which include kanji, katakana, and hiragana in the body of the post or by the results of posts that include co-occurring words in the body of the post. Also, "and" is an estimation of the best time to view cherry blossoms from a post when both the results of posts containing kanji, katakana, and hiragana in the body of the post and the results of posts containing co-occurring words in the body of the post indicate that the post is estimating the best time to view cherry blossoms.

Table 4: Best Viewing Results for Tokyo in 2021

	type	Earlier method	Proposed method
Recall (%)	only sakura	55.6	88.9
	co-word	0.0	77.8
	or	55.6	88.9
	and	0.0	77.8
Precision (%)	only sakura	27.8	44.4
	co-word	0.0	29.2
	or	18.5	29.6
	and	0.0	46.7

Table 5: Best Viewing Results for Tokyo in 2022

	type	Earlier method	Proposed method
Recall (%)	only sakura	50.0	50.0
	co-word	12.5	100.0
	or	62.5	100.0
	and	0.0	50.0
Precision (%)	only sakura	26.7	30.8
	co-word	4.3	23.5
	or	16.1	23.5
	and	0.0	30.8

Table 6: Best Viewing Results for Kyoto in 2021

	type	Earlier method	Proposed method
Recall (%)	only sakura	63.6	63.6
	co-word	0.0	45.5
	or	63.6	63.6
	and	0.0	45.5
Precision (%)	only sakura	41.2	41.2
	co-word	0.0	20.0
	or	25.9	25.9
	and	0.0	33.3

Table 7: Best Viewing Results for Kyoto in 2022

	type	Earlier method	Proposed method
Recall (%)	only sakura	57.1	100.0
	co-word	85.7	85.7
	or	100.0	100.0
	and	42.9	85.7
Precision (%)	only sakura	26.7	38.9
	co-word	33.3	33.3
	or	26.9	30.4
	and	42.9	46.2

Regarding the recall rate precision rate for Tokyo in 2021 in Table 4, 'only sakura', 'co-word', 'or', 'and', the accuracy of the proposed method are better than those of the conventional method. In Table 5, for Tokyo in 2022, the accuracy of the proposed method is better than that of the conventional method for all items, except for the recall rate of 'only sakura'. The recall rate of 'only sakura' is the same for both the conventional and the proposed methods. In Table 6, for Kyoto in 2021 the recall rates of 'co-word' and 'and', and the precision rates of 'co-word' and 'and' the accuracy of the proposed method are better than those of the conventional method. The recall rate of 'only sakura' and 'or' and the precision rate of 'only sakura' and 'or', which are other items, are equal for both the conventional and the proposed methods. In Table 7, for Kyoto in 2022 the recall rates of 'only sakura' and 'and', and the precision rates of 'only sakura', 'or' and 'and' the accuracy of the proposed method are better than those of the conventional method. The recall rate of 'co-word' and 'or' and the precision rate of 'co-word', which are other items, are equal for both the conventional and the proposed methods.

In the two areas for which we estimated the best time to view cherry blossoms this time, the recall rate and precision rates of the proposed method were higher by 35.8% and 14.8%, respectively, compared to the conventional method, indicating the effectiveness of the proposed method. These results demonstrate that the proposed method can provide better accuracy than the conventional method.

5 CONCLUSION

This study demonstrated higher accuracy when estimating the best time to view cherry blossoms using text classification by BERT. To improve the accuracy of estimation, we proposed a method of text classification using BERT to eliminate posts that are unrelated to cherry blossom blooming. When the proposed method was used to estimate the best time to view cherry blossoms, the recall rate and the precision rate were higher by 35.8% and 14.8%, respectively, compared to results obtained using the conventional method, which indicates that the proposed method has the potential to improve accuracy. The loss rate of the text classification model created in this paper was 16.7%, suggesting that there is admit for improvement. Therefore, we would like to use other natural language processing models such as RoBERTa and DistilBERT for comparison and analysis in order to improve accuracy. Using the proposed method to classify text using natural language processing, it should be possible to collect more data in the future without relying on co-occurrence judgments. Therefore, we would like to adapt estimation of the best time to view cherry blossoms posts to areas where posts are few, and extend our estimation to areas with more detailed regional divisions.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 20K12081, 22K13776, and the Telecommunication Advancement Foundation Research Grant.

REFERENCES

[1] X, https://X.com/ (referred in June 2023)

[2] X Developers, https://dev.X.com/ (referred in June 2023)

[3] M. Endo, K. Mitomi, K. Saeki, Y. Ehara, M. Hirota, S. Ohno, and H. Ishikawa, Study of information provided by the best time to see estimation method of phenological observations using tweets, Proceedings of the 12th Annual Conference on Japan Society for Tourism Informatics, Shizuoka, Japan, pp.47-60 (2016).

[4] M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Automatic detection of tourist spots and best-time estimation using social network services, International Workshop on Informatics 2020, pp.65-72 (2020).

[5] Y. Takamori, K. Terada, M. Endo, S. Ohno, and H. Ishikawa, Co-occurring Word Determination Used for Estimating Best Times for Viewing Cherry Blossoms, MMEDIA 2023 Proceedings, pp.1-6 (2023) .

[6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805, (2018).

[7] Google Research, https://research.google/ (referred in June 2023)

[8] J. Sato, Y. Takamori, M. Fujimoto, T. Horikawa, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Verification of the Best Time to See Cherry Blossoms Estimation Method using Time Series Prediction Method, IWIN2022 - Proceedings, pp.95-101 (2022).

[9] J. Meira, J. Carneiro, V. Alonso-Betanzos, et al., Anomaly detection on natural language processing to

improve predictions on tourist preferences, Electronics, Vol.11, No.5, pp.779 (2022).
[10] M. Mathias, A. Moussa, F. Zhou, et al., Optimisation using Natural Language Processing: Personalized Tour Recommendation for Museums, 2014 Federated Conference on Computer Science and Information Systems, IEEE, pp.439- 446 (2014).
[11] T. Kudo, Mecab: Yet another part-of-speech and morphological analyzer, http://mecab.sourceforge.net/ (referred in June 2023)
[12] National Institute for Japanese Language and Linguistics, Electronic Dictionary with Uniformity and Identity, https://clrd.ninjal.ac.jp/unidic/ (referred in June 2023)
[13] Tohoku NLP Group, cl-tohoku/bertbase-japanese-v2 · Hugging Face, https://huggingface.co/cl-tohoku/bert-base-japanese-v2 (referred in June 2023)
[14] B. Salomon, J. Spaans, Demoji, https://github.co-m/bsol-omon1124/demoji (referred in June 2023)
[15] I. Yukino, neologdn, https://pypi.org/project/neolog-dn/ (referred in June 2023)

Relation between time-series forecasting methods and estimation of best times to see cherry blossoms

Asahi Iha*, Kenji Terada**, Masaki Endo**, Tsuyoshi Tanaka***, Shigeyoshi Ohno**, and Hiroshi Ishikawa****

* Electronic Information Course, Polytechnic University, Japan
b20305@uitec.ac.jp
** Information and Communication Engineering, Polytechnic University, Japan
{k-terada, endou, ohno}@uitec.ac.jp
*** Digital Transformation of Skill, Polytechnic University, Japan
t-tanaka@uitec.ac.jp
**** Graduate School of Systems Design, Tokyo Metropolitan University, Japan
ishikawa-hiroshi@tmu.ac.jp

Abstract – Analyzing SNS data enables low-cost phenological observation. Tourism operators can use results of these analyses to plan seasonal events such as cherry blossom viewing tours. In an earlier study, after posts including the word "sakura" (cherry blossom) on the subject of cherry blossom viewing were collected, a time-series analysis was used to predict the number of future posts. The distribution state of the weighted moving average of this predicted time-series data made it possible to estimate the best time to see cherry blossoms in the future. Nevertheless, results of an earlier study that estimated the best time to see cherry blossoms using the number of actual posts were less accurate than the results of a time series analysis that estimated the best time to view cherry blossoms. After investigating this finding, we surmised that rounding the number of posts in time-series prediction might have improved accuracy. Therefore, to improve the accuracy of cherry blossom season estimation, this study first compares the number of posts predicted by time series analysis using DeepAR+ and Prophet, which was used for an earlier study, with the number of actual posts. Then we compared the accuracy of each to ascertain how it affects estimation of the cherry blossom season. We investigate the effects of the numbers of posts on estimation of the cherry blossom viewing season by comparing the accuracy of the two. Based on the findings obtained using the respective methods, we will compare and contrast the methods for improving the accuracy of cherry blossom season estimation through changes in the degree of accuracy of time-series predictions, and will then propose effective methods.

Keywords: Cherry blossoms; machine learning; SNS; text mining; tourism

1 INTRODUCTION

Today, with the development of information technology, many people are finding information via the internet. Particularly with the spread of social networking services (SNSs), users' information is shared. The latest information is diffused instantly. Therefore, SNSs might be used for gathering tourist information. On SNSs, an overflow of tourist information occurs through many postings. Using this information, one can obtain detailed local information as if one were talking to people at the sightseeing spot. One example of a social networking service with shared tourism information is X [1], a system that allows users to share their daily post messages within 140 characters. X is generally used for posting and viewing information about personal events, hobbies, etc. Posting text, photos, videos, and other information on X is called "posting. Among those messages, posts with location information added to an arbitrary setting are called "geotagged posts. Geotagged posts make it possible to share what is actually happening in the real world by including current location information. It is also possible to obtain information related to tourist spots and events through posts made by local people. Therefore, X is anticipated for use as a social sensor for estimation and acquisition of local tourist information in real time.

Using SNSs, it is possible to estimate the seasonality of plants at a low cost. Currently, biological seasonal observations, which record plants and organisms, are made by the Japan Meteorological Agency based on Phenological Observation Guidelines. Nevertheless, phenological observations are costly because they involve visual observations of sample trees. Furthermore, the Japan Meteorological Agency has announced that it will drastically reduce the scale of phenological observations starting in January 2021. Currently, plants of only six types are subject to biological seasonal observation. Therefore, estimating the best time to see cherry blossoms using SNSs is anticipated as a new observation method. We are estimating the best times to see cherry blossoms by extracting information related to cherry blossoms using geotagged posts on X. Endo et al. [2] proposed a method that uses simple moving averages to estimate the best times to view cherry blossoms. This method, which collects geotagged posts that include specific keywords, estimates the best time to see species by obtaining a simple moving average of the number of posts for each date. In other words, estimating the best time to view cherry blossoms is possible in prefectures and localities where a certain number of geotagged posts are displayed. Using this approach, they confirm the real-world seasonality of posts related to the names of organisms in each region. Then they demonstrate the potential for presenting tourism information in real time. Takahashi et al. [3] proposed a method using weighted moving averages to improve the estimation accuracy of the method presented by Endo et al. Compared to the method using a simple moving average, this method can improve the cherry blossom viewing season estimates for each prefecture while maintaining low costs. Horikawa et al. [4][5] applied the estimation method presented by Takahashi et al. and proposed a method for estimating the best time to see cherry blossoms in the future using time-series forecasts based on machine learning. Until that study, estimation of viewing times using actual measurements was only able to estimate the current time. Time-series forecasting is the process of predicting data trends over a future period based on past time-series data. Because cherry blossoms are seasonal plants, time-series forecasting is useful to predict the number of posts related to cherry blossoms for a certain future period. The judgment of when the cherry blossoms are at their best is made by analyzing the relative trends in the number of posts. Therefore, we have developed a method for highly accurate estimation of the best time to view cherry blossoms using time-series predictions of posts. Sato et al. [6] apply time-series forecasting using machine learning by Horikawa et al. to enable the application of a method for estimating the best times to see cherry blossoms nationwide by covering a wide area and by changing the start date of the time-series forecasting. However, the relation between time-series forecasting and estimation of cherry blossom viewing season was unclear. Therefore, this study compares time-series forecasts by Prophet and DeepAR+. Then, we will investigate the relation between the time-series prediction and the estimation of cherry blossom season using the characteristics of the time-series prediction.

The structure of this paper is the following: Section 2 presents a description of research related to this paper; Section 3 presents the proposed method; Section 4 explains the experimentally obtained results and their evaluation; and Section 5 gives some conclusions obtained from this study.

2 RELATED RESEARCH

Silva et al. [7] proposed noise reduction in Neural Network Autoregressive (NNAR) to improve forecasting accuracy. Better prediction of future tourism demand was achieved using the NNAR model.

Colladon et al. [8] proposed a new method for analyzing tourism-related big data and a set of variables that can be integrated into conventional forecasting models. They were able to predict the number of passengers at international airports in Europe, which is useful for the tourism industry.

Takamori et al. [9] proposed a method for estimating the best times to see cherry blossoms at a certain period using posts collected with co-occurring words related to cherry blossoms.post The use of co-occurrence terms enabled them to increase the number of data, thereby improving accuracy. The recall rate was 34.3%. Moreover, recall rate was 4.0% higher than for the results obtained using the conventional method.

Morishita et al. [10] proposed a participatory sensing system called Sakura Sensor, which automatically extracts landscape route information from video captured by a smartphone mounted on a car and shares it with users in real time. In the cherry blossom sensor research, only information on the blooming of cherry blossoms along the road is available because the cherry blossoms are observed by a car. In our research, however, many people who use X post posts from various areas, which allows us to see information on the blooming of many cherry trees.

Research undertaken to extract tourist information using SNS, research on the estimation of cherry blossom viewing times, and research on time-series prediction using machine learning are progressing. This study was conducted to demonstrate the usefulness of a method for estimating future cherry blossom viewing times using time series forecasting, especially using machine learning.

3 PROBLEMS AND METHOD

This chapter presents descriptions of the data and the method used to estimate the timing of cherry blossom viewing, as well as the problems involved.

3.1 Problem

Our research aims at estimating the best time to view cherry blossoms with a high degree of accuracy on a national scale. However, prior research has the issue that "the relation between time-series prediction and estimation of cherry blossom viewing time is unclear".

This is the problem that although time-series forecasting has made it possible to estimate the future timing of cherry blossom viewing, it is not clear how the forecasting of posts affects the estimation of cherry blossom viewing timing. In the previous study, estimation of best time was made immediately using the conventional condition of best time for cherry blossoms after the time-series prediction. In this study, we compare algorithms suitable for estimating cherry blossom viewing time. Next, time-series predictions are made using the algorithm suitable for estimating cherry blossom viewing time, and the predicted results are compared with actual posts to investigate the information obtained by making predictions. Based on the obtained information, we propose a new method by adding new conditions for estimating the best time to see the cherry blossoms. By using the proposed method to estimate the timing of cherry blossom viewing, we can investigate the impact of the predicted posts on the estimation of cherry blossom viewing. We hope that this will solve the problem of "unclear relation between time-series prediction and estimation of cherry blossom viewing season".

3.2 Data preparation

For this study, we used the X Streaming API to collect geotagged posts from Japan that include location information [11]. Then, using the simple reverse geocoding service [12] of the Ministry of Agriculture, Forestry and Fisheries based on latitude and longitude information, the collected geotagged posts were assigned to the prefecture from which they originated. The timing of cherry blossom viewing was estimated by predicting and analyzing the transition of posts including specific keywords from posts in each prefecture. The target regions for the experiment were Tokyo and Kyoto, each of which has a large population, and each of which is famous for cherry blossoms. The key word was "cherry blossom" in kanji, hiragana, or katakana; the analyzed period was February 1, 2015 – February 28, 2022. However, we excluded 2020 post data from our training data because we were unable to collect those data.

3.3 Time series forecast

Based on post information collected in the past, time-series prediction of the number of posts was performed during the estimated cherry blossom viewing period. Because cherry blossoms bloom mainly during March–April, the estimated period of cherry blossom viewing was the two months of March 1 – April 30, 2022. By performing time-series forecasting of the number of posts during this period, it was possible to estimate the best seasons for viewing in two-month increments. The method uses statistical and machine learning algorithms from Amazon Web Services (AWS) [13]. We also use a service called Amazon Forecast [14], which provides very accurate time-series forecasts.

As training data for this study, we use a large dataset of more than 300 days of data for cherry blossoms that bloom around spring of each year. Therefore, the algorithms used from Amazon Forecast are Prophet, which is suitable for seasonal training data, and DeepAR+, which is suitable for large data of more than 300 days. The predicted value is the value which satisfies 50% demand obtained from the weighted quantile loss: a metric used during forecasting with Amazon Forecast. The predicted number of posts is rounded down to the nearest whole number. If the number of posts is negative, then that finding is treated as 0. We use the values obtained from time-series forecasts to estimate the timing of cherry blossom viewing. The data to be trained will be processed for each algorithm so that the algorithm characteristics can be used.

Prophet, an additive regression model, combines the four characteristics of periodicity, trend, event, and error to make forecasts. It is most effective for time series with strong seasonal effects and several seasons of historical data [15][16].

Because this study is collecting posts related to cherry blossom viewing season, the training data will be seasonal, with features more likely to appear around springtime. Prophet is an open source software released by Facebook's Core Data Science team. It is a forecasting method for time series data based on an additive model that takes into account annual, weekly, and daily seasonality as well as holiday effects in the nonlinear trend. The model equation is shown below.

$$y(t)=g(t)+s(t)+h(t)+\varepsilon\_t \qquad (1)$$

y(t) is the predictor variable, g(t) is the trend function, s(t) is the seasonal change, h(t) is the holiday and event term, and ε_t is the error term that follows a normal distribution. When using Prophet, we use data for all months from January through December for forecasting. Although 2020 data are missing, Prophet is trained normally. We use it as-is without making up for the missing data. Amazon Forecast's DeepAR+ is a modified version of DeepAR for time series analysis that jointly trains a single model over several types of time series data [17]. DeepAR+ is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNNs). Therefore, DeepAR+ is a potentially usable algorithm for this study [18]. The characteristic period of the number of posts about cherry blossoms appears during the period from the beginning of February to the end of April each year. However, the characteristic period becomes shorter when viewed on a yearly basis. Therefore, the training data are processed to take advantage of DeepAR+. Data for three months, the period from the beginning of February to the end of April of each year, are extracted. Next, the extracted data are processed with February 28, 2022 as the base date to make them continuous values. To give a specific example of processing, the data for February−April from 2015 through 2021 and for February 2022 were extracted, with February 1 through April 30, 2021 as November 4, 2021 through January 31, 2022, February 1 through April 30, 2019 as August 7, 2021 through November 3, 2018, and February 1 to April 30, 2018 as May 10 to August 6, 2021.

3.4 Best-time estimation method

The condition for estimating the best time to see cherry blossoms by post transition uses the conventional method of weighted moving averages, for which each value is weighted. In the method used for this study, the median value is set as 1. The minimum and maximum values are set respectively as ±0.5 from the median value. To estimate the best viewing period for cherry blossoms, we analyze the frequency of geotagged posts including the key word. The following three values will be analyzed by date and will be used as the basis for estimating the optimal viewing period.

1. 1-year simple moving average of the number of posts
2. 7-day weighted moving average of posts
3. 5-day weighted moving average of posts

To assess the increase in the frequency of geotagged posts including the key words, a one-year simple moving average of the number of posts was obtained. Because posts including key words also include words that are unrelated to actual cherry blossoms, one can ascertain the increase in posts related to actual cherry blossoms by comparing the number of posts with the one-year simple moving average of the number of posts. To analyze the increasing trend of the number of posts, a 7-day weighted moving average and a 5-day weighted moving average are obtained. Because the numbers of posts tend to increase on Saturdays and Sundays, a 7-day weighted moving average is obtained. The formula to be used is shown below. The expression x_y denotes the number of posts x of y days prior.

$$H_{avg7} = \frac{0.5x_7+0.67x_6+0.83x_5+x_4+1.17x_3+1.33x_2+1.5x_1}{7} \qquad (2)$$

One can also find the weighted moving average of 5 days, which is the average number of days between the cherry blossom bloom and full bloom. The formula to be used is presented below. x_y is the number of posts x of y days prior.

$$H_{avg5} = \frac{0.5x_5+0.75x_4+x_3+1.25x_2+1.5x_1}{5} \quad (3)$$

Using these estimation criteria, we can ascertain that cherry blossoms are at their best when the following two conditions are met.

1. The number of posts is greater than the 1-year simple moving average.
2. The 5-day weighted moving average of the number of posts is greater than the 7-day weighted moving average for three consecutive days.

These conditions are used to estimate the best time period to see each attraction. Using conventional methods, the estimation conditions for cherry blossom viewing were defined based on the actual number of posts.

For this study, the maximum number of posts per day in a time-series forecast period is designated as the maximum forecast peak value. The total number of posts for the time-series forecasting period is designated as the total number of posts. The maximum predicted peak value and the total number of posts can only be obtained by performing time-series forecasting. This forecasting is used as a condition for estimating the best time to see the cherry blossoms. We propose two new conditions for estimating the best time to view cherry blossoms. The conventional method uses a one-year moving average because the number of posts in the future is not known. In proposed method 1, the maximum predicted peak value is used to narrow down the estimation period from the one-year moving average, thereby improving the precision rate accuracy. In proposed method 2, the accuracy of the estimated precision rate is improved by estimation from a new angle.

Proposed method 1:
A number of posts that is larger than the predicted maximum number of posts multiplied by 0.3, 0.5, or 0.7, and meeting the estimation criteria of the conventional method is judged to be the best time to view the cherry blossoms.

Proposed method 2:
The ratio of the total number of posts to the cumulative number of posts from March 1 to a certain date is used to determine the time at which the cherry blossoms are at their peak.

3.5 Method for determining the relation between time-series forecasts

To normalize the number of posts obtained from the time series forecast to the actual number of posts, we divide the number of posts per day for the period March 1 − April 29, 2022 by the total number of posts for that period to obtain the ratio of posts per day for the period. The ratio of the number of posts is then evaluated using the mean squared error (MSE). The mean squared error is calculated using the following equation: y_i denotes the ratio of the actual number of posts; $\hat{y}_i$ represents the ratio of the predicted number of posts; and n stands for the total number of period data. Smaller values of MSE indicate smaller errors.

$$MSE = \frac{1}{n}\sum_{i=0}^{n=1}(y_i - \hat{y}_i)^2 \quad (4)$$

Another evaluation is the correlation coefficient (R), which is calculated as follows: x_i is the ratio of the actual number of posts, with $\bar{x}$ representing its average; y_i is the ratio of the predicted number of posts, with $\bar{y}$ representing its average; and n is the total number of period data. An absolute value of R that is closer to 1 signifies stronger correlation; a value closer to 0 represents weaker correlation.

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (5)$$

4 RESULTS AND DISCUSSION

4.1 Results

This chapter presents results obtained using Prophet and DeepAR+ for the cherry blossom viewing season estimation conditions from earlier studies, as well as cherry blossom viewing season estimation using the proposed method. Prediction of the best time to view cherry blossoms can be achieved by combining post information and the best time to view estimation conditions. The accuracy of estimation is evaluated by comparing the estimated results with the correct prediction period, and by evaluating the recall rate and precision rate. The correct prediction period is defined as the period from the cherry blossom blooming date to the full blooming date, as observed by the Japan Meteorological Agency. The cherry blossom bloom date is the first day on which 5−6 or more cherry blossoms are in bloom on a sample tree [19]. The cherry blossom full bloom date represents the first day on which more than 80% of the cherry trees in the sample tree are open. First, we present results obtained for estimating the cherry blossom viewing time based on the number of posts obtained using Prophet as the algorithm for time-series prediction. Figures 1 and 2 respectively present results obtained for Tokyo and Kyoto for 2022: they are estimates of the cherry blossom viewing season based on the number of posts obtained from the time-series forecast.

Figures 3−4 present the results obtained from estimating the best time to view cherry blossoms in Tokyo and Kyoto in 2022, based on the number of posts obtained using DeepAR+ as the algorithm for time-series prediction.

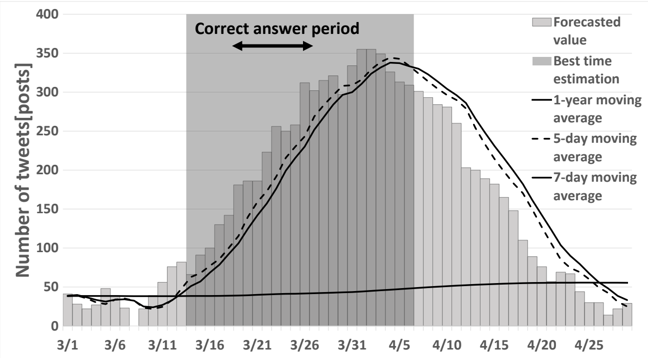


Figure 1: Tokyo Forecast for 2022 (Prophet).

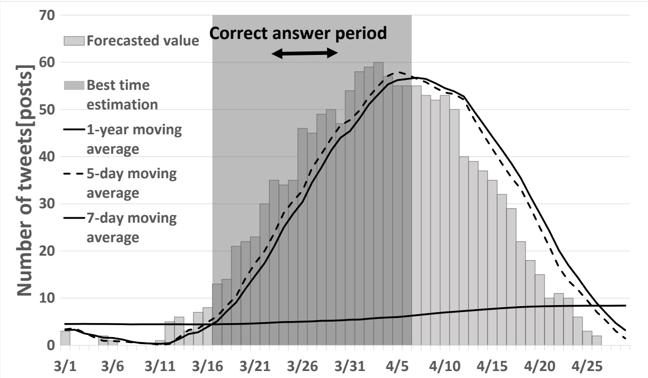


Figure 2: Kyoto Forecast for 2022 (Prophet).

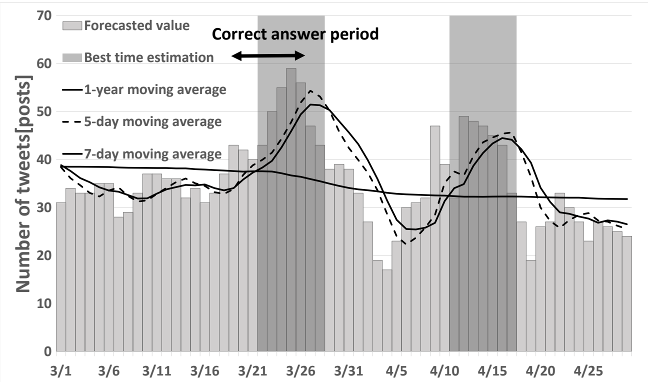


Figure 3: Tokyo Forecast for 2022 (DeepAR+).

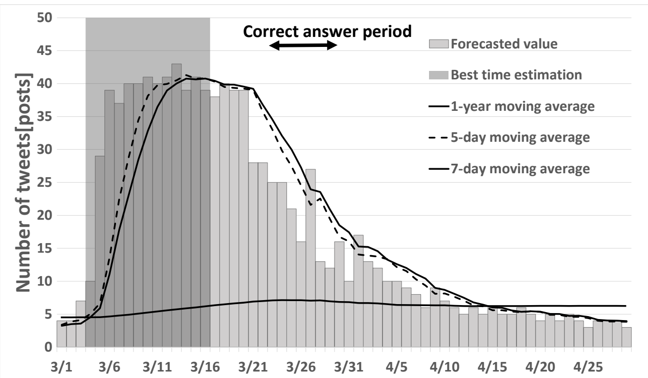


Figure 4: Kyoto Forecast for 2022 (DeepAR+).

The recall rate and precision rate, MSE, and *R* for estimating the cherry blossom viewing season using the number of posts obtained from the Prophet and DeepAR+ time series forecasts are shown respectively in Table 1.

From Figures 1 and 2, Prophet predicts a wide period of time, with the correct response period included in the estimation period. This finding might be attributable to the fact that the number of posts in the time-series prediction increases gradually, reaches a maximum value, and then gradually decreases, thereby making it easier to match the estimation conditions.

Figures 3 and 4 show that DeepAR+ has a distorted graph shape for the number of posts compared to Prophet. In Figure 3, the period differs from the correct period. There are two peaks. It is unlikely that the cherry blossom viewing season will return once it has subsided. In Figure 4, the prediction is made before the correct period. Table 1 shows that DeepAR+ has a lower recall rate score than Prophet has. MSE values are lower for Prophet compared than for DeepAR+ in Tokyo and Kyoto, indicating a smaller error between the actual number of posts and the predicted number of posts. In addition, the value of R is closer to 1 for Prophet than for DeepAR+, indicating strong correlation. However, the higher score for DeepAR+ compared to that for Prophet is the precision rate for Tokyo. Kyoto's precision rate was 0% because of the failure of the prediction. Prophet is the superior prediction method for estimating the best time to see cherry blossoms. Next, Figures 5 through 7 show the results of estimating the best time to see cherry blossoms using Proposal 1 from March 1, 2022 to April 29, 2022. Figure 5, Figure 6, and Figure 7 respectively portray the results obtained using Proposal Method 1 in Tokyo by factors of 0.3, 0.5, and 0.7.

Table 1: DeepAR+ and Prophet ratings

	Area	Recall	Precision	MSE	*R*
Prophet	Tokyo	100.0%	33.3%	8.90e-05	0.682
Deep AR+		75.0%	42.9%	1.32e-04	0.212
Prophet	Kyoto	100.0%	33.3%	8.46e-05	0.794
Deep AR+		0.0%	0.0%	4.96e-04	-0.289

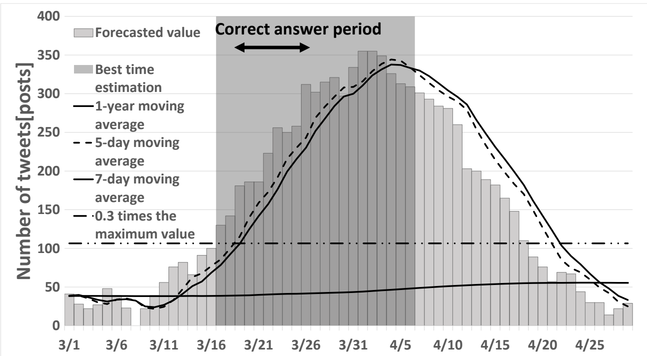


Figure 5: 0.3 times of proposed method 1 (Tokyo).

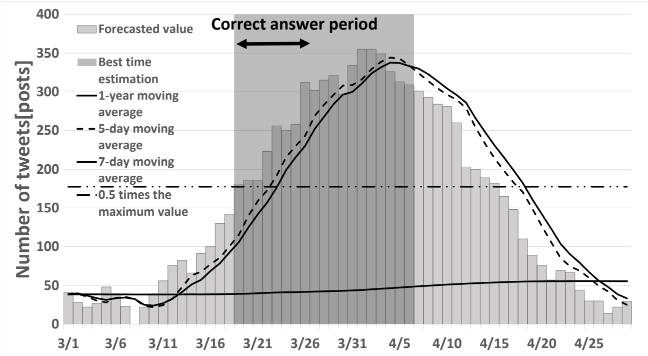


Figure 6: 0.5 times proposed method 1 (Tokyo).

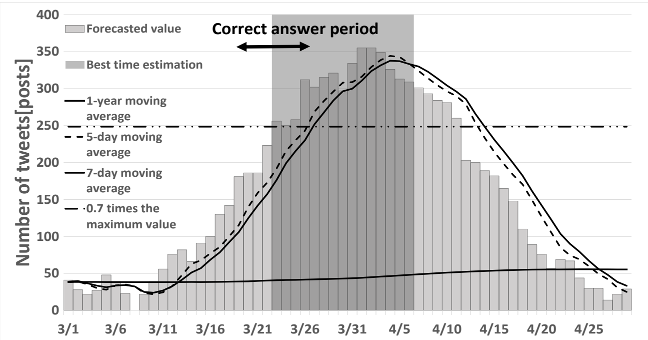


Figure 7: 0.7 times proposed method 1 (Tokyo).

Next, Figures 8−10 portray the results of the viewing period estimated using the proposed method 1 from March 1, 2022 to April 29, 2022. Figure 8, Figure 9, and Figure 10 respectively portray results obtained using proposed method 1 in Kyoto by factors of 0.3, 0.5, and 0.7.

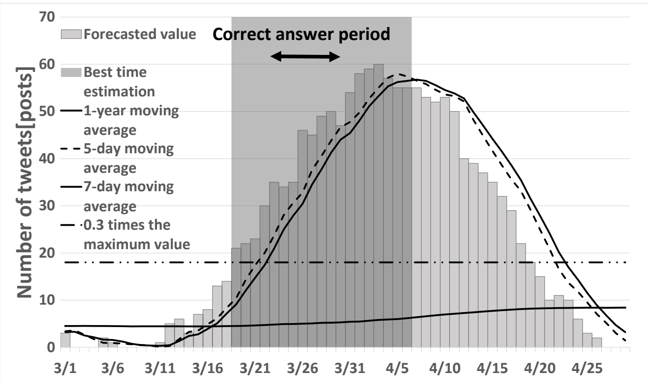


Figure 8: 0.3 times proposed method 1 (Kyoto).

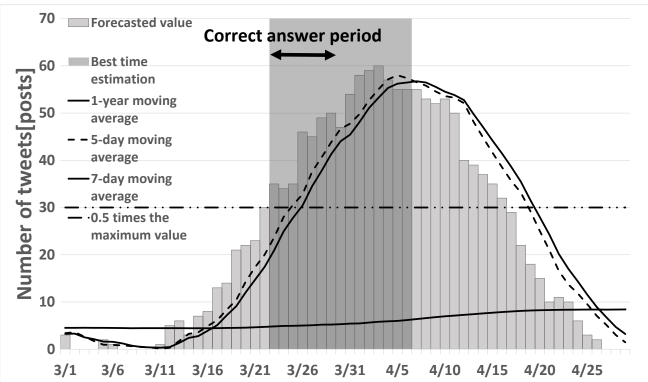


Figure 9: 0.5 times proposed method 1 (Kyoto).

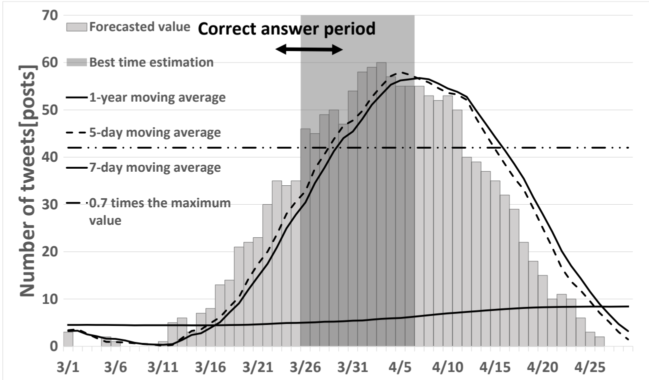


Figure 10: 0.7 times proposed method 1 (Kyoto).

Table 2 shows the recall rate and precision rate for estimating the cherry blossom viewing season using proposed method 1.

Figures 11−12 present results of the estimation of cherry blossom viewing time using proposed method 2 from March 1, 2022 through April 29, 2022. Table 3 presents the recall rate and precision rate of the estimation of cherry blossom viewing season using suggested method 2.

Table 2: Recall rate and Precision rate of viewing time estimation method using proposed method 1

	Area	Recall	Precision
Actual	Tokyo	100.0%	33.3%
0.3		100.0%	38.1%
0.5		100.0%	42.1%
0.7		62.5%	33.3%
Actual	Kyoto	100.0%	33.3%
0.3		100.0%	36.8%
0.5		100.0%	46.7%
0.7		71.4%	41.7%

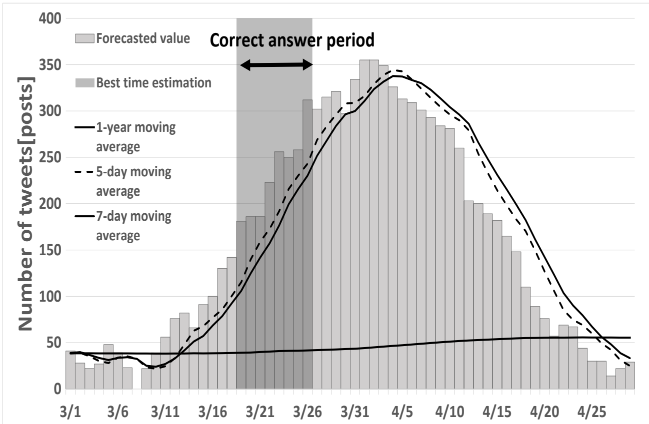


Figure 11: Proposed Method 2 (Tokyo).

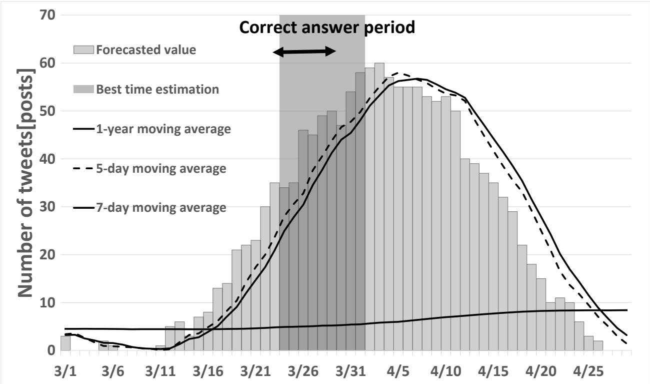


Figure 12: Proposed Method 2 (Kyoto).

Table 3: Recall rate and Precision rate of viewing time estimation method using proposed method 2

	Area	Recall	Precision
Actual	Tokyo	100.0%	33.3%
Proposed Method 2		100.0%	66.7%
Actual	Kyoto	100.0%	33.3%
Proposed Method 2		100.0%	70.0%

Describe the results of comparing Figures 5-12. The recall rate was higher when estimation was done at 0.3x, but the precision rate tended to be low. At 0.5x, both the recall rate and the precision rate were higher. At 0.7x the precision rate was higher, but the recall rate tended to be low. The reason for this result is thought to be that the estimation conditions for cherry blossom viewing at 0.3x were less stringent and the estimation period was wider, thereby yielding the results described above. The result is also thought to be that the estimation conditions for cherry blossom viewing at 0.7x were stringent and the estimation period was narrower, thereby yielding the results described above. The use of a factor of 0.5, which is intermediate between 0.3 and 0.7, is expected to improve the accuracy of the estimation results. Compared to the conventional method, proposed method 2 increased the precision rate by 33.1% in Tokyo and by 36.7% in Kyoto. Results demonstrated that changing the parameters used for estimating the cumulative ratios increased the accuracy. These results indicate that the proposed method can provide higher accuracy than the conventional method.

4.2 Discussion

Other researches that have conducted post forecasting have included forecasting word fads and hits [20][21]. Their common problem is the quantity and quality of data. This is thought to be due to the fact that they are dealing with data based mainly on human behavior. This has led to low forecasting accuracy. On the other hand, my research deals with posts based on natural phenomena. Unlike predicting human behavior, posts based on natural phenomena are periodic. Therefore, it may be said that we were able to make time-series predictions at a practical level.

CONCLUSION

Time-series forecasting using DeepAR+ resulted in distorted graphs and low accuracy. On the other hand, Prophet increased the number of posts in the time-series prediction gradually, reaching a maximum value and then gradually decreasing, which is considered to have increased the accuracy. These results indicate that the time-series forecasting method changes the prediction of the number of posts and affects the accuracy of the estimation of the cherry blossom viewing season.

As a result of time-series forecasting using Prophet, we were able to know the total number of posts for a certain period of time and the maximum value for that period. We proposed a new method for estimating the best time to view the cherry blossoms by adding a new condition for estimating the best time to view the cherry blossoms using the values that can be obtained by the prediction. After the time-series forecasting, the proposed method was able to refine the estimation of the best time to view the cherry blossoms and improve the recall rate and the precision rate. This is thought to be due to the effect of time-series forecasting on the estimation of cherry blossom viewing time.

In the future, we intend to further clarify the relation between time-series forecasts and estimation of cherry blossoms viewing time, and based on this, improve the accuracy of estimation of cherry blossom viewing time and expand the area covered.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 20K12081, 22K13776, and the Telecommunication Advancement Foundation Research Grant.

REFERENCES

[1] X, https://twitter.com/ (referred in June 2023)

[2] M. Endo, K. Mitomi, K. Saeki, Y. Ehara. M. Hirota, S. Ohno, and H. Ishikawa, Study of information provided by the best time to see estimation method of phenological observations using posts, Proceedings of the 12th Annual Conference on Japan Society for Tourism Informatics, Shizuoka, Japan, pp.47–60 (2016).

[3] M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Automatic detection of tourist spots and best-time estimation using social network services, International Workshop on Informatics 2020, pp.65-72 (2020).

[4] T. Horikawa, M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Estimating the best time to see cherry blossoms using SNS and time-series forecasting of post numbers using machine learning, International Workshop on Informatics 2021, pp.37-44 (2021).

[5] T. Horikawa, M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Estimating the Best Time to View Cherry Blossoms Using Time-Series Forecasting Method, Multidisciplinary Digital Publishing Institute, pp.418-431 (2022).

[6] J. Sato, Y. Takamori, M. Fujimoto, T. Horikawa, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Verification of the Best Time to See Cherry Blossoms Estimation Method using Time Series Prediction Method, IWIN2022-Proceedings, pp.95-101 (2022).
[7] E. S. Silva, H. Hassani, S. Heravi, and X. Huang, Forecasting tourism demand with denoised neural networks, Annals of Tourism Research, Vol.74, pp.134-154 (2019).
[8] A. F. Colladon, B. Guardabascio, and R. Innarella, Using social network and semantic analysis to analyze online travel forums and forecast tourism demand, Decision Support Systems, Vol.123, pp.113075 (2019).
[9] Y. Takamori, K. Terada, M. Endo, S. Ohno, and H. Ishikawa, Co-occurring Word Determination Used for Estimating Best Times for Viewing Cherry Blossoms, MMEDIA 2023 Proceedings, pp.1-6 (2023).
[10] S. Morishita, S. Maenaka, D. Nagata, et al., SakuraSensor: Quasi-Realtime Cherry-Lined Roads Detection through Participatory Video Sensing by Cars, Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 695-705 (2015)
[11] X Developers, https://dev.twitter.com/ (referred in June 2023)
[12] Agricultural Research Institute, Simple Reverse Geocoding Service, https://aginfo.cgk.affrc.go.jp/rgeocode/index.html.ja (referred in June 2023)
[13] AWS, https://aws.amazon.com/jp/console/ (referred in June 2023)
[14] Amazon Forecast, https://aws.amazon.com/jp/forecast/ (referred in June 2023)
[15] Prophet, https://docs.aws.amazon.com/ja_jp/forecast/latest/dg/aws-forecast-recipe-prophet.html (referred in June 2023)
[16] S. J. Taylor and B. Letham, Forecasting at Scale, The American Statistician, Vol.72, pp.37-45 (2018)
[17] D. Salinas, V. Flunkert, J.Gasthaus, and T. Januschowski, DeepAR: Probabilistic forecasting with autoregressive recurrent networks, International Journal of Forecasting, Vol.36, pp.1181-1191 (2020)
[18] DeepAR+, https://docs.aws.amazon.com/ja_jp/forecast/latest/dg/aws-forecast-recipe-deeparplus.html (referred in June 2023)
[19] Japan Meteorological Agency https://www.data.jma.go.jp/sakura/data/index.html (referred in June 2023)
[20] A. Khare, Q. He and R. Batta, Predicting gasoline shortage during disasters using social media, Or Spectrum, Vol.42, pp.693-726 (2020)
[21] PF Pai and CH Liu, Predicting vehicle sales by sentiment analysis of tweet data and stock market values, IEEE Access, Vol.6, pp.57655-57622 (2018)

Keynote Speech 2:

Prof. Dr. Hitoshi Aida

(Emeritus Professor
The University of Tokyo)
(Chair: Yoshia Saito)

A Compiler System Supporting Memory Shared by Heterogeneous Machines

Hitoshi AIDA

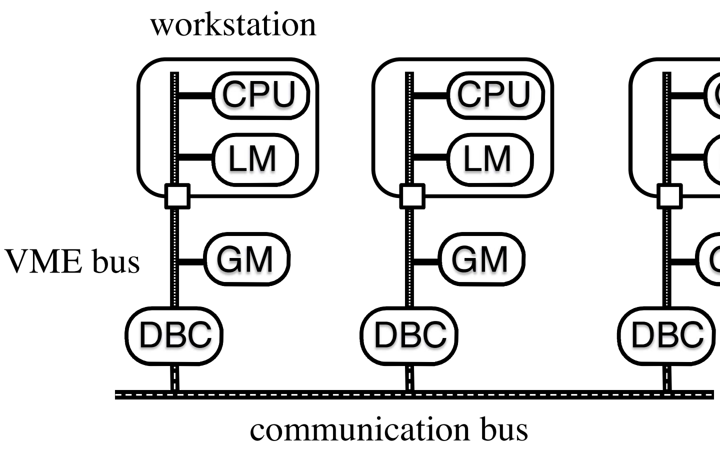
Emeritus Professor

The University of Tokyo

1

Motivation

- Sensor data dispatching system for an electrical power company
  - packet switching system with multicast capability
  - LAN-like mechanism: broadcast and filtering

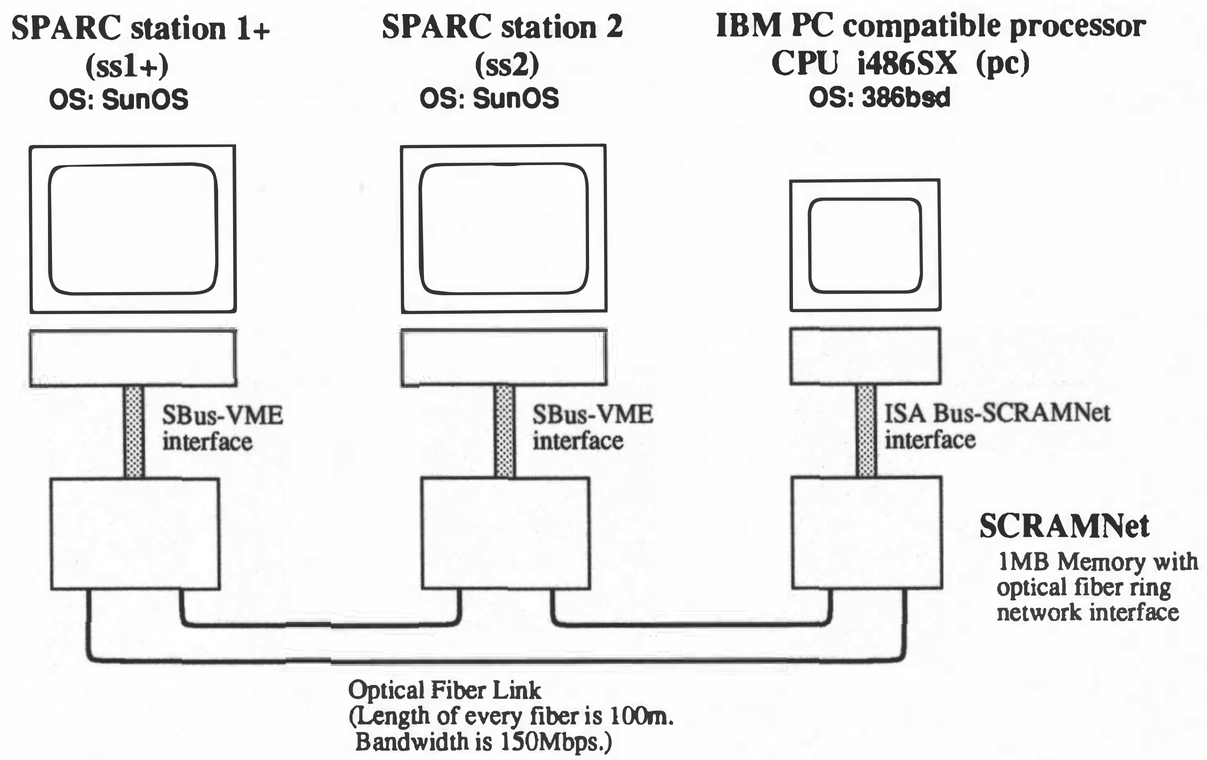


2

CPU used at that time

Sun Workstation	Macintosh
1982 Sun-1 68000 1985 Sun-3 68020 1987 Sun-4 SPARC Separate executable files for Sun-3 and Sun-4 1988 (Sun386i 80386)	1984 Macintosh 128K 68000 1987 Macintosh II 68020
	1994 PowerPC Classic environment(interpreter) Fat Binary 2005 intel X86 Rosetta(run time compilation) Universal Binary 2020 Apple silicon Rosetta 2

3



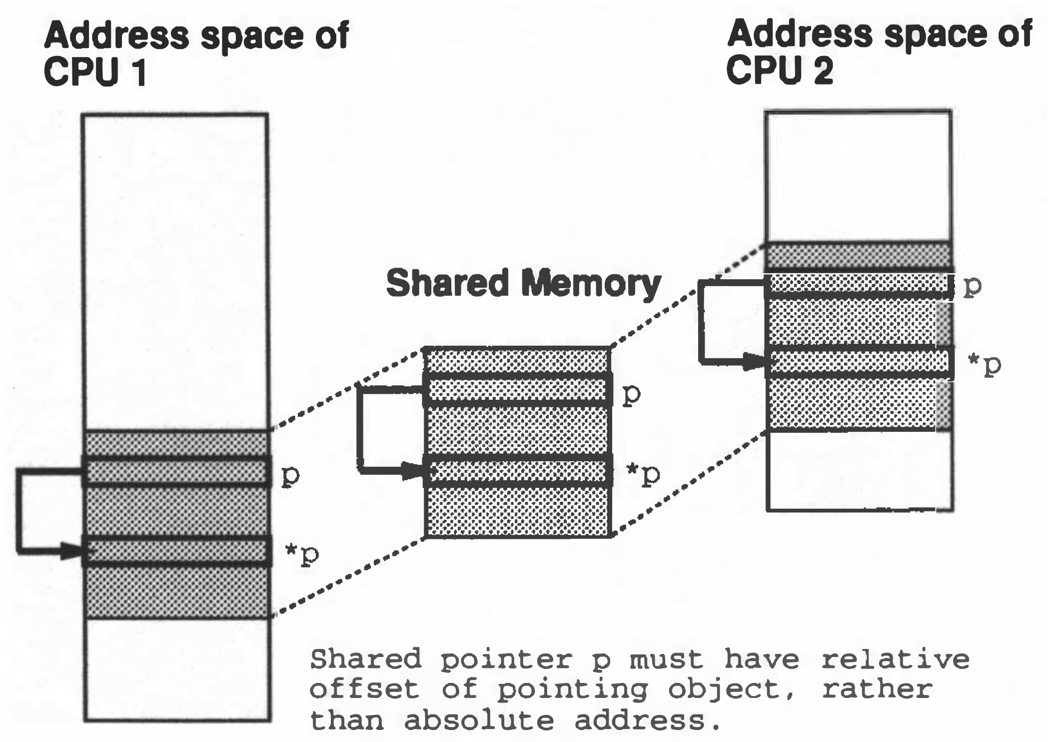
- SONY NEWS(MC68020/68030), OMRON LUNA-88K (MC88100) and others....

4

Difference between machines

- length and alignment of variables
  - `char`:8bits, `short`:16bits, `long`:32bits
  - `int`(and pointers):16bits or 32bits
  - replace `int` (to ex. `long`) to be same length
- byte order of variables
  - big-endian vs. little endian
  - should be translated each time
- address of shared variables
  - mapped to different address depending on OS
  - shared pointers should be expressed as the offset

5



6

How to absorb the difference?

- Using library functions
  - htonl(), ntohl(), …
- Let the compiler system hide the difference
  - which step of the compiler system?
  - how to instruct the compiler system?

7

```
long convert_endian(long data)
{
    return (((data << 24 )& 0x000000ff) |
            ((data << 8) & 0x00ff0000) |
            ((data >> 8) & 0x0000ff00) |
            ((data >> 24) & 0x000000ff));
}

long convert_endian(long data)
{
    union { long l; char c[4] } in, out;

    in.l = data;
    out.c[0] = in.c[3];
    out.c[1] = in.c[2];
    out.c[2] = in.c[1];
    out.c[3] = in.c[0];
    return out.l;
}
```

8

```
or %g0, 255, %l0
sll %i0, 24, %l1
sll %i0, 8, %l2
srl %i0, 8, %l3
srl %i0, 24, %l4
and %l4, %l0, %l4
sll %l0, 8, %l0
and %l3, %l0, %l3
or %l3, %l4, %l4
sll %l0, 8, %l0
and %l2, %l0, %l2
or %l2, %l4, %l4
sll %l0, 8, %l0
and %l1, %l0, %l1
or %l1, %l4, %o0
```

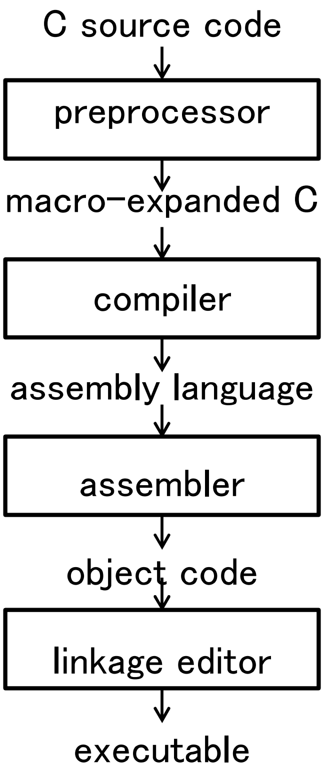
SPARC instruction

```
bswap %eax
```

i486 instruction

9

Compilation steps



10

Types in standard C

$$\underset{\substack{\uparrow\\ \text{type-qualifier}}}{\begin{pmatrix}\text{const}\\ \text{volatile}\end{pmatrix}}\underset{\substack{\uparrow\\ \text{storage-class-specifier}}}{\begin{pmatrix}\text{auto}\\ \text{register}\\ \text{static}\\ \text{extern}\end{pmatrix}}\underbrace{\begin{pmatrix}\text{signed}\\ \text{unsigned}\end{pmatrix}\begin{pmatrix}\text{short}\\ \text{long}\end{pmatrix}\begin{pmatrix}\text{char}\\ \text{int}\\ \text{float}\\ \text{double}\end{pmatrix}}_{\substack{\text{type-specifier}\\ \downarrow}}$$

defines semantics and allowed operations of the data e.g. bitwise logical operators may not be applied to `float` or `double`.

11

Type specifier

- defines the semantics and allowed operation of the data
- *C provides a number of operators for bit manipulation; these may not be applied to* ***`float`*** *or* ***`double`****.* (THE C PROGRAMMING LANGUAGE 2.9)

→operational type

12

Representation types

- Data with same meaning could be stored differently in the memory.
  - $-128 \leqq x \leqq 127$ → char/short/long/float/double
  - char x; → ASCII/EBCDIC/…
  - big-endian/little-endian

$$\begin{pmatrix}\text{const}\\ \text{volatile}\end{pmatrix}\begin{pmatrix}\text{auto}\\ \text{register}\\ \text{static}\\ \text{extern}\end{pmatrix}\begin{pmatrix}\text{ASCII}\\ \text{EBCDIC}\\ \text{big\_endian}\\ \text{little\_endian}\\ \vdots\end{pmatrix}\begin{pmatrix}\text{signed}\\ \text{unsigned}\end{pmatrix}\begin{pmatrix}\text{short}\\ \text{long}\end{pmatrix}\begin{pmatrix}\text{char}\\ \text{int}\\ \text{float}\\ \text{double}\end{pmatrix}$$

↑

representation types

13

Representation type of pointers

- Segmented address machines might distinguish between `near` pointers and `far` pointers.
- In shared memory environment, absolute addresses of the shared variables are different among machines, thus pointers should be expressed as either `offset` from the base address of shared memory or displacement from the pointers themselves (`relative`).

14

Representation type of pointers

pointer:

    * *type-qualifier-list*$_{opt}$

    * *type-qualifier-list*$_{opt}$ *pointer*

type-qualifier-list:

    type-qualifier

    type-qualifier-list type-qualifier

- Theoretically, whether pointers are represented by absolute/offset/relative values and which byte order those values are represented are independent.

15

Representation type `standard`

- To simplify, we use only one representation type `standard` for the data and pointers stored in the shared memory.

```
standard int *standard p;
```

- Storing `standard` data in the local memory is allowed.

16

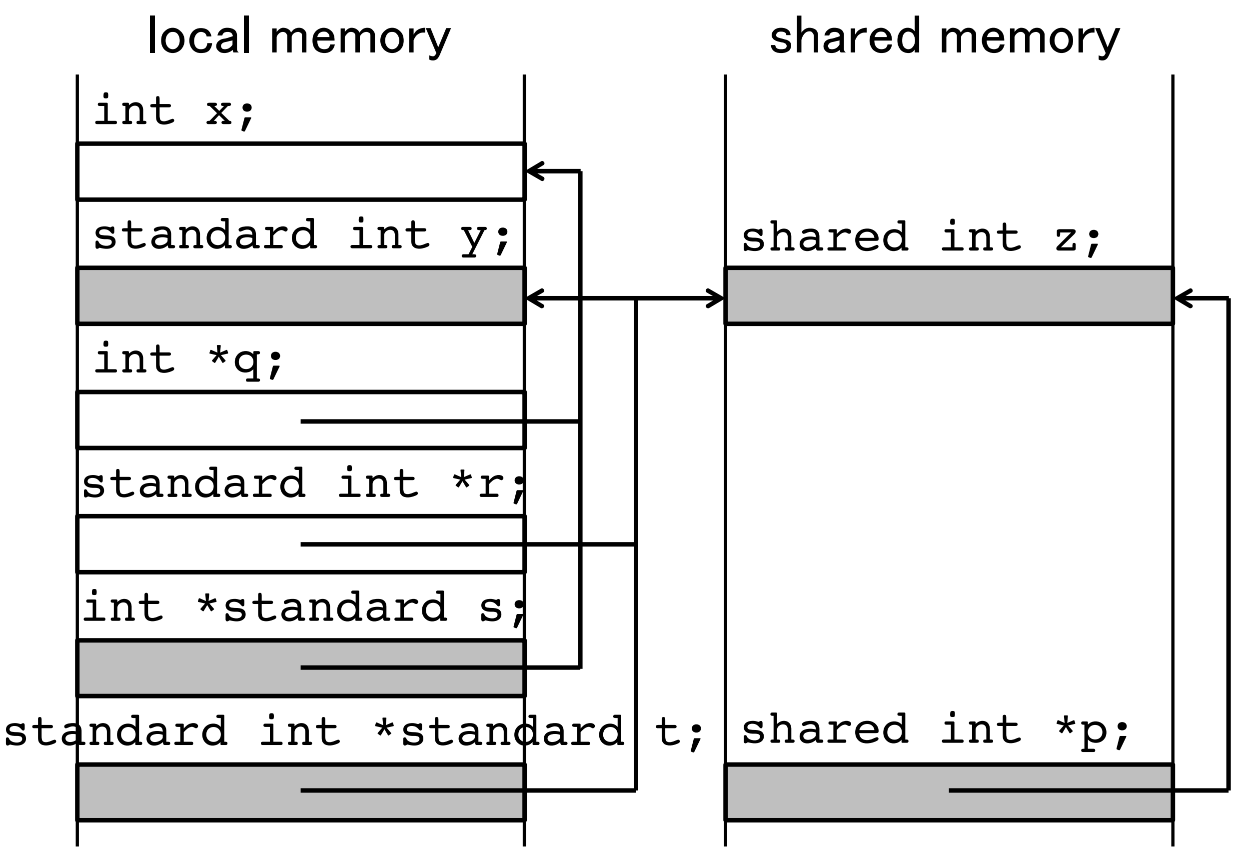
Storage-class `shared`

- Storage class `shared` is used to specify variables to be located in the shared memory.
- Contraints:
  - All the data and pointers stored in the shared memory should be `standard` representation.
  - All the pointers stored in the shared memory should point into the shared memory.

→Storage class `shared` implies `standard` representation

```
shared int *p;
→ shared standard int *standard p;
```

17



18

Implementation by preprocessor

```
standard int *standard p;
int x;

    x = *p;
```

```
void *_GM;  /* base address of shared */
long p;
int x;

    x = _stol(*(long *)(_GM + _stol(p)));
```

19

If standard is relative ....

```
standard int *standard p;
int x;

    x = *p;
```

```
long p;
int x;

    x = _stol(*(long *)
              ((void *)&p + _stol(p)));
```

20

Alignment of members in structures

```
standard struct {
    char c,
    int i, j;
    double d;
} x;
```

```
struct {
    char c, _dummy1[3];
    long i, j, _dummy2;
    double d;
} x;
```

21

Layout of `shared` variables

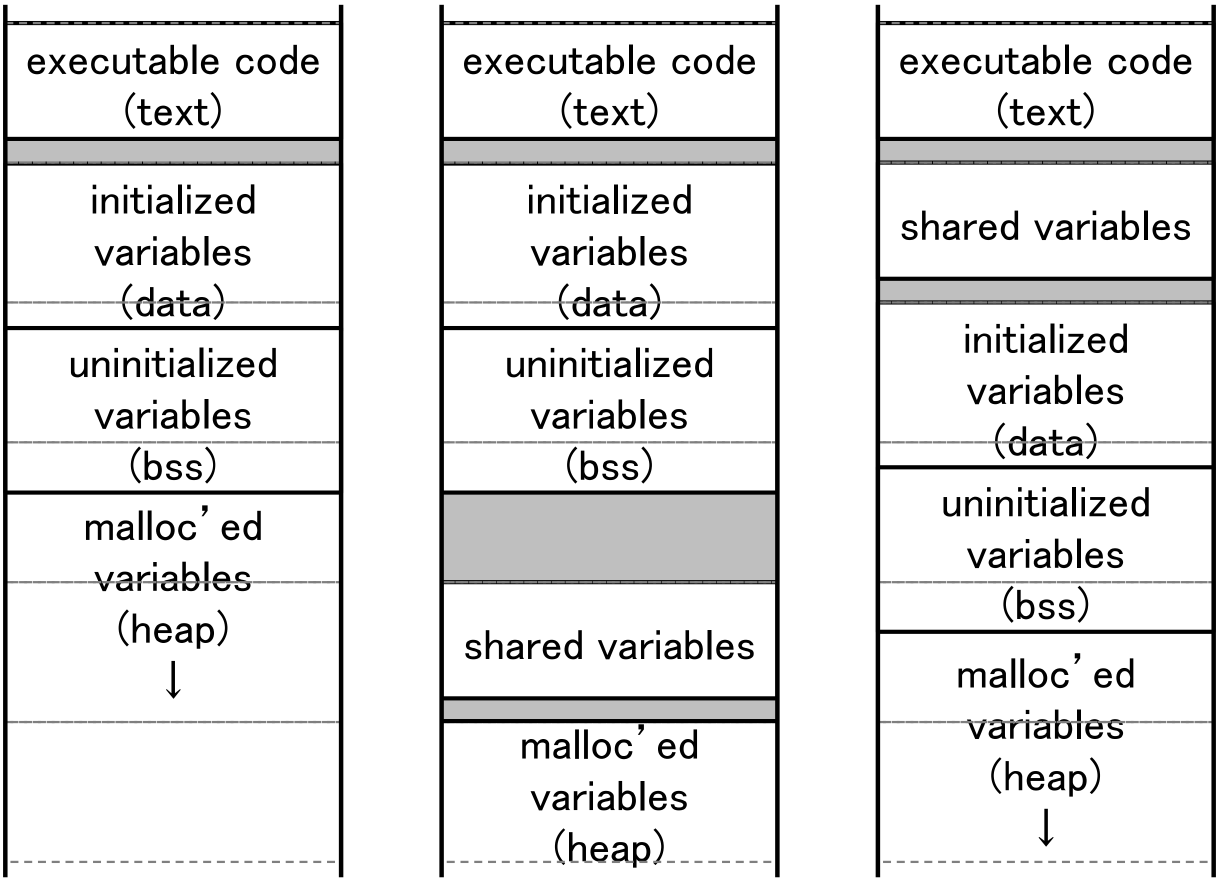
- The order of variables in the memory is not necessarily same as the order in the source file.
  - The order might be different in separately compiled files.
  - Compiler usually groups the variables by their alignment requirements.
  - Layout of variables are usually determined by the linkage editor.
- If separately compiled, layout of shared variables may become different among machines.

22

Layout of `shared` variables (continued)

- A compilation step is added to fix the layout of shared variables which satisfies the alignment requirement of all machines, and which is fed to the linkage editor.
- Shared memory is mapped to the right address during startup initialization.

23



24

Prototype implementation

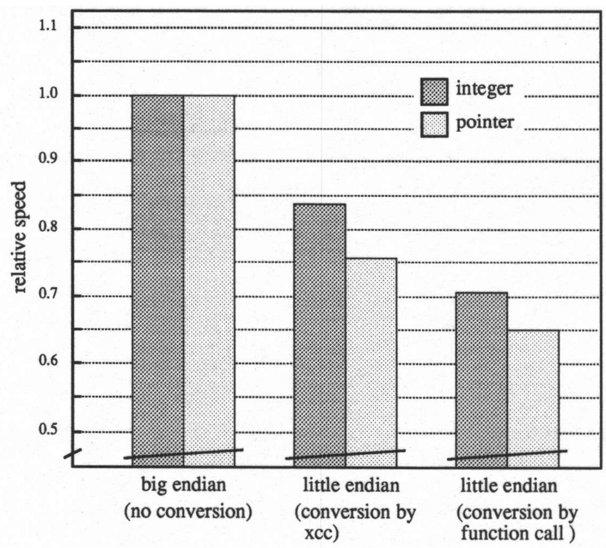
- Sun SPARCstation 1+, 2
  - SunOS 4.1.3
- CANON AXi/V 486
  - IBM PC compatible
  - 386bsd-0.1

	endian	**Alignment (in bytes)**				
		char	short	long	float	double
SPARCstation	big	1	2	4	4	8
i486pc	little	1	2	4	4	4

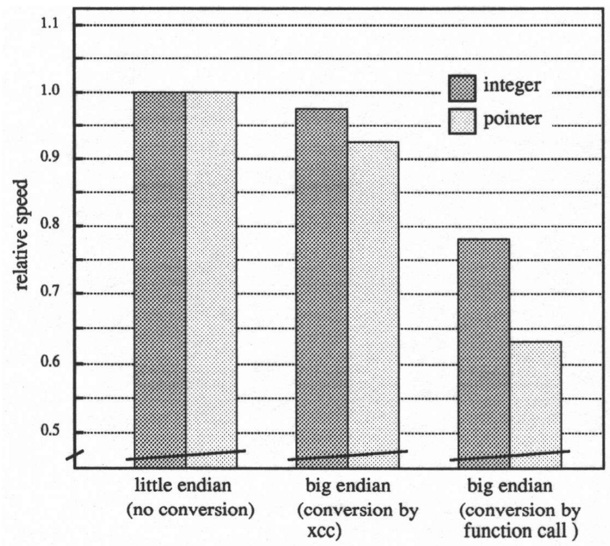
25

Overhead of endian conversion

- Batcher's Merge-Exchange Sort*



SPARCstation



i486pc

* Donald E. Knuth: *The Art of Computer Programming*, volume 3, Addison-Wesley, 1973. 26

Number of instructions in the innermost loop

	SPARCstation		i486pc	
	integer	pointer	integer	pointer
Total number of instructions	52	86	33	42
Number of instructions for endian conversion	30	62	2	8

27

Topics skipped

- detailed discussion on struct/union
- typedef
- bit field
- memory management of shared memory
- mutual exclusion

28

Summary

- A compiler system to support memory shared by heterogeneous machine is discussed
- Representation type `standard` and storage class `shared` are added to language C.
- A prototype compiler is implemented on SPARCstation(SunOS 4.1.3) and IBM compatible PC(386bsd-0.1).
- Overhead of endian conversion is small for the machines having byte-swap instruction.

29

Session 5:
Multimedia and Communication
(Chair: Yuichi Tokunaga)

Comparative of Experienced and Inexperienced People by Shoot Form Analysis from Free Throws of Basketball using 2-D Joint Position information and Slow Motion

TETSUTO MUKAI *, SHOTA ISHIGAKI **, and KEIICHI ABE *,
* Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
** Q'SFIX CO.,LTD.
mukai.tetsuto@gmail.com, abe@he.kanagawa-it.ac.jp

Abstract -In this study, we analyzed the shooting form of basketball players, using the OpenPose software to acquire two-dimensional (2D) joint position information without using markers. Slow-motion imaging proved to be effective for the posture analysis of the free throw motion. The shooting form of experienced players was compared with that of inexperienced players to obtain skeletal information that is useful for determining an optimum shooting form.
Keywords: Basketball, 2D Joint Position information, shoot form.

1. INTRODUCTION

Basketball is a very popular sport that has spread to more than 1 million people in Japan, including basketball enthusiasts. Shooting is essential to score points in basketball. However, the shooting ability in basketball varies from person to person, with some players having a high probability of scoring a goal and others having only a low probability. Even the players who have the same experience and practice significantly vary in shooting ability. Therefore, this study aims to improve the shooting rate of basketball players based on scientific findings.

In previous studies, there are technologies using OpenPose[1][2] and Kinect[3] for estimating the human skeleton from video images and for form analysis during sports. Kinect technology requires the use of a marker and a depth-sensor camera to acquire skeletal information, while OpenPose can acquire 2D Joint Position information from a camera image without using a marker or a depth-sensor camera. Therefore, we believe that OpenPose is a suitable analysis method for analyzing dynamic sports, such as basketball, and we focused on 2D Joint Position information acquisition using OpenPose technology in this study. This study utilized this two-dimensional skeletal information acquisition technology to analyze the optimal shooting form from the free-throw line in basketball by comparing the posture information of experienced and inexperienced basketball players to determine which form is most likely to score a goal. The effectiveness of slow-motion imaging of the subjects' shooting form was also evaluated, and the details are described in this paper.

2. RELATED WORKS

Microsoft's Kinect[3] is a representative technology for acquiring 2D Joint Position information. Kinect requires a marker to be attached to the body, making it difficult to use for sports such as basketball, which involve vigorous movements. By the way, OpenPose [1] is a technology for analyzing the human posture without using markers. OpenPose, developed by Zhe Ca et al., is a system that uses deep learning to estimate a person's skeleton. Using OpenPose, the 2D coordinates of 18 joint positions of a person can be obtained without using markers. This enables detailed analysis of the coordinates of characteristic points on the human body. In a previous study[2] of sport analysis using OpenPose, shooting scenes are detected in soccer videos and classified into three levels of skill from beginner to expert based on skeletal estimation. Another study[4] analyzed the spike form of a volleyball player by skeletal estimation and investigated the relationship between the spike form and athletic ability. There is a study [5] from the basketball free throw line using OpenPose. In this study, we analyzed the form when the knees were bent to the maximum during free throw shooting, the width of the knees at the time of release, and the height of the arms. We are proposing a technology to display in real time. However, it is not described in detail as an ideal form in this study. In this study, we investigated the ideal shooting form of basketball players by comparing the posture of skilled and inexperienced players based on 2D skeletal analysis of their shooting from the free throw line.

3. 2D JOINT POSITION INFORMATION ACQUISITION SYSTEM

Figure 3-1 shows an overview of the 2D Joint Position information acquisition system used in this paper, and Table 3-1 shows the specifications of the equipment and software used in this system. The system uses a smartphone and a high-precision camera to acquire video of the subject shooting from the free-throw line, and then performs posture analysis based on 2D skeletal information using OpenPose, which is open-source software. As shown in Figure 3-1, the hardware used in this study was a Jetson NANO equipped with both a GPU and CPU.

Table 3-1: Specifications of equipment and software used.

Equipment＆Software	Specification	Remarks
Jetson Nano A02	CPU：ARM CortexA57(4Core,1.43GHz) GPU:NVDIA Maxwell128Core (CUDA) Memory:4GB	
OS	Ubuntu 18.04.6LTS	
Software of 2-D Joint Position information	tf-pose-estimation(TensorFlow)	

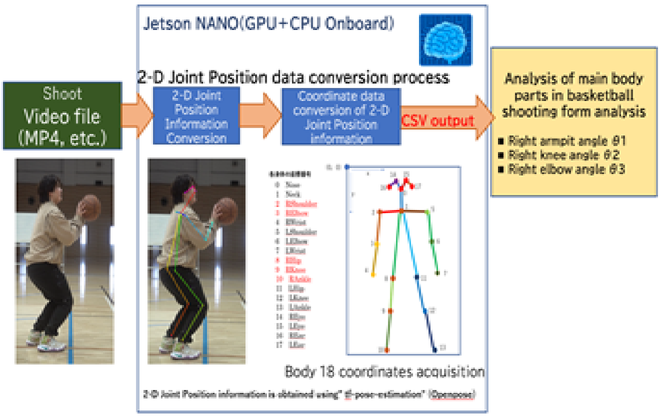


Figure 3-1:Overview of 2-D Joint Position acquisition system.

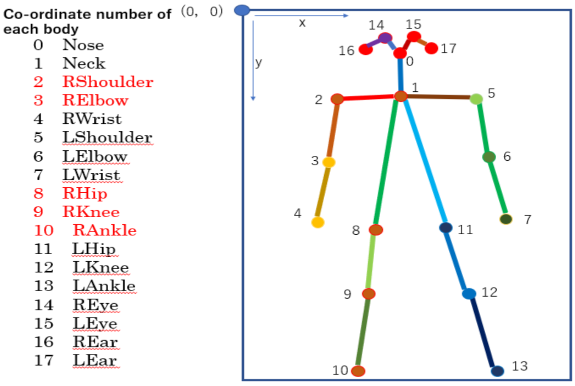


Figure3-2:Skeleton information analyzed by OpenPose.

The 2D joint position information was obtained in the Comma Separated Value (CSV) format from still images with 2D skeletal information analysis and 2D Joint Position information (X-axis and Y-axis coordinate data) for every two frames. The upper left corner of the still image is the origin of coordinates. In this study, to enable detailed analysis of the armpit angle, elbow angle, and knee angle, which are important for shooting from the free-throw line in basketball, we extracted and analyzed the data for the right shoulder, right elbow, right wrist, right hip, right knee, and right ankle from the body part data actually acquired.

4. Analysis Methods

4.1 Body part analysis

In this paper, we evaluated the shooting analysis from the free throw line by comparing the data of experienced basketball players with those of inexperienced players. Experienced players in this study are those who still belong to a basketball club or those who used to belong to a basketball club but have been away from basketball for a short period of time (less than one year), and all other players are treated as inexperienced players. In analyzing basketball shooting from the free throw line in this paper, we considered the armpit angle, knee angle, and elbow angle to be important factors, based on our analysis of explanations by experienced basketball players. As shown in Figure 4-1, we transformed the 2D skeletal information into these three parameters and analyzed the posture during shooting. In this paper, we calculated the right armpit angle θ1 from the two coordinate data of the right shoulder and the right elbow obtained by OpenPose, as shown in Figure 4-1. The right knee angle θ2 was calculated from the three coordinate data of the right hip, right knee, and right ankle. The right elbow angle θ3 was calculated from the three coordinate data of the right shoulder, right elbow, and right wrist using the inverse sine function and the cosine theorem. Figure 4-1 shows the calculation method and the coordinate numbers of the acquired 2D skeletal data. First, the angle θ1 of the right armpit with reference to the right shoulder was calculated using equations (4-1) and (4-2) according to the sign of H1, where H1 is the height of the right shoulder from the right elbow, and D1 is the horizontal distance between the right shoulder and the right elbow as shown in Figure 4-1. Since the origin in the OpenPose analysis image is at the upper left corner, the angle θ1 of the right armpit can be calculated using the positive and negative sign of the parameter H1.

When H1≥0,

$$\theta 1 = \sin^{-1}\left(\frac{D1}{L}\right) \qquad (4\text{-}1)$$

When H1<0,

$$\theta 1 = 180 - \sin^{-1}\left(\frac{D1}{L}\right) \qquad (4\text{-}2)$$

Next, the right elbow angle θ3 was calculated from the cosine theorem by equation (4-3), using the lengths L from the right shoulder to the right elbow, L2 from the right elbow to the right wrist, and L3 from the right shoulder to the right wrist.

$$\theta 3 = \cos^{-1}\left(\frac{L^2 + L2^2 - L3^2}{2 \cdot L \cdot L2}\right) \qquad (4\text{-}3)$$

Next, the right knee angle θ2 was calculated using the cosine theorem by equation (4-4), where, as shown in Figure.4-1, a is the length from the right hip to the right ankle, b is the length from the right knee to the right ankle, and c is the length from the right hip to the right knee.

$$\theta 2 = \cos^{-1}\left(\frac{b^2 + c2^2 - a^2}{2 \cdot b \cdot c}\right) \qquad (4\text{-}4)$$

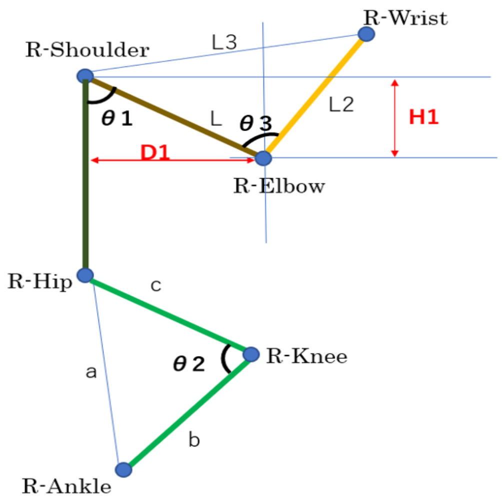


Figure 4-1:Method of calculating angles.

4.2 Methods of form analysis during shooting

According to the Basketball Medical Support Lab website[5], the important points in the flow of motion from the ball set to the release are the posture at the "motion break", which is between the set and release motions, as well as the posture at the "release". The definitions of "motion break" and "release" in this paper are as follows. First, the definition of "motion break" focuses on the knee extension motion during the shooting motion. Figure 4-2 shows a graph of the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 of an experienced player at the time of a successful shooting, obtained by OpenPose. The horizontal axis is the number of frames. As shown in Figure 4-2, the gray graph is the right knee angle θ2, and the point where the right knee fully extends again after the angle reaches the minimum value is the point where the shot is released. As shown in Figure 4-2, the point where the right knee begins to extend again after the right knee angle θ2 reaches its minimum angle is defined as the "motion break" in this paper. On the other hand, as shown in Figure 4-2, the "release" is defined as the point where the development movement of the right knee angle θ2 is near its maximum value and the ball is away from the hand. In the case of an NBA professional player active in the U.S., it is suggested that at the "motion break," the right arm is lifted to an armpit angle of 90°, the elbow rises further toward the release, and the elbow joint extension motion begins at the same time. In other words, the shoulder joint and elbow joint move simultaneously after the 90° angle of the armpit. Therefore, in this paper, we analyze the shooting form from the free throw line in detail by simultaneously calculating the subject's right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 using 2D Joint Position information acquisition during the "motion break" and "release" motions.

4.3 Demonstration experiment method

In our previous study [6], the subjects were six right-handed males, two of whom had basketball experience, and the remaining four were inexperienced. The shooting location was the Arena Gymnasium of the Kanagawa Institute of Technology, and the shooting video was taken from the free throw line of the basketball. We used an iPhone8puls camera (specifications: 4k, 60fps), and the camera was set up at 45 degrees from the subject.

4.4 Discussion of previous studies

Figure 4-3 shows an image of 2D Joint Position information acquired as a result of the analysis of the past study by using the commercially available Vision Pose from NEXT-SYSTEM. In the previous study[6] , detection errors often occurred during skeletal information acquisition in the release motion, making the analysis difficult. Figure 4-4 shows the image in which errors occurred during skeletal information acquisition, and Table 4-1 shows the values of the errors in the analysis of 2D Joint Position information acquisition during the motion break and release motion. The image in Figure 4-4 is the image at the time of release. It can be seen that the skeletal detection points of the right elbow and right wrist are displaced from the right arm in the foreground of the image and cannot be detected correctly. Table 4-1 shows that the detection errors are larger at the angle of the right armpit and especially during the period of the rapid movement from the motion break to the release. This is thought to be due to the fact that the arm swinging motion becomes faster when releasing the ball, causing the video image to be blurred during the analysis. In the previous study[6] , the distance between the subject and the camera was too far, which made the subject appear to be smaller, and the blurring of the analyzed images during fast arm movements such as during release is considered to have caused a detection error when acquiring skeletal information. For this reason, the authors considered that it would be better to place the camera closer to the subject and shoot from right beside the subject. Therefore, based on the results of the previous studies, we devised the following improvement measures (1) to (4) to reduce the detection error in skeletal information acquisition, and verified them in a demonstration experiment.

1) Shoot at a close distance so that the subject appears larger
2) Shoot from right beside the subject instead of 45 degrees from the subject
3) Change from normal speed to slow motion shooting
4) Increased frame rate from 60 fps to 120 fps

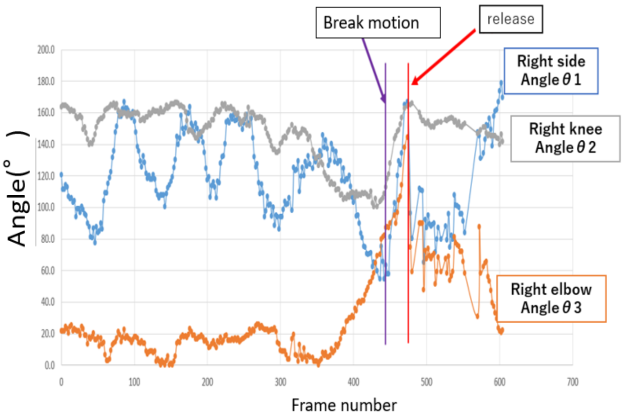


Figure 4-2: The right armpit angle θ1, right knee angle θ2 and right elbow angle θ3 of an experienced player.

Table4-1:Analysis error in 2-D joint position information.

Item	(1) Break motion			(2) Release		
	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)
Measured Value(°)	88.6	119	75	132	153	154
True Value(°)	75	119	75	115	155	140
Error(°)	13.6	0	0	17	-2	14
Error rate(%)	18	0	0	14.8	-1.3	10

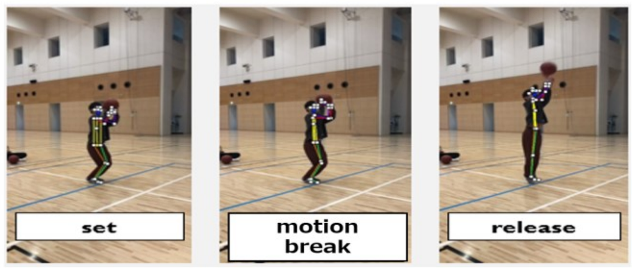


Figure 4-3:2-D joint position information image[6].

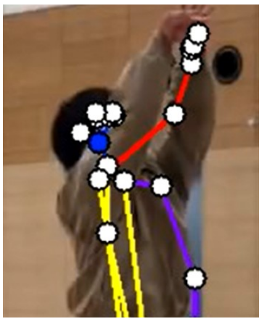


Figure 4-4:Error in 2-D joint position information.

Table5-1 Specifications of Camera

Camera	Number of effective pixels in photography	Number of pixels for video recording, etc
α7III(ILCE-7M3K/SONY)	24.2 megapixels	2.0 megapixels (1920x1080 pixels), 60 fps Slow-motion shooting 120 fps.

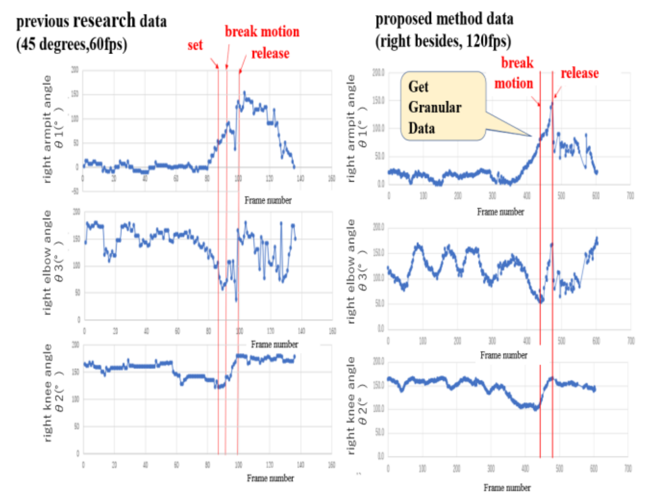


Figure 5-1: Comparison of data obtained with the proposed method and those obtained in previous study.

5. DEMOSTRATION EXPERIMENT

We conducted a demonstration experiment by implementing improvement measures based on the results of the previous studies. In this demonstration experiment, the video shooting was conducted twice, with four subjects (including one with basketball experience) for the first shooting and eight subjects (including five with basketball experience) for the second shooting. All subjects were right-handed males. The location of the shooting was inside the Arena Gymnasium of Kanagawa Institute of Technology. A high-performance camera was used to shoot a slow-motion video of the subject throwing a ball from the center of the free throw line, 4.225 m from the basketball goal. The camera was set up on the extension of the free throw line at a distance where the subject shooting the ball was seen on the entire screen of the camera, and the camera was fixed on a tripod so that the subject was right beside the camera. The camera used for this shooting was SONY α 7 Ⅲ. The specifications of this camera are shown in Table 5-1. This high-performance camera has a slow-motion shooting function, which was used in this study.

5.1 Evaluation of the proposed methodology

In this section, we evaluate the effectiveness of acquiring 2D Joint Position information using OpenPose on videos captured with the proposed method (improvement measures) described in section 4.4. Figure 5-1 shows a comparison between the proposed method and previous research data. The horizontal axis of the graphs in Figure 5-1 is the number of frames, and the vertical axis shows the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 acquired with 2D Joint Position information. The video shot using the method proposed in this paper was shot in slow motion at a frame rate of 120 fps, more than two times that of the previous studies, so the number of acquired frames increased as shown in Figure 5-1. In addition, the graphs of the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 obtained with the 2D Joint Position information are very detailed, and in particular, the data for each angle during the release can be obtained accurately without missing any data. This confirmed the effectiveness of the proposed method, and we proceeded with the analysis under these conditions.

5.2 Motion analysis of experienced and inexperienced basketball players

In this study, the motion analysis of the experienced and inexperienced players was conducted using the videos of their successful shooting. Figure 5-2 shows still images of experienced and inexperienced players shooting from the free throw line, converted into 2D Joint Position information by OpenPose. The images in Figure 5-2 are divided into three parts: set, motion break, and release, with the number of frames indicated for each part. Table 5-2 shows the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 calculated from the 2D Joint Position information data acquired during the "motion break" and "release" phases of the successful shooting by experienced and inexperienced players. The data of NBA player in Table 5-2 were manually calculated from the video images on the Internet [5]. As shown in Table 5-2, the armpit angle θ1 of Experienced Player 1 at the motion break is 87°, which is close to the NBA player 90°, while the value of θ1 is as high as 116° for Inexperienced Player 1. On the other hand, the armpit angle θ1 of Experienced Player 1 at the time of release was 144°, which was close to the NBA player's 140°, while that of Inexperienced Player 1was lower at 125°. The results show that there is a difference in the values of θ1 and θ2 between the experienced and inexperienced players at the "motion break" and "release," and that the experienced player's posture and movement are closer to those of the NBA player in the values of θ1 and θ2. Figure 5-3 shows the graphs of the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 against the number of frames on the horizontal axis for the experienced player. Red vertical lines indicate the number of frames for the set, motion break, and release motion. The graph of the knee angle θ2 in Figure 5-3 shows that at the motion break, the right knee angle θ2 once reaches its minimum value and the knee begins to extend from the minimum value, while the release motion starts at the motion break and ends with the right armpit angle θ1 being at its maximum value.

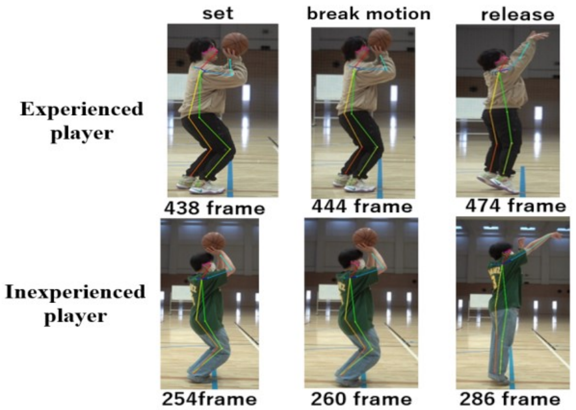


Figure5-2:Shooting motion of experienced and inexperienced players.

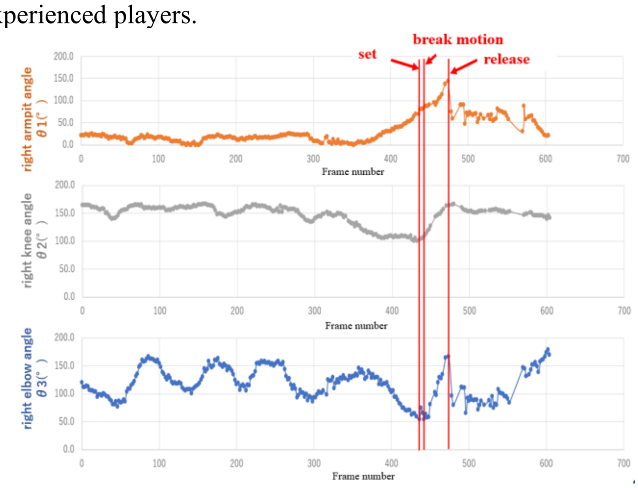


Figure 5-3: Armpit angle θ1, knee angle θ2 and elbow angle θ3 of Experienced Player 1 plotted against the number of frames.

Table 5-2:Comparison between experienced and inexperienced players.

Subject	(1) Break motion				(2) Release				Remarks
	Frame No	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Frame No	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	
Experienced player 1	444	87	113	63	474	144	164	167	Shota
Inexperienced player1	260	116	106	78.5	286	125	176	173	Koku
NBA player		90.0		60.0		140		180	

5.3 Comparative evaluation of Experienced and Inexperienced Players in successful shooting

Next, a comparative evaluation of experienced and inexperienced subjects at the time of the successful shooting was conducted. Four subjects, one experienced and three inexperienced, participated in this study. Table 5-3 shows the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 calculated from the two-dimensional skeletal data obtained at the "motion break" and "release" during the successful shooting. Data from NBA players were also included for reference. Δθ1, Δθ2, and Δθ3 in Table 5-3 indicate the amount of change from the motion break to the release. As shown in Table 5-3, the right armpit angle θ1 showed a large difference between the experienced and inexperienced players. Experienced Player 1's right armpit angle θ1 at the motion break was around 90°, and the right armpit angle θ1 at the time of release was around 140°. These values are comparable to those of the NBA player. By the way, the values of θ1 and θ3 of Inexperienced Player 3 are similar to those of Experienced Player 1. However, as shown in Figure 5-4, when we look at the series of shooting motions, Experienced Player 1 shoots with his arms closed and his left hand on the ball, while Inexperienced Player 3 shoots with his arms open and in the same position. Inexperienced Player 3's shooting form is different from that of Experienced Player 1. However, when Inexperienced Player 3 shoots with both hands, the values of θ1 and θ3 happened to be comparable to those of Experienced Player 1. Next, the amount of change in the right elbow angle θ3 from the "motion break" to the "release" is examined. The Δθ3 of Experienced Player 1 is 104°, which is close to that of the NBA player (120°). The Δθ3 values of Inexperienced Players except for Inexperienced Player 3 were less than 100°. Table 5-3 also shows the number of frames in which the right armpit angle θ1 was 90±5° during the transition from "motion break" to "release" in the movie of a successful shot, and the presence or absence of the angle. It was confirmed that the number of frames in which the right armpit angle θ1 was 90±5° was more than one for all subjects except for Inexperienced Player 1. Table 5-3 shows the success rates of the experienced and inexperienced shooters. The success rate was calculated as the percentage of shots that resulted in a goal in five throws, but when all of the five throws resulted in failure, the player was asked to throw the ball again until it went in. The results show that the success rate of Experienced Player 1 was 75%, while the other inexperienced players had a success rate of less than 20%.

Table5-3: Comparison of θ1, θ2 and θ3 between experienced and inexperienced players in successful shooting

Subject	(1) Break motion			(2) Release			Difference			The number of frames in which the right armpit angle θ1 was 90±5° during the transition from (1) to (2)	Remarks
	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Δθ1 (°)	Δθ2 (°)	Δθ3 (°)		
Experienced player 1	87	113	63	144	164	167	57	51	104	6	Shota
Inexperienced player1	116	106	78.5	125	176	173	9	70	95	0	Koku
Inexperienced player2	42	119	101	123	179	176	81	78	75	1	hama
Inexperienced player3	93	101	31	106	165	145	13	64	114	3	sano
NBA player	90.0		60.0	140		180	50.0	-	120	-	

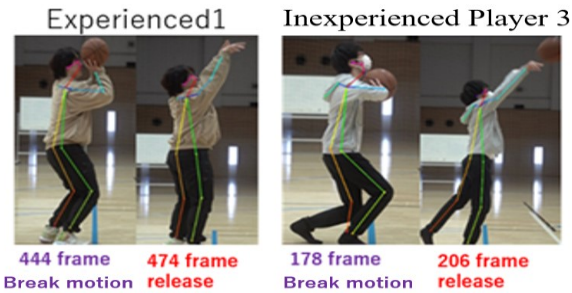


Figure5-4:Comparison of shooting motion between Experienced Player 1 and Inexperienced Player 3.

Table5-4 Comparison of θ1, θ2 and θ3 between experienced and inexperienced players in failed shooting.

Subject	(1) Break motion			(2) Release			Difference			The number of frames in which the right armpit angle θ1 was 90±5° during the transition from (1) to (2)	Remarks
	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Right armpit angle θ1(°)	Right knee angle θ2(°)	Right elbow angle θ3(°)	Δθ1 (°)	Δθ2 (°)	Δθ3 (°)		
Experienced player 1	83	127	99	129	157	169	46	30	70	0	Shota
Inexperienced player1	116	126	91	131	177	175	15	51	84	0	Koku
Inexperienced player2	60	118	97	123	176	180	63	58	83	0	hama
Inexperienced player3	101	118	103	155	178	152	54	60	49	0	sano
NBA player	90.0		60.0	140		180	50.0	-	120	-	

5.4 Comparative evaluation of experienced and inexperienced shooters in failed shoots

Next, Table 5-4 shows a comparison of the experienced and inexperienced players at the time of shooting failure. Table 5-4 shows the results of calculating the right armpit angle θ1, right knee angle θ2, and right elbow angle θ3 based on the 2D Joint Position information at the "motion break" and "release.” The right armpit angle θ1 was not 90±5° for both experienced and inexperienced shooters when they failed to shoot, and the number of frames in which the right armpit angle θ1 was 90±5° during the transition from the "motion break" to the "release" was zero. Compared with Table 5-3, which shows the results of experienced and inexperienced players at the time of successful shooting, it is considered that for successful shooting from the free throw line, the right armpit angle θ1 should be around 90° at the "motion break" and the number of frames during which the right armpit angle θ1 reaches 90±5° until the "release" should be at least one. The right elbow angle θ3 is also noteworthy. In the difference Δθ3 during the transition from the "motion break" to the "release", Experienced Player 1 had a value of 104° for the successful shooting and 70° for the unsuccessful shooting. Therefore, in shooting from the three-throw line, the amount of movement of the right arm during the transition period from the "motion break" to the "release" is also considered to be important.

6. CONSIDERATION

Chapter 6 discusses efficient shooting from the free throw line, based on the findings of this study, with reference to the "Basketball Instruction Manual"[9] published by the Japan Basketball Association. This "Basketball Instruction Manual" describes shooting in basketball as follows: "The area that the ball can pass through the hoop depends on the angle of incidence of the ball in relation to the hoop. If the angle of incidence is large, the ball will enter the hoop more easily, but the loop needs to be higher, and the distance will be less. If the angle of incidence is too small, the area through which the ball can pass through the hoop is reduced, making it unsuitable. To obtain the ideal loop and distance, it is best to shoot at a projection angle of 45 to 50 degrees.” In other words, this "Basketball Instruction Manual" [8] states that the angle of throw of the ball at the release should be 45° to 50°, which is considered to be the ideal angle. Therefore, if the ball is thrown from a right armpit angle of 90° in the "motion break," the right armpit angle θ1 at the release is considered to be 135° to 140°. This figure is supported by the fact that the right armpit angle θ1 at the release for the NBA players shown in Table 5-4 is 140°.

7. SUMMARY

In this paper, we used OpenPose to perform a two-dimensional posture analysis of the basketball shooting form of experienced and inexperienced players from the free throw line. This time, a posture analysis was performed using videos taken in slow motion. As a result, it was confirmed that experienced players were characterized by the armpit angle θ1, which is within around 90° at the motion break and around 135° at the release.

Also, for inexperienced players trying to efficiently acquire the shooting form of experienced players, it is good to be aware of the armpit angles θ1, knee angle θ2 and elbow angle θ3 and of shooting with a good coordination between the upper and lower body forces. In addition, when the two-dimensional skeletal information posture was evaluated with and without slow motion, it was confirmed that the slow-motion shooting particularly improves the accuracy of the 2D Joint Position information at the time of release.

Based on the findings of this research, we would like to prove through demonstration experiments that when inexperienced players are taught effective practice methods, their form becomes similar to that of experienced players, and that their shooting rate increases.

REFERENCES

[1] Zhe Cao, Tomas Simon, Shih-EnWei, Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", <https://arxivorg/abs/1611.08050>, (Referred:2022-04-29).

[2] N. Takuma, Y. Ikuko, H. Hitoshi, " Classification of soccer skills using OpenPose", The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020, Vol. 3M5-GS-12-05, pp. 1-4, 2020.

[3] Y. Iida, S. Uchino," Verification of accuracy of marker type joint angle measurement during free throw-like movement using KINECT v2 sensor ", Vol. 4, The Japan Journal of Basketball Studies, pp. 55-63, (2018).

[4] Kira Yusei, "Study of the Correlation between Success of Volleyball Spike and Exercise Ability using Skeletal Estimation", Vol. ISS-SP-029, p. 188, (2020).

[5] M. Nakai,Y. Tsunoda, H. Hayashi, H. Murakoshi, "Prediction of Basketball Free Throw Shooting by OpenPose", New Frontiers in Artificial Intelligence, JSAI-is AI 2018 Workshops, (2018). <https://doi.org/ 10.1007/978-3-030-316051_31>, (Referred:2022-07-20).

[6] Basketball Medical Support Lab , ” Japan national basketball team! Shot analysis of #8 Rui Hachimura!”, <https://www.baske-medical-support.com>, (Referred:2022-04-29).

[7] Shota Ishigaki, Keiichi Abe," Comparative of Experienced and Inexperienced People by Shoot Form Analysis from Free Throws of Basketball using 2D Joint Position information " IPSJ Symposium series DICOMO Multimedia, Distributed, Cooperative, and Mobile Symposium, (DICOMO 2022), 8E-3, pp. 1659-1663, (2022).

[8] NBA all-time free-throw success rate rankings The number one free-throw success rate all-time in a season is Steph Curry (as of 29 October 2020 locally) , <https://ikeple.com/nba-historical-records007/>, (Referred:2023-02).

[9] Japan Basketball Association, "Basketball Instructional Manual”, pp. 63-65, (2006).

Disaster Victim Impact Analysis System using a Communication Failure Emulator

Kei Hiroi†, Akihito Kohiga‡, Sho Fukaya††, and Yoichi Shinoda‡

†Disaster Prevention Research Institute, Kyoto University, Japan
‡Japan Advanced Institute of Science and Technology, Japan
††Suwa University of Science, Japan
hiroi@dimsis.dpri.kyoto-u.ac.jp

Abstract - This paper analyzes the prediction of communication failures and their impact on information communication during a flood event using a communication failure emulator. Damage estimates related to communication in the event of a disaster are indicated by the number of disrupted lines. However, this method does not allow disaster victims to know when they will not be able to use the system and what kind of situation they will be in. In these days when information communication has become commonplace, such a situation could have a tremendous impact on the damage. Therefore, we propose a system that emulates communication failures in the event of a disaster and analyzes the impact from the user's perspective. We evaluate the prediction accuracy of the emulation of communication failures and discuss its impact.

Keywords: flood disaster, communication failure, federation system

1 Introduction

1.1 Background

The number of disasters and the number of human casualties is steadily increasing in the world due to the complex intertwining of changes in the global climate system and global warming caused by greenhouse gases[1] [2]. Urban populations are expected to reach over 6 billion by 2050, with 68% of that population concentrated [3] in urban areas at risk of disaster. In particular, flood disasters have the highest number of human casualties, with about 3.39 million affected in 2018 [4], and the projections show further increases in these numbers.

Smart city technology is expected to realize resilient cities. Smart city technology uses Internet of Things (IoT) sensors to collect various data on cities, such as traffic flow, people flow, and weather, and uses AI to determine and predict the supply and demand of people living in cities in real time [5].The aim of smart cities is to create resilient cities that make our lives in cities with high disaster risks better.

Advances in cyber-physical systems, the (IoT), cloud computing, and other software technologies are actively contributing to the development of smart cities. Smart cities offer the basic capabilities of integrating sensors, actuators, and other devices into the physical environment of the city through a variety of technologies and computational techniques. Especially, simulation technology, one of the basic technologies of smart cities, playing an important role in service demand forecasting in physical spaces and prediction analysis in virtual spaces.

1.2 Motivation

Simulation is a traditional fundamental technology for imitating real-world events. Simulation techniques can be broadly classified into physical simulations (i.e., that physically model actual objects or physical phenomena) and logic simulations (i.e., that approximate state changes with continuous or discrete modeling). Case study application of simulation technologies are contributing to the study of production process management and productivity improvement measures in the supply chain, as well as to the upgrading of agricultural management systems based on weather data in the field of agriculture. These manufacturing and agriculture smart city objectives can be said to be relatively simple simulations and system optimizations. For example, in agriculture, the change in weather can be represented by a physical phenomenon model, and the growth of agricultural products can be represented by a real model given conditions. In other words, these objects and phenomenon can be described by physical simulation.

By contrast, in the field of disaster researches, various types of simulations are required to explain complicated phenomena. The magnitude of disaster damage is determined by the disaster phenomena and the vulnerability of elements, such as urban spatial structure and human behavior, against them. Therefore, in order to extend the smart city to the field of disaster prevention and contribute to realizing resilient cities, it is necessary to handle various physical and social phenomena. The problem is that the simulations of each phenomenon to be handled are based on different models. Disaster phenomena are described by a physical model, and human behavior is described by a discrete model such as a multi-agent model. Conventional simulation technology has evolved such as to be specialized for one certain model. That is, these different models and technologies are represented by different simulations and systems. Therefore, in order to realize a cyber-physical system for disaster prevention, dealing with analysis results obtained from different models and technologies in an integrated manner is a major challenge.

2 Related Work

2.1 Literature Review

One of the simulation techniques is physical simulation for reproducing disaster phenomena; for flood damage, this means a flood analysis simulator that estimates the occurrence of flood damage and its expansion process. This is a simulator that derives the flood water value in several-meter grid units by numerical analysis using input data such as precipitation, land use, building coverage, sewer pipes, and manholes. In addition to flood damage, many simulations have been developed to accurately reproduce various disaster phenomena such as tsunami [6] and wildfire [7]. In these simulations, each generation mechanism is modeled by a physical formula and then reproduced on a computer; these simulations are actually used for risk estimation and damage estimation in many regions.

Another approach is a social simulation that models the behaviors that can be taken by evacuees. An evacuation simulator reproduces evacuation behavior in various situations, such as when a disaster event occurs. There are several approaches to behavior modeling, including an empirical approach and an agent-based approach [8]. A number of models have been developed, including an evacuation model that dynamically selects destinations [9], an indoor evacuation model that considers collisions between evacuees [10], an evacuation model that considers evacuation behavior affected from nearby evacuees [11], and a crowd evacuation model [12]. Nguyen et al. [13] used an evacuation simulation in disaster phenomena, considering the behavior of smoke diffusion during a fire and providing directions to evacuees, and discussed its use in formulating efficient evacuation plans.

A new approach in the concept of the cyber physical system is a simulation that models physical elements of the real space onto the virtual space. Virtual Singapore produces an urban space model that integrates terrain information, buildings, and social infrastructure information throughout Singapore onto a virtual space [14][15]. Furthermore, a simulation environment is being constructed by integrating various real-time data (traffic information, car / people location information, etc.) into that urban space model. Behind these technologies is the evolution of technologies developed for various applications, such as IoT. Cyber-physical systems extend the IoT concept further to facilitate the interaction of the smart city's cyber and physical spaces. [16] attempted to detect landslide using IoT concept, [17]-[19] adopted their cyber physical system to disaster responce. In disaster prevention, construction of an elaborate urban structure model means that disaster phenomena can be reproduced with high accuracy, which can improve the usefulness of feedback to the physical spaces. This is because the parameters of the physical model used for the physical simulation of disaster phenomena are defined based on the data of the urban space model. Moreover, the collection of real-time data by the IoT has the potential to improve the accuracy of both disaster phenomena simulation and social simulation.

The goal of handling different models and technologies in an integrated manner is achieved by a strategy to develop a large framework that incorporates these simulations and systems as a single function. Integrated Emergency Response Framework (iERF) [20] is such an integrated framework. This was proposed as a framework to integrate various tools used in emergencies such as simulation and visualization. High Level Architecture (HLA) is an approach to standardize different simulators/systems for distributed simulation, used when developing a simulation for a larger purpose by combining several simulations [21]. HLA develops a distributed simulation by providing various services such as data distribution and time synchronization through middleware. Dahmann et al. [22] have achieved the interoperability of various simulations by examining the specifications of a common technical architecture for simulations using HLA. The architecture was implemented as a prototype to interoperate multiple simulators, including evacuation, information provision systems, and emergency operations, in an earthquake disaster [23]. In addition, a distributed simulation platform has been constructed for evacuation simulation in the event of fire disasters caused by earthquakes [24].

2.2 Limitations and Contributions

Although the abovementioned integration strategies provide theoretical support for coordination between different models to represent complex real spaces, there remain several issues to be improved. First, it is necessary to realize information communication environment in virtual space. Considering the feedback of the prediction results, actual communication needs to be simulated, but this has rarely been discussed in disaster response terms. Second, while ensuring scalability, the cost of system improvement for synchronization with new simulations and systems should be reduced. This is due to reducing obstacles to quickly adopting new observation systems and prediction technologies created by feedback from the virtual space of a cyber-physical system. Third, social simulation depends on information and decision making. In order to reproduce human behavior in a disaster situation more precisely, a cyber-physical system needs to incorporate accessibility to information as a parameter of the simulation.

A federation strategy using the information bus was proposed in our previous paper [25] to solve the problems above. The contributions of the federation strategy are summarized as follows:

- Federation technology enables the synchronization and sequential sharing of computations between simulations / systems directly into a cyber-physical system. This can connect information technology in a virtual space and create a more elaborate physical space. In addition, the components themselves can be exchanged between virtual space and physical space via federation, and application to various test beds can be expected. We named this progressive simulation.

- Our platform also supports the behavior of computer networks that affect the accessibility of information, and it evaluates the damage mitigation effects of that behaviour. We define the technique that imitates the behavior of computer networks as computer emulation. Computer emulation automatically verifies computer network outages or destruction due to disasters and can more accurately reproduce physical space events caused by complex factors.

This paper proposes a system that uses this federation strategy to create scenarios that include the impact of communication failures in floods.

3 Proposal of Disaster Victim Impact Analysis System

This paper aims to propose a system that uses the federation strategy to create scenarios that include the impact of communication failures during floods. It is well known that communication failures occur during disasters such as floods. Communication failures during disasters include power outages, broken communication systems, and traffic congestion caused by concentrated traffic. In particular, communication failures caused by power outages and physical damage to communication systems occur frequently during floods. In recent years, 70% of floods have resulted in failures due to damage to communication systems. Such physical failures are considered to have a significant impact on the disaster victims because their recovery requires physical system relocation.

Particularly in disasters such as floods, where damage spreads gradually, it is relevant to the accessibility of information to the disaster victims before and during the disaster. Therefore, it is important to consider strategies including communication failures in providing and obtaining information. However, preliminary disaster assessments are based on large estimates, such as "X million network access lines will be out of service." In preparation for network outages, there have been studies on advance deployment of systems that enable early recovery and alternative networks such as delay tolerant network [26]-[28]. On the contrary, we believe that it is essential to consider how the occurrence of such communication failures affects the access to information for disaster victims during a disaster. In the 2020 floods in Kumamoto, Japan, communication failures occurred as the rainstorms progressed, making it difficult to obtain information. Optical fibers at bridge piers were flushed away by the rising river water, and communication was damaged even in areas where no flooding had occurred.

Human behavior during disasters depends largely on what kind of information affected people have obtained. In the current society, which relies heavily on information and communication technology, we should assess what kind of situation the victims' access to information will be in and how it will affect them by simulating the situation in advance as a strategy for disaster response. Thus, in this paper, we develop a system to create scenarios that simulate communication failures by using our federation strategy, which allows multiple simulators to be implemented in a cooperative manner. The system incorporates multiple factors such as the impact of natural disasters, human behavior, the impact of disasters on communication networks, and information services as a flood analysis simulator, evacuation simulator, communication failure emulator, and evacuation routes information system, respectively, and creates scenarios of human impact by emphasizing the dependencies among these factors(Figure 1). Using this system, we confirm that it is possible to generate multiple scenarios (optimistic scenarios and worse scenarios) related to communication failures.

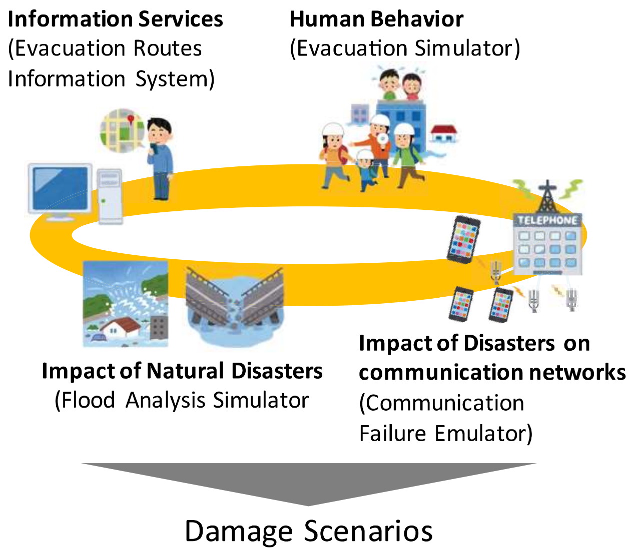


Figure 1: System Concept

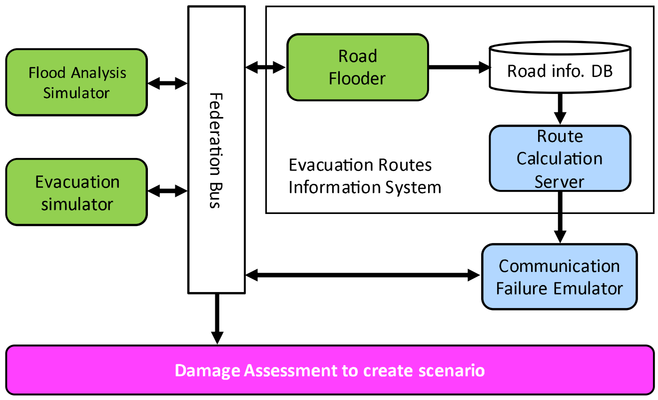


Figure 2: System Configuration

4 System Configuration

4.1 Overview

This paper simulates and analyzes human impacts from disaster phenomena simultaneously with information service technology. The incorporation of multiple factors into a simulation model requires complex and massive development efforts, although our federation strategy allows for the cooperative implementation of these multiple simulations, It can also be expected to be used as a mechanism for evaluating the accuracy and usefulness of the information service in a disaster situation.

Our system consists of multiple simulators operating in a cooperative manner. Four simulators / systems (flood analysis simulator, evacuation simulator, evacuation route service system, and communication failure emulator) exchange data with each other considering their operation timing, and progressively compute their own simulations using data from other simulators(Figure 2).

4.2 Federation Bus

The federation bus is responsible for their timing control and data exchange. The federation bus consists of a federation manager, which is responsible for time-synchronized progress control, and an MQTT broker, which is responsible for Publish-Subscribe message exchange. It creates scenarios by changing the alert timing, cooperative operation of these multiple simulators / systems, and analyzing the behavior of disaster victims and their access to information as calculated by the evacuation simulator.

4.3 Flood Analysis Simulator

Our federation strategy can allow us to use a conventional flood analysis simulator. However, in this paper, we use data from API[1], which performs flood calculations, due to the difficulty of obtaining the input data for the flood calculations. For the scenario area, Hitoyoshi City, Kumamoto Prefecture, we regard the flood data as calculated at specific time intervals and connect it to a federation bus. Road Flooder progressively calculates the impassable roads, according to the flooded value from the flood data. The calculation result is transmitted to the road management DB through the federation platform, and the road information is updated. Road information management DB stores the road information and the shelter information, which are used for the evacuation route calculation in the target area.

4.4 Evacuation Simulator

Evacuation simulator is developed using a multi-agent simulator. At the beginning of simulation, evacuation simulator generates the evacuee agents, according to the population distribution which is set in the map. Determination of the each behavior can be set the parameters for each agent as evacuation characteristics. In other words, we can set the evacuation behavioral characteristics of each agent according to age, disaster experience, and ability to access information. Besides, the shelter where to which each agent evacuates is set as the initial value for each agent. The total number of evacuee agents can be set on the evacuation simulator.

Evacuation simulator calculates the behavior of the evacuee agents, depending on the flooded value calculated from the flood analysis simulator. The evacuee agents determine their behaviors on the basis of their own location, walking speed, or information. Evacuee agents have their own smartphones. When local agency sends the evacuation information, evacuee agents move in accordance with their own evacuation behavioral characteristics. For example, an agent who has better evacuation behavior characteristics starts to evacuate to the shelter immediately after receiving information. We configure several types of the evacuation behavioral characteristics; the agent who start the evacuation several minutes after receiving the information, or who start if their surrounding agents start to move.

Evacuee agent consults the evacuation routes information system to determine the safe evacuation route after their decision, then starts to move to the shelter. During their evacuation, they confirm the flooded value $v_{t,k}$ of the surrounding grids at each simulation step. If $v_{t,k} = 0$, they move to the next grid on their route. If $V_{t,k}$ exceeds the threshold, their evacuation routes are flooded; in consequence, they re-ask the other safe route to information system. This threshold is used to set the flooded value as impassable for a given road for each agent. The evacuation simulator calculates the amount of damage and data access status in a progressive manner using data from evacuation agents.

[1]https://suiboumap.gsi.go.jp/

4.5 Evacuation Routes Information System

Before the evacuation, each evacuee agent searches the safest evacuation route to the shelter using their smartphone. Evacuation routes information system search and provide the evacuation route . The system consists of road flooder, road information management DB, and route calculation server.

When the evacuee agents search for an evacuation route, the route calculation server excludes impassable roads calculated by the road flooder and computes them according to the current location of the agent and of the nearest shelter. By inputting the location of the evacuee agents and the destination node ID, the route calculation server determines a route by weighted Dijkstra method in consideration of distance and cost, and it returns a list of nodes on the route. In the evacuation simulation, the location of the evacuee agents and the ID of the destination node are transmitted by MQTT via the federation bus. Upon receiving the data, the smartphone emulator receives a list of nodes from the route calculation server using the REST API and transmits the list to the evacuation simulator using MQTT.

The route calculation server generates a weighted undirected graph $G_N = (N, N_{next})$ with the distance between nodes as the weight. Simultaneously, a set node N_{next} is selected from a table of links connecting to the set of nodes N. The initial value of the cost of the graph G_N is configured to the distance between the nodes; however the ID of the nodes and links that have become impassable due to the flood expansion can be updated to the graph G_N via the road information management DB. After the river floods, if the grid on a agents' route becomes impassable, the node ID and link ID included in the grid are transmitted to the road management DB, and the cost of the corresponding nodes and links of the graph G_N is updated. After the cost is updated, when the agent inquires of the evacuation route via their smartphone, the route calculation server returns the detour route.

4.6 Communication Failure Emulator

Our communication failure emulator is an emulator that simulates communication failures related to smartphones in floods. The emulator consists of a mobile base station, a mobile relay station, a wired communication line between the mobile base station and the relay station, and a power supply to the mobile base station and the relay station. The mobile base station transmits radio waves within a defined range and is used for information access by evacuee agents within the range. If the flood water level rises in the flood analysis simulator and the mobile base station itself or the power supply function of the mobile base station is flooded, the mobile base station loses its communication function. The wired communication lines between the mobile base station and the relay stations are connected by the shortest path. When there is damage to the communication line due to flood flow, the mobile base station to which it is connected loses its communication function. The mobile base stations and the relay stations are supplied with electric power, and when a power outage occurs due to flood, the power supply is stopped and the communication function is disconnected at the same time. Channels and alternative means of power supply to the mobile base station are not considered in this paper.

The communication failure emulator sends the availability status of the mobile base station via MQTT over the federation bus. Depending on the location of the evacuee agent, it determines which base station to use and transmits the availability status of that base station back to the evacuation simulator. If the mobile base station is active, evacuee agents in the evacuee simulator will have access to routes to the shelter and alerts. If the base station is not available, the evacuee agents will not be able to access information. With the above steps, the communication failure emulator simulates information access in the event of a flood.

4.7 Damage Scenario

The previous section describes a system that uses a federation strategy to simulates and analyzes human impacts from disaster phenomena simultaneously with information service technology. We use this system to generate scenarios of communication failures during floods. The flood analysis simulator uses data of the Kuma River flooding that occurred on July 4, 2020 in Hitoyoshi City, Kumamoto Prefecture, Japan. In this flood, the water level exceeded the flood hazard level at six water level observation stations, and overflows from the river occurred at 34 locations along the main Kuma River. Regarding telecommunications, 88, 70, and 111 mobile base stations were out of service in Kumamoto Prefecture by NTTdocomo, KDDI, and Softbank, respectively.

The evacuation simulator generates evacuee agents according to the population distribution in Hitoyoshi City and sets them to move to the nearest evacuation center. For the communication failure emulator, we visually survey the actual mobile base station locations within Hitoyoshi City and set them up in the emulator in the same arrangement as the actual base and relay stations. The communication area of the base station is set to a fixed radius. When the mobile base stations in the communication failure emulator are in operation, our evacuation routes information system presents the shortest route to the shelter, excluding flooded roads, in response to access by the evacuee agent in the evacuation simulator. We create scenarios in this system by changing alerts timing as trigger the start of evacuation and the time of the simulation. Using the scenarios, we analyze the possibility of creating multiple scenarios and the effects of communication failures from the viewpoint of disaster victims.

5 Conclusion

This paper proposed a system that uses the federation strategy to create scenarios that include the impact of communication failures during floods. Communication failures caused by power outages and physical damage to communication systems occur frequently during floods. Physical failures are considered to have a significant impact on the disaster victims because their recovery requires physical system relocation. Particularly in disasters such as floods, where damage spreads gradually, it is relevant to the accessibility of information to the disaster victims before and during the disaster. Therefore, it is important to consider strategies including communication failures in providing and obtaining information.

This paper focued how the occurrence of such communication failures affects the access to information for disaster victims during a disaster.. We develop a system to create scenarios that simulate communication failures by using our federation strategy, which allows multiple simulators to be implemented in a cooperative manner. Using this system, we confirmed that it is possible to generate multiple scenarios (optimistic scenarios and worse scenarios) related to communication failures.

Our system can simulate and analyzeshuman impacts from disaster phenomena simultaneously with information service technology It consists of multiple simulators operating in a cooperative manner. Four simulators / systems (flood analysis simulator, evacuation simulator, evacuation route service system, and communication failure emulator) exchange data with each other considering their operation timing, and progressively compute their own simulations using data from other simulators. The incorporation of multiple factors into a simulation model requires complex and massive development efforts, although our federation strategy allows for the cooperative implementation of these multiple simulations, It can also be expected to be used as a mechanism for evaluating the accuracy and usefulness of the information service in a disaster situation.

After we developed four simulators / systems, we generate scenarios of communication failures during floods. Creating scenarios, we used data of the Kuma River flooding that occurred on July 4, 2020 in Hitoyoshi City, Kumamoto Prefecture, Japan. We set actual mobile base station locations within Hitoyoshi City up in our communication failure emulator. Four simulators/systems were simultaneously operated to simulate flood water value by the flood analysis simulator and to simulate the accessibility of information by the evacuee agents generated by the evacuation simulator. Using the scenarios, we confirmed the possibility of creating multiple scenarios and analyzed the effects of communication failures from the viewpoint of disaster victims.

Acknowledgement

This work was supported by JST, PRESTO Grant Number JPMJPR2036, and the commissioned research(No. 05401) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

[1] Intergovernmental Panel on Climate Change Fifth Assessment Report (AR5), available at: https://www.ipcc.ch/report/ar5/, Accessed 19/2/2019.

[2] Milly, P. Christopher D., Wetherald, Richard T., Dunne, K. A., Delworth, Thomas L, Increasing Risk of Great Floods in a Changing Climate, Nature, Nature Publishing Group, Vol.415, No.6871, page514, 2002.

[3] United Nations Department of Economic and Social Affairs, 2018 Revision of World Urbanization Prospects, 2018.

[4] United Nations Office for Disaster Risk Reduction, 2018:Extreme weather events affected 60m people, available at: https://www.undrr.org/news/2018extreme-weather-events-affected-60m-people, Accessed 19/2/2019.

[5] Yamagata, Y., Maruyama, H., Urban Resilience -A Transformative Approach-, Springer International Publishing, 2016.

[6] Makinoshima, F., Imamura, I., Abe, Y., Enhancing a Tsunami Evacuation Simulation for a Multi-scenario Analysis using Parallel Computing, Simulation Modelling Practice and Theory, Vol.83, pp.36–50, 2018.

[7] Beloglazov, A., Almashor, M., Abebe, E., Richter, J., and Charles, B. S. K., Simulation of Wildfire Evacuation with Dynamic Factors and Model Composition, Simulation Modelling Practice and Theory, Vol.60, pp.144–159, 2016.

[8] Ouyang, M., Review on Modeling and Simulation of Interdependent Critical Infrastructure Systems, Reliability engineering & System safety, Vol.121, pp.43–60, Elsevier, 2014.

[9] Pel, A. J., Bliemer, M. CJ., Hoogendoorn, S. P., A Review on Travel Behaviour Modelling in Dynamic Traffic Simulation Models for Evacuations, Transportation, Vol.39, No.1, pp.97–123, Springer, 2012.

[10] Chen, L., Tang, T.Q., Huang, H.J., Wu, J.J., Song, Z., Modeling Pedestrian Flow Accounting for Collision Avoidance during Evacuation, Simulation Modelling Practice and Theory, Vol.82, pp.1–11, 2018.

[11] Li, W., Li, Y., Yu. P., Gong. J., Shen, S., The Trace Model: A Model for Simulation of the Tracing Process during Evacuations in Complex Route Environments, Simulation Modelling Practice and Theory, Vol.60, pp.108–121, 2016.

[12] Shendarkar, A., Vasudevan, K., Lee, S., Son, Y.J., Crowd Simulation for Emergency Response using BDI Agents based on Immersive Virtual Reality, Simulation Modelling Practice and Theory, Vol.16, No.9, pp.1415–1429, Elsevier, 2008.

[13] Nguyen, M. H., Ho, T. V., Zucker, J.D., Integration of Smoke Effect and Blind Evacuation Strategy (SEBES) within Fire Evacuation Simulation, Simulation Modelling Practice and Theory, Vol.36, pp.44–59, Elsevier, 2013

[14] Soon, K. H., Khoo, V. H. S., CityGML Modelling for Singapore 3D National Mapping. In Proceedings og The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol.42, page37, 2017

[15] Ignatius, M., Wong, N. H., Martin, M., Chen, S. Virtual Singapore Integration with Energy Simulation and Canopy Modelling for Climate Assessment. In IOP Conference Series: Earth and Environmental Science, Vol.294, No.1, pp.012018, IOP Publishing, 2019.

[16] Liu, Z., Tsuda, T., Watanabe, H., Ryuo, S., Iwasawa, N. Data Driven Cyber-physical System for Landslide Detection, Mobile Networks and Applications, Vol.24(3), pp.991–1002, 2019.

[17] Estrela, V. V., Hemanth, J., Saotome, O., Grata, E. G., Izario, D. R., Emergency Response Cyber-physical System for Flood Prevention with Sustainable Electronics, In Brazilian Technology Symposium, Springer, pp.319-328, 2017.

[18] Mariappan, R., Reddy, P. N., Wu, C, Cyber Physical System using Intelligent Wireless Sensor Actuator Networks for Disaster Recovery, In Proceedings of 2015 International Conference on Computational Intelligence and Communication Networks (CICN), pp.95–99, IEEE, 2015.

[19] Nishiyama, H., Suto,K., Kuribayashi, H., Cyber Physical Systems for Intelligent Disaster Response Networks: Conceptual Proposal and Field Experiment, IEEE Network, Vol.31, No.4, pp.120–128, 2017.

[20] Jain S., McLean, A Framework for Modeling and Simulation for Emergency Response, In Proceedings of the 2003 Winter Simulation Conference, Vol.1, pp.1068–1076, 2003.

[21] Kirov, G., Zlateva, P., Velev, D., Software Architecture for Rapid Development of HLA-Integrated Simulations for Critical Infrastructure Elements under Natural Disasters, International Journal of Innovation, Management and Technology, 6.4, 244, 2015.

[22] Dahmann J. S., Morse K. L., High Level Architecture for Simulation: An Update, In Proceedings of 2nd International Workshop on Distributed Interactive Simulation and Real-Time Applications, pp.32-40, 1998.

[23] Fiedrich, F., An HLA-based Multiagent System for Optimized Resource Allocation after Strong Earthquakes, In Proceedings of the 38th conference on Winter simulation Winter Simulation Conference, pp.486–492, IEEE, 2006.

[24] Hwang, S., Starbuck, R., Lee, S., Choi, M., Lee, S., Park, M., High Level Architecture (HLA) Compliant Distributed Simulation Platform for Disaster Preparedness and Response in Facility Management, In Proceedings of the 2016 Winter Simulation Conference, pp.3365–3374, IEEE, 2016.

[25] Hiroi, K., Inoue, T., Akashi, K., Yumura, T., Miyachi, T., Hironaka, H., Kanno, H., Shinoda, Y., ARIA: Interactive Damage Prediction System for Urban Flood using Simulation and Emulation Federation Platform, In Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, pp.284–287, ACM, 2019.

[26] Misumi, M., Kamiyama, N., Evacuation-Route Recommendation Using DTN with Evacuee Attributes in Disasters, In Proceedings of the 2021 IEEE Wireless Communications and Networking Conference, pp.1–7, IEEE, 2021.

[27] Fauzan, M., Purboyo, W. T., Setianingsih, C., IBR-DTN to Solve Communication Problem on Post-Disaster Rescue Mission, In Proceedings of the 2019 International Seminar on Intelligent Technology and Its Applications, pp.24–28, IEEE, 2019.

[28] Bhattacharjee, S., Basu, S., Roy, S., Das Bit, S., Best-effort Delivery of Emergency Messages in Post-disaster Scenario with Content-based Filtering and Priority-enhanced PRoPHET over DTN, In Proceedings of the 8th International Conference on Communication Systems and Networks, pp.1–7, 2016.

"I'm Going Right" or "Please Go to the Right"?: Disambiguation in Arrow Display on Mobile Robots to Avoid Collision with Passersby[1]

Yo Kuwamiya†, Taketo Imagawa‡, and Minoru Kobayashi‡

†Graduate School of Advanced Mathematical Sciences, Meiji University, Japan
‡Faculty of Interdisciplinary Mathematical Sciences, Meiji University, Japan
{yo.kuwamiya, taketo.imagawa}@koblab.org, minoru@acm.org

Abstract - In recent years, mobile robots that actively interact with humans have started to be used within human living areas. When the robot and pedestrians pass each other, they might collide because it is difficult for a pedestrian to predict the robot's movement from its appearance. We are considering a method of indicating the robot's direction of movement by a horizontal arrow which is displayed on a flat display attached to the robot's body. However, this approach has the problem that pedestrians may interpret the displayed arrow as "indicating the direction passing pedestrians should go". To avoid misinterpretation, we are researching what kind of arrows to use and how to display them to convey the direction of movement with an arrow. In this research, we investigated the recognition of a horizontal arrow presented by a mobile robot. At the same time, we varied the timing and position of the display by using a subjective video in which the cameraman passed the mobile robot. The results showed that an arrow presented by the robot just before passing by pedestrians tended to be interpreted as "indicating the direction passing pedestrians should go".

Keywords: Arrow, Mobile Robot, Direction of Movement, Nonverbal Communication, Metaphorical Design

1 INTRODUCTION

In recent years, mobile robots that actively interact with humans have started to be used within human living areas to solve labor shortages in industrial fields and to reduce human-to-human contact in order to prevent the spread of infectious diseases. Such robots are expected to move without causing harm or anxiety to those around them. However, it is difficult for a human to predict the robot's movement from its appearance. Therefore, even if the mobile robot recognizes the people around it for movement, the people around it do not recognize it and become fearful of the robot's movements. When pedestrians pass by a mobile robot, they might collide with or feel fear from the robot because it is difficult for a pedestrian to predict which direction the robot will go. Therefore, mobile robots must visually indicate the intended direction of movement.

Robots currently under development vary widely depending on their use and design. Hence, it is promising to use arrows, a socially widespread sign, to communicate the robot's direction of movement. Some studies have proposed a method using arrows to indicate the robot's movement [1]–[4]. In these methods, an arrow pointing in the robot's direction of movement is displayed on the top of the robot [1] or projected on the floor in front of the robot [2]–[4].

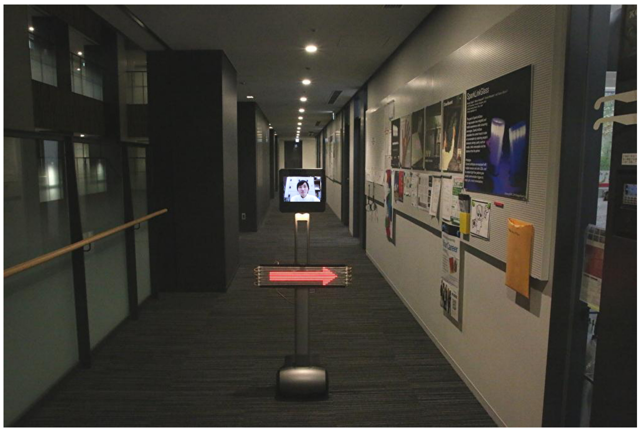


Figure 1: Image of the proposed method. The robot indicate its direction of movement using a horizontal arrow. This method has the problem that the interpretation of the arrow varies depending on individuals.

We are proposing a method of indicating the robot's direction of movement using a horizontal arrow which is shown on a flat display attached to the robot's body to avoid collisions when a pedestrian and mobile robot pass each other. Figure.1 shows an image of the proposed method. This method has the problem that the interpretation of the arrow varies depending on individuals. If a pedestrian passes by a mobile robot that displays a horizontal right arrow as in Figure.1, the direction to avoid differs from person to person depending on how they interpret the arrow. Those who avoid the left side of the robot interpret the arrow as " indicating the robot's direction of movement ". Conversely, those who avoid the right side of the robot interpret the arrow as " indicating the direction passing pedestrians should go ".

In this research, we investigated how the interpretation of a horizontal arrow changes depending on when and where the robot displays the arrow when a mobile robot and a pedestrian pass each other. Based on the findings of this research, we aim to develop a method that communicates the direction of the robot's movement without misunderstanding. This paper presents the results of the research and the outlook for the future of the research.

[1]This work was supported by JSPS KAKENHI Grant Numbers JP18K11410,JP22H03635.

2 RELATED WORK: ACTION ANNOUNCEMENTS OF MOBILE ROBOTS

2.1 Method using non-verbal communication

When people pass each other, they use non-verbal communication such as line of sight [5] and head direction [6] to communicate the direction they will go. There are some related studies that proposed a motion announcement that applies such a communication method [7]–[9]. Matsumaru et al. [7] proposed a method of indicating the robot's movement by displaying eyeballs on an Omni-directional Magicball(R) display attached to the top of the robot's head and transmitting the direction of movement by the position of the displayed eyeballs and the speed by the degree of eye-opening. Similarly, in [8], [9], a method has been proposed in which a face is mounted on the upper part of a robot and the direction of travel is transmitted by the direction of the face.

Equipping robots with facial and eye expression devices has the effect of improving the emotional affinity with the robot. This technique is particularly used when the robot moves autonomously and interacts with people. However, it is difficult to apply such a method to all robots. This is because the shapes and uses of robots vary greatly from one individual to another. To give an example, the mobile telepresence robot, Beam [10] (see Figure.1), moves freely and interacts with people around it. It displays the user's face, which is captured from the user's device, on the screen at the top of the robot. People who interact with the robot recognize the user's face as the robot's face. Therefore, displaying human-like expressions such as eyes or a face in addition to the user's face may confuse humans interacting with the robot.

2.2 Method using arrows

Some studies have proposed using arrows to announce the robot's direction of movement [1]–[4]. Matsumaru et al. proposed a method of displaying an arrow pointing in the direction of travel on a flat panel display attached to the top of the robot [1] and a method of projecting an arrow on the floor in front of the robot [2]. Methods of projecting an arrow pointing in the robot's direction of travel on the floor have been proposed in several further studies [3], [4].

Arrows make it easier for pedestrians to understand the robot's direction of movement. In addition, a method using arrows can be applied to a variety of robots with large individual differences in shape, usage, and movement because it is not a communication method based on social rules or common sense such as blinkers [11]. There are people who do not understand the rules for blinkers, and if they are put on by a robot, it may not be thought of as a blinker. In contrast with the spoken communication method [12], an arrow can communicate the direction of movement to people who speak different languages because an arrow is a globally used sign. Furthermore, it is possible to convey additional information such as the speed of movement by flexibly changing the thickness and length of the arrows.

3 METHOD USING A HORIZONTAL ARROW DISPLAYED ON THE BODY

We are proposing a method of indicating the robot's direction of movement by a horizontal arrow which is displayed on a flat display attached to the robot's body. The advantages of the method are as follows.

Easy to implement — The use of displays to present an arrow is feasible without the need for large equipment or complex systems.

No restriction on arrow display position — In related studies using an arrow [1]–[4], the authors proposed a method of using the top surface of the robot and the floor surface as arrow display locations. In contrast, the proposed method can display an arrow on the robot's body. This feature is free from restrictions on the location of output devices and arrow display positions.

High visibility — The proposed method has higher visibility than the method that displays the arrow pointing forward [1]–[4]. Therefore, people in a wide area around the robot can be made aware of the robot's movements.

4 RESEARCH SUBJECT

4.1 Perception of the horizontal arrow presented by the robot

To communicate the robot's direction of movement by an arrow, it is necessary to indicate clearly what the arrow represents. In related studies using the arrow [1]–[4], it points forward so that the pedestrian understands it indicates the direction of travel of the robot. On the other hand, in our proposed method it is not clear whether the arrow indicates the direction of travel of the robot or a passing pedestrian. This uncertainty may cause a collision between the robot and the pedestrian.

4.2 Consideration of research subject

To convey the direction of travel of the robot, it is necessary to display the arrow in such a way that the pedestrian can easily interpret the arrow as "indicating the direction in which the robot will move". Therefore, we need to know how people perceive the arrow when it is attached to the mobile robot. However, there are no known studies that have investigated the perception of arrows attached to moving objects, because they are usually attached to stationary objects.

In this research, we investigate the recognition of an arrow attached to a mobile robot from the viewpoints of display timing and display position. Based on the findings of this research, we aim to realize a method of displaying the arrow that resolves the research problem.

4.3 Research contents

In this research, we first investigated the difference in interpretation when the robot presented an arrow at a close distance and a far distance from a passing pedestrian. Second, we similarly investigated the interpretation of the arrow when it was presented in nine positions, multiplied by three different heights (top, center, and bottom of the robot) and three different left-right positions (left, center, and right of the robot).

This paper describes the details and results of the first experiment in Chapter 5, and the second experiment in Chapter 6.

5 INVESTIGATION OF PERCEPTION BY THE TIMING OF DISPLAY

This chapter describes the details and results of the experiment that investigated the perception of a horizontal arrow presented by a mobile robot in terms of the timing of the presentation. We experimented by creating a subjective video in which the cameraman passed the mobile robot.

5.1 Experiment details

We conducted an experiment in which the participants watched a video of the robot moving straight ahead and responded to questions asking in which direction participants intended to go to avoid the robot.

Participants. 20 persons (14 males and six females).

Experimental environment. Participants responded to a form created using Microsoft Forms [13] on their PCs.

Robot. We used a mobile telepresence robot, Beam [10]. By operating the robot from a PC, users of this robot can move it freely and communicate with people around the robot. Although the robot was originally designed to display the user's face on the screen at the top of the robot, we hid this with black paper to eliminate the possibility that the face and the line of sight of the user might affect the participant's judgment of the direction of movement. We equipped the robot with an LED panel display to present an arrow. The display was an 11-pixel high and 72-pixel wide LED screen controlled by M5Stack [14], manufactured by M5Stack Technology Co., Ltd. Figure.2 shows the dimensions of the robot, the display, and the arrow.

Experimental video. Figure.3 shows captured images of the experimental video. Figure.4 shows the length and width of the hallway and the composition of the experimental video. The hallway width was 1.8 m, and the length was 8.5 m. The robot and the cameraman moved straight along the centerline at the same speed from both ends as soon as the video started. The robot displayed an arrow at a point between 2.5 m and 6.5 m (1.0 m intervals) from the cameraman. The speed of the robot was approximately 0.5 m/sec. Therefore, the point at which the arrow was displayed was from one second to five seconds before the video ended. The video finished when the distance between the robot and the cameraman was 1.5 m. We informed the participants that the robot would start avoiding action at the end of the video.

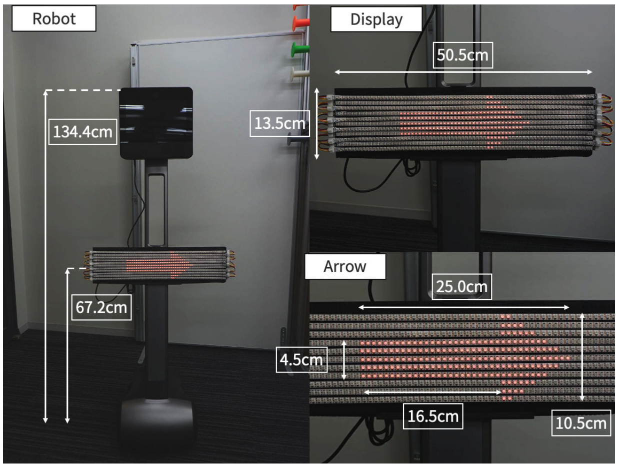


Figure 2: Dimensions of the robot, display, and the arrow used in the experiment.

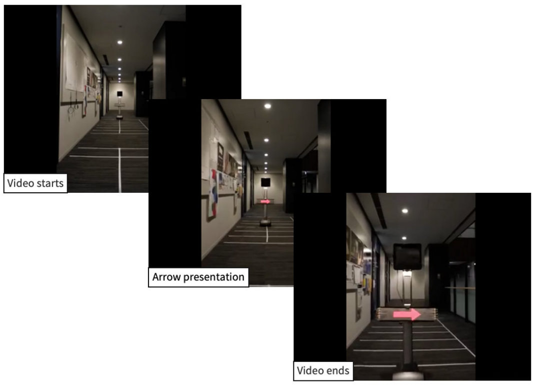


Figure 3: The captured images of the experimental video.

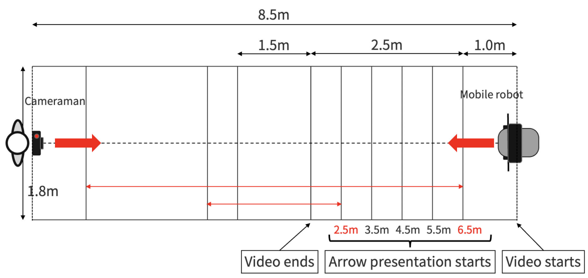


Figure 4: Length and width of the hallway and composition of the experimental video.

Instructional contents. We gave the participants the following instructions to explain the situation in the experimental video.

- You will pass a mobile robot.
- When the robot gets close to a certain distance from you, it will show a horizontal arrow. (There are also videos without displaying an arrow.)
- The arrow may be "pointing in the direction you should go, or it may be "pointing in the direction the robot will go".

- After the video ends, the robot passes you.

Experimental procedure. We experimented as follows.

1. Participants watched the experimental videos. The videos were presented in a random order for each participant.
2. Participants were asked to choose which direction they wanted to go to avoid the robot from five options: {left, likely left, I cannot judge, likely right, right}.
3. Each participant repeated this process for a total of three sets of 12 videos (36 trials per participant). One set consisted of five videos with right-pointing arrows at each timing, five videos with left-pointing arrows at the same timing, and two videos with no arrow displayed. The order of the videos in each set was randomized. After each set, we asked participants to respond with their opinions and impressions in the free-response section.

Evaluation method. We weighted the five choices sequentially by {-2, -1, 0, 1, 2}, and calculated the mean and percentage of responses at each presentation timing.

5.2 Results

Figure.5 shows the average of the responses, and Figure.6,7 shows the percentages of responses by timing. When the robot presented the right arrow, the mean values of the responses at 2.5 m and 3.5 m, which were presented at close distances, were 1.0 and 0.93, indicating a tendency for more respondents to avoid the right side. Similarly, when the robot presented the left arrow, the values of the responses were -0.73, and -0.80, indicating that more participants tended to avoid the left side. On the other hand, when the robot presented the arrows at a far distance of 6.5 m, the value was close to zero. As can be seen from Figure.6,7, the reason for the value being close to zero was not because most participants chose "cannot judge", but because the number of participants who chose "left" and "right" was about the same.

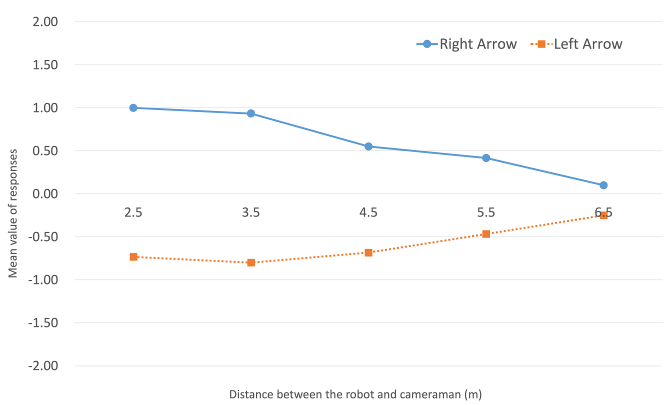


Figure 5: Mean value of responses by the timing of presentation.

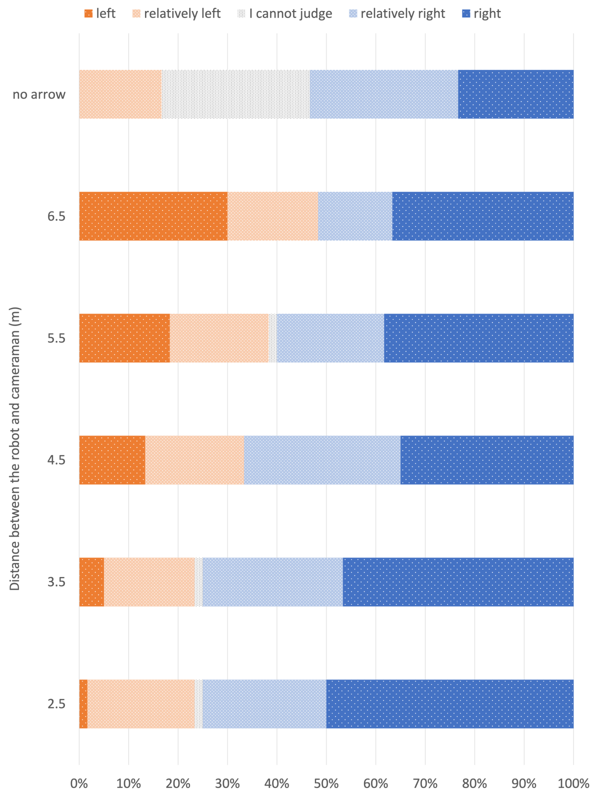


Figure 6: Percentage of responses by the timing of presentation. (Right Arrow)

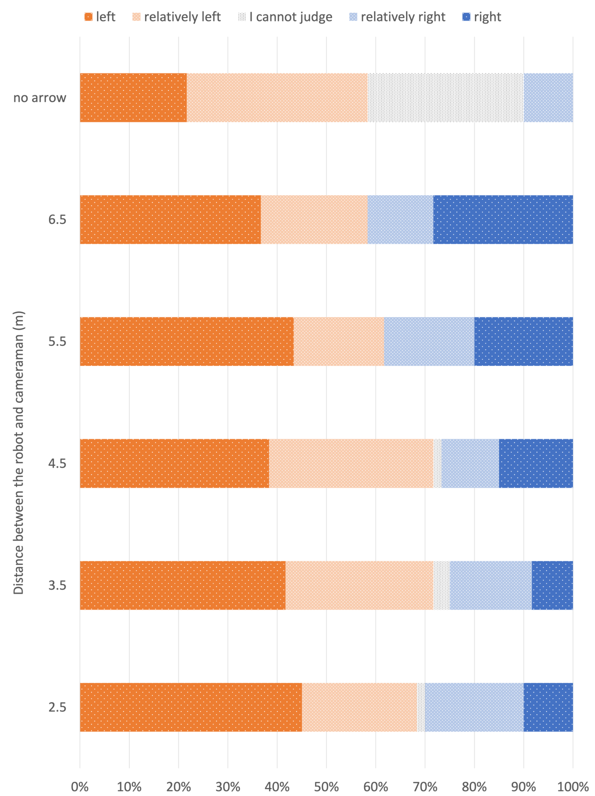


Figure 7: Percentage of responses by the timing of presentation. (Left Arrow)

5.3 Consideration

In this section, we discuss the results of the experiment. When the arrows were displayed at a distance closer than 4.5 m from the cameraman, many participants responded that they intended to go in the direction the arrows were pointing. This indicates that many participants recognized the arrows as "indicating the direction passing pedestrians should go" when the arrows are displayed just before they pass each other. Possible reasons for this result are the usual use of arrows and a problem with space on one side of the hallway used for the experimental video.

First, the arrows that people usually see when walking are used to indicate the direction to a specific place. Therefore, we consider that if the robot displays an arrow just before passing a pedestrian, it is likely to be reflexively judged as " indicating the direction passing pedestrian should go " . On the other hand, if the arrows are presented from a far distance, many people begin to recognize the possibility that the arrow indicates the direction of travel of the robot. This is because they have more time to judge the meaning of the arrow. As a result, the number of people who interpreted the arrows as indicating the direction of the robot increased to be about the same as the number of people who interpreted it as indicating the direction passing pedestrians should go.

Second, the hallway used in this experiment has a space on one side, as shown in Figure.3. In addition, the videos in which the robot presented the left arrow were created by reversing the videos in which the robot presented the right arrow, so in all the videos there was a space in the direction the arrow was pointing. In this experiment, participants watched the two videos in which no arrows were presented and there was a tendency to avoid the direction in which the space was located in both videos (Figure.6,7 no arrows). Furthermore, we obtained the comment, "Because the hallway was asymmetrical, I chose the direction with more space" in the free-response section.

6 INVESTIGATION OF PERCEPTION BY DISPLAY POSITION

This chapter describes the details and results of the experiment that investigate the perception of a horizontal arrow presented by a mobile robot in terms of its presentation position, using the same method as the first experiment.

6.1 Experiment details

Participants. 20 persons (17 males and three females).
Robot. As in the first experiment, we used Beam [10]. An LED panel display can be attached to the robot at three different height positions, as shown in Figure.8. These positions are located at the top, center, and bottom of the robot. The LED panel display can display a right arrow on the left where the arrow tip is in the middle, in the center where the arrow center is in the middle, and on the right where the arrow starts in the middle. Equally, the left arrow is displayed at the left, center, and right of the display. Figure.9 shows the display position of the arrow.

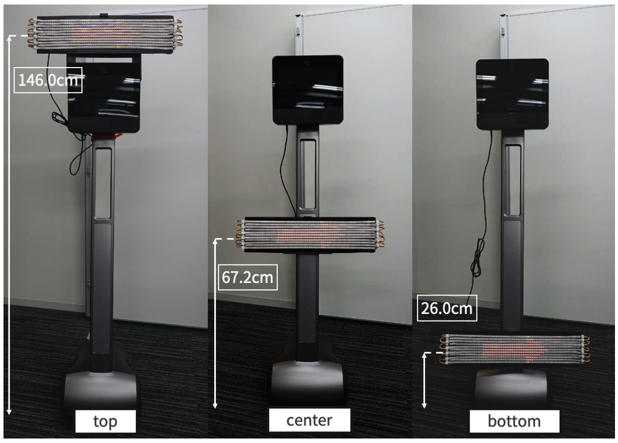


Figure 8: The height positions of the LED panel display.

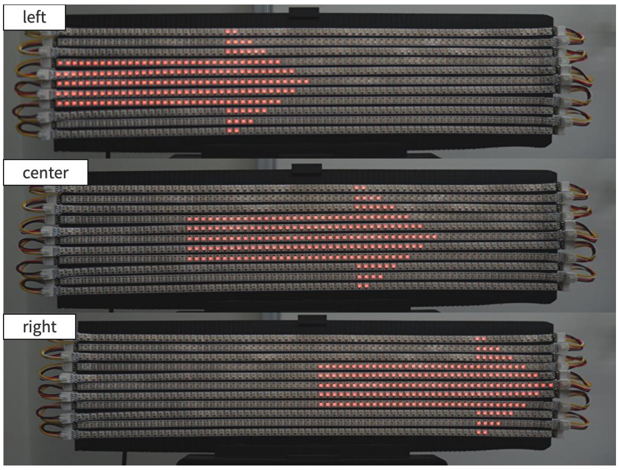


Figure 9: The display position of the arrow.

Experimental video. Figure.10 shows the captured images of the experimental video. To eliminate the problem with the space of a hallway in the first experiment, we shot the videos at a place where the spaces on both sides of the robot were of the same size. Figure.11 shows the length of the path traveled by the robot and cameraman and the composition of the experimental video. The robot and cameraman move straight along the centerline at the same speed from both ends as soon as the video starts. The speed of the robot is about 0.5 m/sec. The robot presents an arrow when the distance from the cameraman reaches 3.5m. We determined the distance regarding the maximum social distance which is proposed by E. T. Hall [15], because a participant in the first experiment commented, "It is hard to see the arrow when it is presented from a far distance, and I don't think it is presented toward me". The robot displayed an arrow at one of nine positions, multiplied by three different heights (top, center, and bottom of the robot) and three different left-right positions (left, center, and right of the robot). As in the first experiment, the video ended when the cameraman and the robot approached 1.5 m. We informed the participant that the robot would begin to avoid them at the end of the video. The instructional content explaining the situation in the experimental video was the same as in the first experiment.

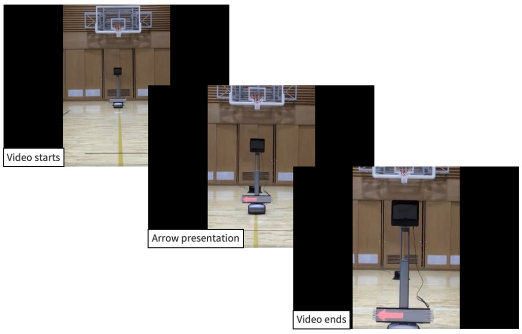


Figure 10: The captured images of the experimental video.

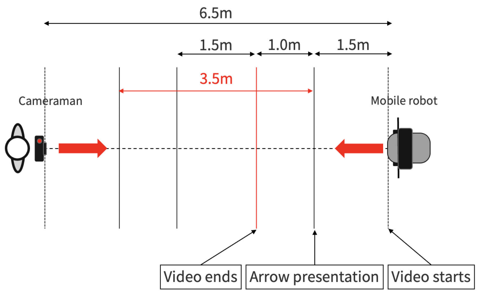


Figure 11: The length of the path traveled by the robot and cameraman and the composition of the experimental video.

Experimental procedure. The experimental procedure was the same as in the first experiment. We presented a total of 21 videos. These videos included nine videos with right arrows, nine videos with left arrows, and three videos with no arrows. The participants repeated the experiment for a total of three sets (63 trials per person).

Evaluation method. The evaluation method was the same as in the first experiment. We asked the participants to state the direction they intended to avoid, choosing from five options {left, relatively left, I cannot judge, relatively right, right}, and then we calculated the mean and percentage of the weighted values in order {-2, -1, 0, 1, 2}.

6.2 Results

Table.1,2 show the mean values. Figure.12,13 shows the percentage of responses by display position. The mean of the responses was above zero for all positions in which the right arrow was presented (Table.1), and many participants tended to avoid the right side of the robot. The same was true for the left arrow, which was below zero in all positions (Table2), and many respondents tended to avoid the left side of the robot. We obtained the comment from the free-response section that “it was difficult to understand when the right arrow was presented to the left and the left arrow to the right”. When the robot displayed right arrows to the left and left arrows to the right, a few participants chose “I cannot judge”, and a few participants answered “right” or “left” with certainty. We found that the display which placed the tip of the arrow in the middle of the robot made the interpretation difficult.

Table 1: Mean value of responses by presentation position. (Right Arrow)

	Left	Center	Right
Up	0.733	0.417	0.533
Center	0.067	0.750	0.883
Bottom	0.450	0.500	0.667

Table 2: Mean value of responses by presentation position. (Left Arrow)

	Left	Center	Right
Up	-0.733	-0.333	-0.633
Center	-0.783	-0.650	0.000
Bottom	-0.583	-0.683	-0.450

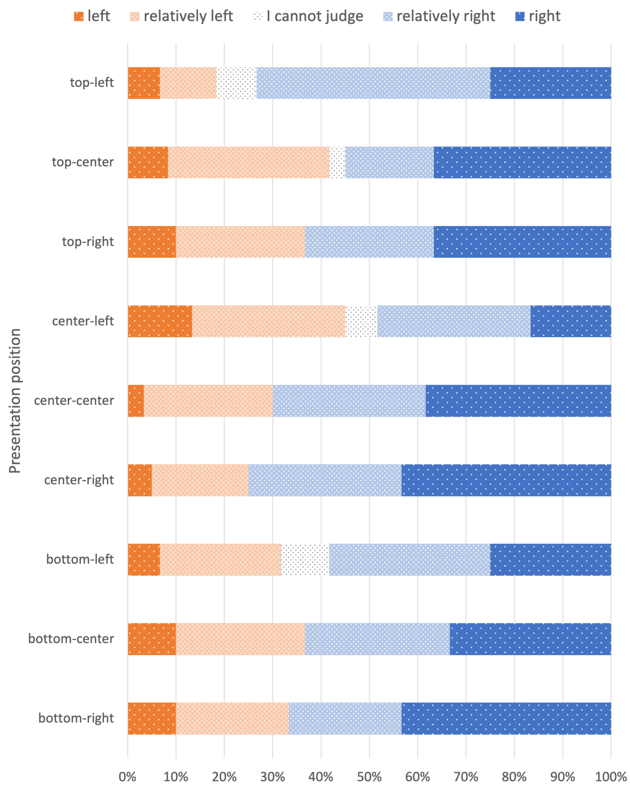


Figure 12: Percentage of responses by the position of the presentation. (Right Arrow)

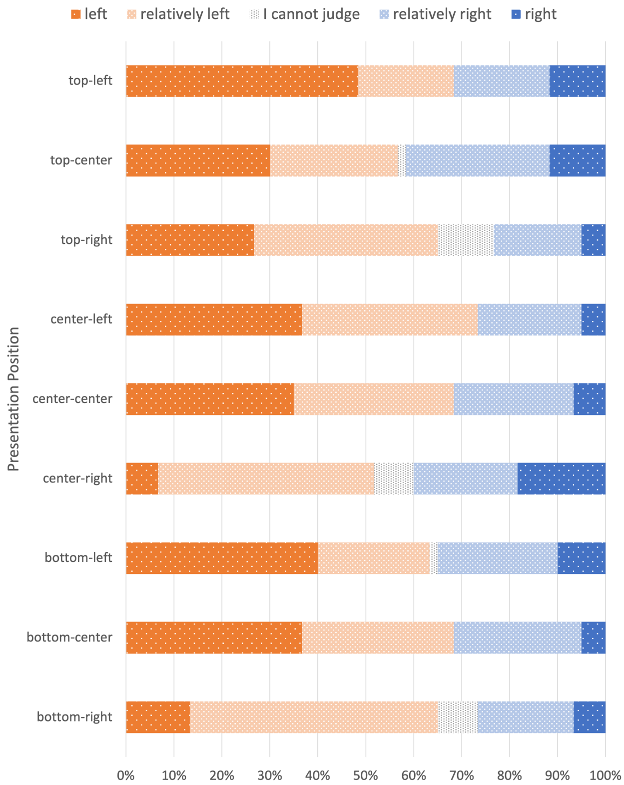


Figure 13: Percentage of responses by the position of the presentation. (Left Arrow)

6.3 Consideration

In this section, we discuss the results of the experiments. The results of the experiment showed that participants tended to avoid the direction indicated by the arrow no matter where it was displayed. In the first experiment, we considered that when deciding the direction in which they intended to go, a large number of participants might be influenced by the space on one side of the hallway. However, in the second experiment, many participants also avoided the direction of the arrow. We prepared three videos without arrows, and the mean values of the responses to these videos were -0.1, 0.0, and 0.1, confirming that there was no effect of the experimental video environment. These results indicate that when the robot presented the arrow just before passing by a pedestrian, the arrow was often perceived as "indicating the direction passing pedestrians should go" even when the problems with the space of a hallway were eliminated.

7 CONTINUITY OF ARROW DISPLAY

From the two experiments conducted in this research, we found that an arrow presented by the robot just before passing by pedestrians tended to be interpreted as " indicating the direction passing pedestrians should go " , regardless of the arrow's display position. Regarding this knowledge, we consider that a mobile robot should always display an arrow to convey the direction of its movement because the arrow presented just before passing by a pedestrian tended to be misinterpreted.

In the proposed method using arrows [1]–[4], an arrow pointing forward was displayed when going straight ahead. On the other hand, we were considering a method of displaying a horizontal arrow just before the robot passes a pedestrian. We believe that this difference is the reason why the robot could not transmit the direction of movement in the experiment. If a passerby can recognize that the arrow indicates the robot's direction of movement when it is moving in a straight line, they can also make judgments without confusion when it turns. We considered this analogy to be crucial in establishing the meaning of the arrow. Therefore, the robot should display the arrow not only when turning but also when moving straight ahead to make the surrounding people aware that the robot is using the arrow to indicate the direction of travel.

In our proposed method, however, the robot presents the arrow on a display attached to the body, so the arrow pointing forward cannot be displayed. Therefore, we will address how to display an arrow when moving straight ahead on a flat display in future research.

8 CONTINUOUS ARROW EXPRESSION

In this chapter, we consider two methods of continuous arrow presentation to indicate that the robot is moving forward and turning right or left.

8.1 Down arrow

The first method entails using use the down arrow to indicate that the robot is moving forward. Down arrows are sometimes used on information boards to represent arrows extending toward the rear of pedestrians. Therefore, when the robot displays a down arrow, a passing pedestrian may recognize it as indicating the direction in which the robot is moving. Based on this possibility, we consider a method in which the down arrow indicates the robot's forward motion, and the arrow rotates to match the angle of the turning motion. Figure.14 shows an image of the method.

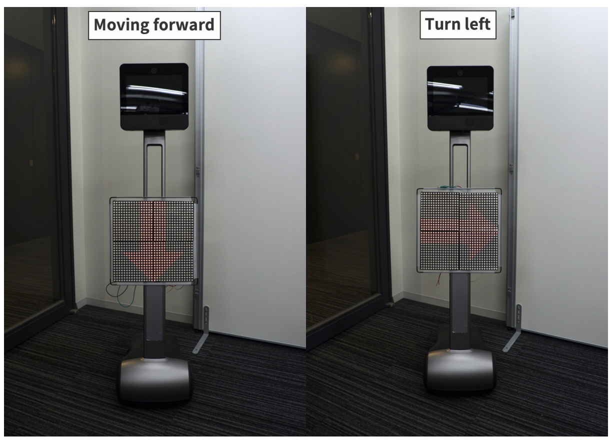


Figure 14: Image of the method using use the down arrow to indicate that the robot is moving forward. The method in which the down arrow indicates the robot's forward motion, and the arrow rotates to match the angle of the turning motion.

8.2 Double-headed arrow

The second method entails using a double-headed arrow to indicate the robot's motion. When the robot is moving forward, it displays a double-headed arrow where both sides are the same length, and when the robot is turning, it displays a double-headed arrow where one side extends in the direction in which the robot will go. Figure.15 shows an image of the method.

This method was inspired by bicycle hand signals. Bicycles do not have signals like blinkers, so cyclists indicate which direction they will take by extending their hands in the direction they are going. Although hand signals are socially known rules, the indication method is easy to understand visually. We consider that a double-headed arrow can also indicate the direction of the robot's movement in the same way as hand signals. In addition, the extending movement of the arrow can express the angle of the turn.

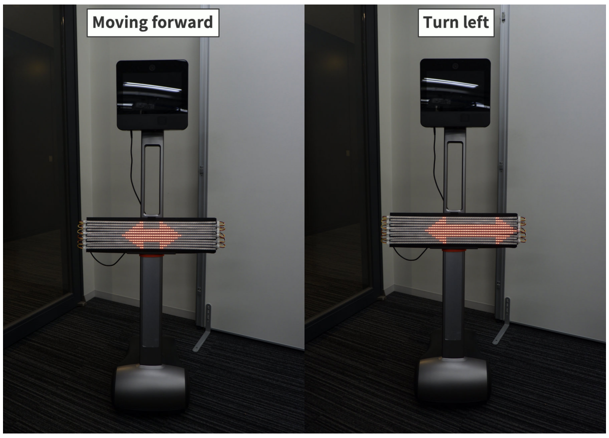


Figure 15: Image of the method using a double-headed arrow. The method of communicating the direction of movement by using an arrow extending in the direction in which the robot turns.

8.3 Future perspective

We will verify whether these methods can convey the direction of movement of a mobile robot in an experiment in which a pedestrian passes by the robot. In the experiment, we will compare methods with no indication and a horizontal arrow that the robot presents just before passing by a participant, and evaluate these methods by observing the participants' behavior and evaluating impressions of the robot.

9 CONCLUSION

We are studying a method to avoid collisions when a mobile robot and a pedestrian pass each other by indicating the robot's direction of movement with a horizontal arrow attached to the robot's body. In this research, we investigated the perception of arrows presented by a mobile robot in terms of the timing and position of arrow presentation through an experiment using a subjective video of passing by a robot. As a result of the two experiments, we found that an arrow presented by the robot just before passing by pedestrians tended to be interpreted as "indicating the direction passing pedestrians should go", regardless of their display position. Based on this knowledge, we believe the robot needs to always display an arrow to indicate its direction of movement. We considered methods of continuous arrow presentation using a down arrow and a double-headed arrow.

REFERENCES

[1] Matsumaru. T, Mobile Robot with Preliminary-announcement and Indication Function of Forthcoming Operation using a Flat-panel Display, 2007 IEEE Int. Conf. on Robotics and Automation (ICRA'07), ThA4.4, pp.1774-1781 (2007).

[2] Matsumaru. T, Mobile Robot with Preliminary-announcement and Display Function of Forthcoming Motion using Projection Equipment, ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp.443-450 (2006).

[3] Coovert. M. D, Lee. T, Shindev. I and Sun. Y, Spatial augmented reality as a method for a mobile robot to communicate intended movement, Computers in Human Behavior, Vol.34, pp.241-248, (2014).

[4] Shrestha. M. C, Onishi. T, Kobayashi. A, Kamezaki. M and Sugano. S, Communicating Directional Intent in Robot Navigation using Projection Indicators, 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp.746-751 (2018).

[5] Nummenmaa. L, Hyönä. J and Hietanen. JK, I'll Walk This Way: Eyes Reveal The Direction of Locomotion and Make Passersby Look and Go the Other way, Psychol Sci., vol.20, no.12, pp.1454-1458, (2009).

[6] Ueda. S and Kitazaki. M, Collision Avoidance Affected by Walker's Head Direction in a Virtual Environment, HCI International 2013 - Posters' Extended Abstracts, CCIS 374, C. Stephanidis, Ed. Springer, Berlin, Heidelberg, pp.727–731 (2013).

[7] Matsumaru. T, Iwase. K, Akiyama. K, Kusada. T and Ito. T, Mobile Robot with Eyeball Expression as the Preliminary-Announcement and Display of the Robot's Following Motion, Autonomous Robots, Vol.18, No.2, pp.231-246 (2005).

[8] Yamashita. S, Ikeda. T, Shinozawa. K and Iwaki. S, Evaluation of Robots that Signals a Pedestrian Using Face Orientation Based on Moving Trajectory Analysis, 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1-8 (2019).

[9] Lyu. J, Mikawa. M, Fujisawa. M and Hiiragi. M, Mobile Robot with Previous Announcement of Upcoming Operation Using Face Interface, 2019 IEEE/SICE International Symposium on System Integration (SII), pp.782-787 (2019).

[10] Beam, https://gobe.blue-ocean-robotics.com/beam-to-gobe accessed 11 June 2023

[11] Shrestha. M. C et al, Exploring the use of light and display indicators for communicating directional intent,

2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), pp. 1651-1656 (2016).
[12] Muramatsu. S, Higashi. S, Chugo. D, Yokota. S and Hashimoto. H, Consideration of the preliminary announcement function for the human friendly service robot, IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, pp. 5868-5872 (2016).
[13] Microsoft Forms, https://www.microsoft.com/en-us/microsoft-365/online-surveys-polls-quizzes accessed 11 June 2023
[14] M5Stack, https://m5stack.com/ accessed 11 June 2023
[15] E. T. Hall, The Hidden Dimension, Doubleday, New York (1969).

A Broadcast-based Online Machine Learning System for Environmental Value Prediction

Tomoki Yoshihisa[*,**], Rikuto Kotouge[**]

[*]Department of Data Science, Shiga University, Japan
[**]Graduate School of Engineering Science, Osaka University, Japan
yoshihisa@biwako.shiga-u.ac.jp

Abstract -Various online machine learning systems have been studied recently. In ordinary online machine learning systems, machine learning servers cyclically collect training data from data sources such as connected vehicles, smart phones, etc. By collecting only the data that contribute to improve the performances of the machine learning models, the times needed for the data collections in each cycle are reduced without the performance degradations. However, it is difficult for the data sources to judge whether their data contribute to the improvements or not because they do not have the latest learned models. Hence, in this paper, we propose a broadcast-based online machine learning system. In our proposed system, the server broadcasts the latest model to the data sources. Each data source sends its obtained data to the server if the data can contribute to the performance improvement with a high probability.

Keywords: federated learning, neural networks, distributed computing.

1 INTRODUCTION

Various artificial intelligence (AI)-based information systems have recently attracted much attention. AI technologies have various functions and are used in various fields. Many of these artificial intelligence technologies use machine learning. Machine learning is a technology for learning decision criteria from data. The most critical aspect of machine learning is prediction accuracy, which is impacted by the type and size of training data used to train machine learning models. Two learning frameworks are used in machine learning: online learning and batch learning.

In online learning, the machine learning models are continuously trained each time data is obtained. Therefore, these models are constantly considering the latest situations. The advantage of the online machine learning system is that the model can be trained while collecting training data, so the model is always reflecting the latest situation. The learning server in the above example can provide more up-to-date data than the offline machine learning system used in batch learning. Therefore, online machine learning systems are often used when the latest situation changes frequently.

However, the online machine learning model is trained in time intervals after collecting a specific data size. If these time intervals are long, the prediction accuracy of online machine learning models too can be negatively impacted. Although these time intervals can be reduced by reducing the data size required for the next training, the model cannot

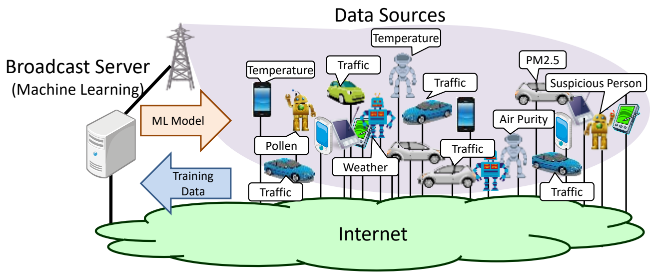


Figure 1: A brief image of the broadcast-based online machine learning system

be sufficiently trained with less data because there will be a larger machine learning loss. As the loss increases, the accuracy of the model decreases. Therefore, selecting such data as the system collects training data is brilliant. In this approach, to reduce the number of training data to be collected, it is necessary to compute the result of the loss function before the system collects training data. That is, each data source must calculate the results of its training data. This calculation requires the parameters of the machine learning model.

Therefore, in this paper, we propose an online machine learning system that uses data broadcasting to reduce the delay in reflecting observed environmental values in a machine learning model[1]. In the proposed system, the parameters of the machine learning model are periodically broadcasted to the data source, and if the data source calculates the loss and the result is greater than a given threshold, the training data is sent for machine learning. With the above approach, the proposed system reduces time by reducing the data size collected. The main novelties of this report are as follows: 1) an online machine learning system using data broadcasting, 2) a method for selecting highly efficient training data, 3) experimental evaluation of the proposed system.

The rest of the paper is organized as follows. Section 2 presents related works on the proposed system. Section 3 describes an online machine learning system using data broadcasting. Section 4 expounds on the proposed training data selection method, and Section 5 evaluates the proposed system.

2 RELATED WORK

[1] The work reported in the paper was supported by a Grants-in-Aid for Scientific Research (B) numbered 21H03429, 22H03587, 22H03700, (C) numbered 18K11316, 20K11829, 22K12009.

2.1 Data Collection in Sensor Networks

Various data collection techniques in sensor networks have been studied in the past decade.

Among them, a data collection technique has been proposed that performs aggregation such as averaging and summing [1,2]. In this technique, each sensor device aggregates its observed values and the values received from other sensor devices. Aggregation also reduces the amount of data to be collected and enables faster data collection. In the system proposed in [3], when the user sends a probabilistic query, the system collects sensor data based on the order determined by the probabilistic model of sensor values. The system can reduce the time required to aggregate sensor data by creating an efficient aggregation plan based on the submitted query. However, these techniques aggregate sensor data, so the sink cannot collect every raw data.

Several data collection techniques have been proposed to reduce data collection latency in sensor networks. In a system called Ken, the base station generates a probabilistic model of the sensor network and collects only model update information from the physical sensor network [4]. In this system, even though the collected values are probabilistically correct, the data collection latency is reduced using a probabilistic model.

A practical probabilistic model for health care systems has been proposed in [5]. The article [6] proposes a query processing technique to reduce the waiting time. The authors consider the predicates of the submitted query and construct the order in which queries are executed to reduce the waiting time to collect related sensor data. However, the application of these techniques is not online machine learning, and they do not account for machine learning loss.

2.2 Communication Traffic Reduction

Several techniques have been proposed to reduce the communication traffic for machine learning. In some cases, the data size sent from the data source can be reduced by converting the data types to those required for machine learning. Such transformations are performed on the data source and are referred to as fog/edge computing [7-10]. The article [10] proposes a machine learning-based sensor data analysis system using the fog computing paradigm. This system converts data sources into data suitable for feed-forward neural networks, reducing the data size. The literature confirms that the proposed system can reduce the data size sent from the data source in practical situations. In addition, the loss of accuracy of the neural network in this system is small, but this system does not consider the loss in training the machine learning model.

Reduction of communication traffic can also be achieved by data cleaning. Data cleaning is a technique to remove dirty data such as incomplete or inaccurate data from observed data. The Sample-and-Clean proposed in [11] and Progressive Data Cleaning proposed in [12-15] can reduce the computational cost of data cleaning. These studies indicate that the losses on these systems may increase because the target application is not machine learning.

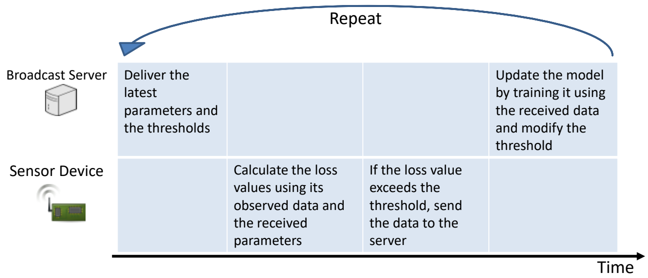


Figure 2: The procedure of the assumed system

ActiveClean, proposed in [15], is a data cleaning system targeted at machine learning. The system trains a machine learning model using the original and cleaned data, checks the accuracy of both models, and cleans the data so as not to degrade the accuracy. However, ActiveClean does not consider the time required to collect training data, which is the objective of this report.

3 BROADCAST-BASED ONLINE MACHINE LEARNING SYSTEM

In this chapter, we describe our proposed broadcast-based online machine learning system.

3.1 Summary

This study aims to rapidly reduce the loss of online machine learning by effectively selecting the training data to be collected. We propose an online machine learning system to achieve this goal in a data broadcasting environment.

A brief image of the system is shown in Figure 1. In our proposed system, we set an upper bound on the number of data sources that can send training data to a computer for online machine learning (broadcast server). The upper limit is set according to the communication bandwidth and the learning time. To keep the number of data sources close to the upper limit, the broadcast server adjusts the loss threshold and broadcasts the data.

As explained in Section 1, in systems that optimize machine learning models using gradient descent, training data with a large impact on the loss have large gradients, and the loss decreases significantly after learning. Therefore, training data with large losses are preferentially transmitted in our proposed system. The weight parameters of the machine learning model are necessary to calculate the loss of each data source.

Moreover, the proposed system uses data broadcasting to distribute these weight parameters to all data sources in a short time. The broadcasting server receives the training data sent from the data source and starts training the machine learning model when it finishes collecting the training data. When the training is complete and the weight parameters are updated, the updated weight parameters are broadcasted together with the threshold values, and the subsequent training data collection begins.

3.2 Assumed Data Broadcasting Environment

Recently, with the development of wireless technology, the data broadcasting environment has attracted much attention. Two data distribution channels exist in a data broadcasting environment: a broadcasting channel and a communication channel. The image is shown in Figure 1.

The broadcasting server is equipped with broadcasting equipment to broadcast data to the source. The broadcasting devices are assumed to broadcast data using wireless communication protocols such as WiFi, ISDBT, and Bluetooth. In this case, data sources located in the area where radio waves for wireless communication can reach can receive the broadcasted data. A data source that is not encountering communication errors can receive broadcast data. The broadcasting server is connected to the Internet and can receive the data sent from the data sources. A broadcasting server can be a computer for training machine learning models in a data broadcasting environment for machine learning.

Data sources are devices with various sensors such as smartphones, smart cars, and sensors installed in buildings. These devices can continuously acquire sensor data and transmit them to other computers via the Internet. The broadcasting server can also receive data broadcast from broadcasting equipment.

3.3 Advantages and Disadvantages

The most significant advantage of data broadcasting is that the broadcasting server can simultaneously distribute the same data to all data sources. When the same data is delivered over a communication channel, the broadcasting server sends the data to each data source sequentially (unicast). In the case of sequential transmission, the communication time is proportional to the number of data sources, and the more data sources there are, the longer it takes. Therefore, the time required for the delivery of the broadcast channel is shorter than that for the communication channel.

One of the disadvantages of data broadcasting is that both the broadcasting server and the data source needs to manage the channel. Therefore, the required computational resources and power consumption increase. However, this is not a significant disadvantage because wireless communication modules in recent desktop computers and smartphones can receive broadcast and unicast data.

For example, a typical WiFi module can receive data sent to a broadcast IP address and data sent to a unicast IP address.

3.4 Application to Machine Learning

As described above, the advantage of data broadcasting is that it reduces the time required for the broadcasting server to deliver the same data to all data sources. In other words, the data sources can share the data in a short time. Machine learning data sources do not need data because they only send training data to train machine learning models. However, the proposed system shares the parameter data of the training model. Data for machine learning include training data, model data, and parameter data, which need to be prepared by the system side. Therefore, data broadcasting is effective for sharing data for machine learning.

3.5 Application Example

Consider a broadcast-based online machine learning system for predicting rain at a random location. The procedure of the assumed system is shown in Figure 2. The broadcasting server and the data source are connected via the Internet.

The data source is a sensor terminal with sensors mounted on various IoT devices such as Raspberry Pi. These sensor terminals are equipped with 5G communication modules connected to the Internet by wireless communication protocols.

The broadcasting server broadcasts data to the data source over the 5G network using an IP address. We construct the neural network by setting the period, the loss threshold, and the upper threshold of the data. Upon receiving the weight parameters of the model, the data source checks whether the loss exceeds the threshold value. If the threshold is exceeded, the data (latitude, longitude, and rain) is sent to the broadcast server.

The broadcast server starts training the online machine learning model with the received training data. If the number of training data is more than the upper limit, the threshold value for the next broadcast is increased to reduce the data size. Otherwise, if the data size is small, the threshold is reduced, and the data size is increased. When the training is completed, the broadcasting server broadcasts the machine learning model's updated weight parameters and the data source's threshold values.

4 PROPOSED METHOD

This chapter describes the details of the proposed data collection method.

4.1 Summary

The proposed broadcast-based online machine learning system aims to improve accuracy by preferentially collecting data with large errors rapidly. Therefore, we select the data with the largest error among all the data.

If we know the errors of all the data and then train the data with large errors, we can optimally train the data with large errors. However, the proposed system limits the number of data sources transmitting data due to the limitation of communication bandwidth and learning time, so the broadcasting server cannot know the error of all data. Therefore, the data source needs to judge whether the observed data has a large error without the error information of other data sources. The broadcast server gives the threshold value for this decision.

By transmitting data if the error is larger than the threshold value, the data source can determine whether the observed data has a large error without error information from other data sources. If this threshold is too small, the data size to be transmitted will be too large, and the upper limit will be exceeded. Meanwhile, if the size is too small, less data will be transmitted, and the accuracy will decrease due to the decrease in the training data; to prevent this, the broadcast server dynamically adjusts the threshold value.

4.2 Phases

In online machine learning, the training of the machine learning model is repeated every time the computer obtains training data. This iterative period is the learning interval in our proposed broadcast-based online machine learning system, as shown in % Figure 2.

A learning interval consists of the following two stages. We assume that the time managed by the broadcasting server and the data source are synchronized. Time synchronization can be easily achieved using the Network Time Protocol (NTP).

- Broadcasting Phase: The broadcasting phase is when the broadcasting server trains the machine learning model and updates the weight parameters and thresholds. The broadcasting server broadcasts the data sources' weight parameters and threshold values. In order for the broadcasting server to learn all the collected data, the time of the broadcasting phase should be longer than the time needed for learning.
- Collection Phase: Each data source observes environmental values and generates training data. Upon receiving the model weight parameters, the data source calculates the loss and then sends it to the broadcast server if it is greater than a threshold value. In our proposed system, the data sources with large losses preferentially send training data to reduce learning loss.

4.3 Broadcast Server Behavior

The broadcast server receives the training data sent from the data source. If the received training data is larger than the upper limit, the threshold value for the next broadcast is increased to reduce the data size. However, if it is small, the threshold is reduced, and the data size is increased.

When it is time to move to the broadcast phase, the broadcast server starts training the online machine learning model. Since the broadcasting server and the data sources are synchronized in time, the start of the broadcasting phase can be recognized.

When the training is finished, the broadcasting server broadcasts the machine learning model's updated weight parameters and the data sources' threshold values. Then the next learning interval is started.

4.4 Data Source Behavior

Each data source always observes the environmental values. Upon receiving the initial values of the model weight parameters and threshold values from the broadcast server, the collection phase begins, and the errors are calculated using the model weight parameters and the observed values.

The data is sent to the broadcast server if the calculated error exceeds the threshold. The updated values are received from the model weight parameters and threshold values when broadcasted from the broadcast server. This process is repeated periodically.

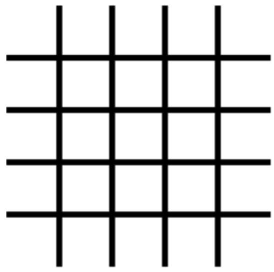


Figure 3: The road map for the evaluation

4.5 Time Synchronization

The time of the broadcasting server and the data source are synchronized. The broadcasting server broadcasts the parameters and threshold values of the learning model every specific time according to a predefined cycle. Therefore, it is necessary to complete the learning process before the broadcast. If enough bandwidth is available, this learning time determines the number of stations that can receive the broadcast. The broadcasting server broadcasts the data without training if the training is not completed by the next cycle. The collection phase begins when the data source receives the model weight parameters and threshold values. When the data is received in the collection phase, it moves to the broadcasting phase, and the broadcasting server learns the data.

5 EVALUATION

The evaluation results of the proposed system are presented in this section. This evaluation presents the results of the evaluation of the proposed system under the rain scenario.

5.1 Configuration

In this experiment, we assumed a system of the section 3 and performed simulations. We considered a connected car equipped with sensors for rain or temperature always connected to a 5G network.

The broadcasting server and the vehicle are connected to the Internet and can communicate.

The data source and the broadcasting server share the online machine learning model and start learning.

The car driving simulation was performed using SUMO, an open-source traffic flow simulator. The map used in the simulation is shown in Figure 3.

The figure shows a 1000 m square area for the simulation. Cars were randomly selected from one endpoint for departure and arrival points, respectively, and were allowed to run for 1000 [s] with one car per second inflow. When the car arrives at an intersection, it chooses its course with equal probability and proceeds. There are always approximately 100 cars running on the map.

The online machine learning model is LSTM, a recurrent neural network suitable for predicting time series data. The input data of this model is continuous location information, and the output data is environmental data such as rain, assuming the example described in the subsection 3.5. The input data is continuous two-dimensional data, and the output is one-dimensional data. Each data source has all the input and output data (training data) of the neural network and computes the loss using its observed training data. In this experiment, the number of hidden layers is 2, and the number of neurons in each hidden layer is 64. The uniform random algorithm gives the initial values of the weight parameters of the neural network. The activation function of each layer is assumed to be a hyperbolic tangent. The gradient descent algorithm used to optimize the weight parameters is RMSProp.

Since it took approximately 10 [s] to train 1000 epochs of the neural network model on the author's computing environment, we assume that the time required to train one epoch is 10 [ms]. In order to systematically evaluate the system, the observed rain by a vehicle located at (*x*, *y*) at time *t* is given by 1/(1+exp(-0.05(x+y+t-2000))). Value for rain ranges from 0 to 1.

In this experiment, the locations on the map shown in Figure 3 are clear at the initial stage (*t* =0), The upper right (1000, 1000) indicates that rain clouds are approaching, and rain begins to fall in the area covered by the clouds. The clouds move to the lower left (0, 0), and the rain changes from 0 to 1 at the point of cloud cover. Therefore, at the initial stage (*t* =0), rain at all locations is 0, and at the end of the simulation (*t* =1000), rain is 1 because all regions have rain.

The broadcast period is 10 [s], and the model's updated weight parameters and threshold values are broadcasted to the sensor terminals every 10 [s]. The initial value of the threshold is 0.1 and is adjusted every 10 [s].

If the transmitted data size is more than the upper limit, the threshold is increased by a factor of 1.1 to reduce the size of the subsequently transmitted data; if the number is less than the upper limit, the threshold is increased by a factor of 0.9 to increase the size of the subsequently transmitted data.

To evaluate the accuracy of the proposed method, which selects and collects lossy data, we compared the results of the proposed method with those of the randomly selected data. In the proposed method, data with an extensive root square error between the predicted and measured values are preferentially collected as data with a significant loss.

5.2 Results

A graph of the results of the proposed method and random selection is shown in Figure 4. The vertical axis indicates the mean value of the error between the predicted rain and the measured rain (prediction error) for all vehicles, and the horizontal axis indicates the time *t* [s]. Since no rain falls at the beginning of the simulation and rain is 0 at all locations, the prediction errors of both the proposed method and random selection are minimal, indicating good accuracy. As the rainfall gradually changes from 0 to 1, the prediction error becomes larger, and the accuracy decreases. The

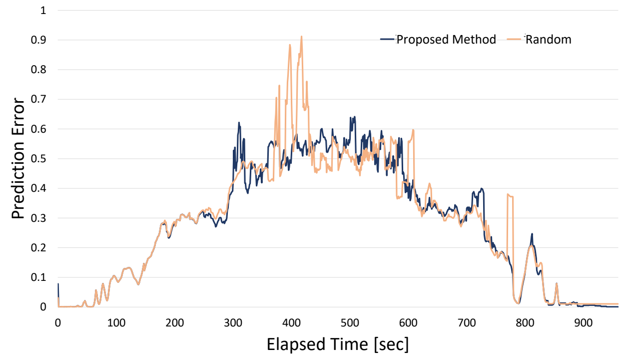


Figure 4: The elapsed time and the prediction error (rain scenario)

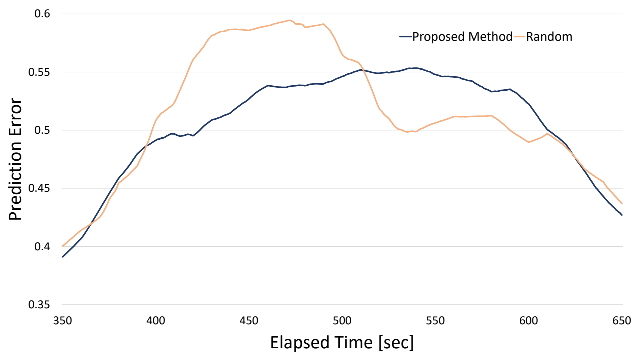


Figure 5: The moving average for 100 [s] of the prediction error (rain scenario)

prediction error becomes very large at approximately 400 [s]. However, the proposed method can keep the prediction error small. This is because the proposed method improves the accuracy at high speed by preferentially collecting data with significant losses. At the end of the simulation, it is raining at all locations, the proposed method and random selection have minimal prediction errors and good accuracy.

Because of the large variation of the forecast error in Figure 4, the vertical axis is changed to the moving average of the forecast error for 100 [s], and a graph showing the period from 350 [s] to 550 [s], when the variation of the forecast error of the proposed method and random selection is large, is shown in Figure 5.

As shown in Figure 5, the prediction error of the proposed method is smaller at approximately 400 [s]. However, at approximately 500 [s], the prediction error in random selection is smaller than that of the proposed method, and the accuracy of the proposed method is lower than that of random selection. This is because while the proposed method can easily reflect changes, it predicts too large a change. All data in this experiment were trained with large losses, but this can probably be improved by partially training data at random.

The experimental evaluation results will show that the proposed method can reduce the prediction error by approximately 10% compared to random selection. In this experiment, the proposed method can improve the prediction accuracy even when sudden rain is assumed.

6 CONCLUSION

In this study, we proposed an online machine learning system under a data broadcasting environment to reduce the delay in reflecting observed environmental values into the machine learning model. In the proposed system, the data broadcasting server periodically broadcasts the model weight parameters and threshold values, and only the data sources of training data whose loss is larger than the collection threshold sends training data to the data broadcasting server. This method reduces the number of data to be collected and latency.

Future issues include adjusting the threshold value and evaluating the proposed method on complex maps. We could not evaluate the method by adjusting the threshold value in the present experimental evaluation.

REFERENCES

[1] Y. Yin, L. Long, and X. Deng: Dynamic Data Mining of Sensor Data, IEEE Access, vol. 8, pp. 41637–41648 (2020).

[2] J. Considine, F. Li, G. Kollios, and J. Byers: Approximate Aggregation Techniques for Sensor Databases, in Proc. IEEE International Conference on Data Engineering (ICDE), pp. 449–460 (2004).

[3] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong: Model-driven Data Acquisition in Sensor Networks, in Proc. International Conference on Very Large Databases, volume 30, pp. 588–599 (2004).

[4] D. Chu, A. Deshpande, J. M. Hellerstein, and W. Hong: Approximate Data Collection in Sensor Networks using Probabilistic Models, in Proc. International Conference on Data Engineering (ICDE), pp. 48–59 (2006).

[5] N. Nikzad, J. Yang, P. Zappi, T. S. Rosing, and D. Krishnaswamy: Model-driven Adaptive Wireless Sensing for Environmental Healthcare Feedback Systems, in Proc. IEEE International Conference on Communications (ICC), pp. 3439–3444 (2012).

[6] A. Deshpande, C. Guestrin, W. Hong, and S. Madden: Exploiting Correlated Attributes in Acquisitional Query Processing, in Proc. IEEE International Conference on Data Engineering (ICDE), pp. 143–154 (2005).

[7] E. I. Gaura, J. Brusey, M. Allen, R. Wilkins, D. Goldsmith, and R. Rednic: Edge Mining the Internet of Things, IEEE Sensors, vol. 13, no. 10, pp. 3816–3825 (2013).

[8] A. Jain, E. Y. Chang, and Y.-F. Wang: Adaptive Stream Resource Management using Kalman Flters, in Proc. ACM SIGMOD International Conference on Management of Data, pp. 11–22 (2004).

[9] H. M. Raafat and A. Tolba: Homoscedasticity for Defect Detection in Homogeneous Fat Surface Products, Textile Research Journal, vol. 85, no. 8, pp. 850–866 (2015).

[10] H. M. Raafat, M. S. Hossain, E. Essa, S. Elmougy, A. S. Tolba, G. Muhammad, and A. Ghoneim: Fog Intelligence for Real-Time IoT Sensor Data Analytics, IEEE Access, vol. 5, pp. 24062–24069 (2017).

[11] J. Wang, S. Krishnan, M. J. Franklin, K. Goldberg, T. Kraska, and T. Milo: A Sample-and-Clean Framework for Fast and Accurate Query Processing on Dirty Data, in Proc. ACM SIGMOD International Conference on Management of Data, pp. 469–480 (2014).

[12] Y. Altowim, D. V. Kalashnikov, and S. Mehrotra: Progressive Approach to Relational Entity Resolution, the VLDB Endowment, vol. 7, no. 11, pp. 999–1010 (2014).

[13] T. Papenbrock, A. Heise, and F. Naumann: Progressive Duplicate Detection, IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 5, pp. 1316–1329 (2015).

[14] M. Yakout, A. K. Elmagarmid, J. Neville, M. Ouzzani, and I. F. Ilyas: Guided data repair, arXiv preprint arXiv:1103.3103 (2011).

[15] S. Krishnan, J. Wang, E. Wu, M. Franklin, and K. Goldberg: Activeclean: Interactive Data Cleaning for Statistical Modeling, the VLDB Endowment, vol. 9, pp. 948–959 (2016).

<u>Session 6:</u>
<u>Systems and Applications 2</u>
(Chair: Kei Hiroi)

DNN-based Fault Localization with Virtual Coverage based on Number of Executions

Takuma Ikeda†, Hitoshi Kiryu†, Shinpei Ogata†, and Kozo Okano†

†Shinshu University, Japan

Abstract - Automatic fault localization is a technique that helps to reduce the costly task of program debugging. Among the existing approaches, Spectrum-based fault localization (SFL) shows promising results in terms of scalability. SFL calculates the suspiciousness values of each statement from the coverage information provided by test cases. One Deep Neural Network (DNN)-based SFL approach that uses virtual coverage has been proposed. This approach uses a DNN model that takes code coverage as input and classifies whether the test result of an input code coverage is Pass or Fail. Virtual coverage is code coverage that expresses that only certain code blocks are executed. We propose a new virtual coverage based on the number of executions. Our idea is that by using execution count-based virtual coverage, which better reflects actual execution patterns, higher accuracy can be achieved than existing approaches. In this paper, we evaluate our proposed approach using three projects available on Defects4j. As a result of the evaluation, we confirmed that using execution count-based virtual coverages improve the accuracy by up to 4.2 points compared to the existing approach.

Keywords: Spectrum-based Fault Localization, DNN, Virtual Coverage

1 INTRODUCTION

In software development, fault localization is a costly task. Testing and debugging are reported to account for up to 75% of the development cost[12]. Automatic fault localization is an effective technique to reduce the cost of program debugging. Among the existing methods, Spectrum-based fault localization (SFL) has shown promising results in terms of scalability, lightweight, and language-agnostic[11][9][10].

In recent years, several deep learning-based approaches have been proposed for locating faults, and the learning capability of DNNs is effective in locating faults, showing better identification results than conventional SFL techniques (Ochiai, Tarantula)[16]. As one of the deep learning-based SFL approaches, the approach using virtual coverage has been proposed[14][15][16][7]. An overview of this approach[7] is shown in Figure 1. The virtual coverage used in this approach is shown in Figure 2. The approach in Figure 1 takes test coverage as input and learns a Deep Neural Network (DNN) model that classifies whether the input test execution coverage is Pass or Fail. Next, the virtual coverage shown in Figure 2 is input to a learned DNN model, and a DNN model outputs a score indicating the suspiciousness of failure for each code block.

The virtual coverage in Figure 2 is generated from the test execution coverage and therefore treats the statements commonly executed in all test cases as code blocks. Therefore, the size of some blocks can be too large for some programs, which has a negative impact on the accuracy of SFL. We propose a new virtual coverage that can be used with the existing approach shown in Figure 1. Our proposed virtual coverage is created based on the number of executions (execution count reports) of each statement. Since it is based on the number of executions, it is possible to divide the source code into more code blocks than the existing virtual coverage shown in Figure 2. Figure 3 shows the creation of existing virtual coverage and the creation of our proposed virtual coverage. In Figure 3, t_1, t_2, and t_3 represent the three test cases, and $s_1, s_2, ...,$ and s_5 represent statements in the source code. The v_1, v_2, v_3, and v_4 represent the virtual coverages created in each approach. Execution count reports describe counts of execution of each statement at test runtime. In Figure 3, the existing approach divides source code into three code blocks (virtual coverages), and our approach divides source code into four code blocks. The idea of our proposed virtual coverage is that the accuracy of SFL is improved by dividing source code based on execution count information.

We evaluated our proposed virtual coverage on the three projects (Math, Lang, Chart) available on Defects4j[5]. As a result, we confirmed that the proposed virtual coverage improves the accuracy by up to 4.2 points compared to the existing virtual coverage. Applying the Wilcoxon Signed-Rank Test to the experimental results, we confirmed that the proposed approach is significantly more accurate than the existing approach.

The rest of the paper is organized as follows. Section 2 describes the background of this paper. Section 3 describes the proposed approach. Section 4 describes the experimental setup and Section 5 discusses the experimental results. We conclude in Section 6.

2 BACKGROUND

2.1 Statistical SFL Techniques

Tarantula[4] and Ochiai[3] are representative statistical SFL techniques. These techniques use test coverage collected during test execution to compute suspiciousness scores, which indicate the suspiciousness of failure for each statement. Several approaches[2][8][13] have been proposed to compute suspiciousness scores similar to Tarantula and Ochiai.

This paper discusses deep learning-based SFL approaches, which differ from statistical SFL approaches in the idea of fault localization. In this paper, statistical SFL techniques are not further discussed.

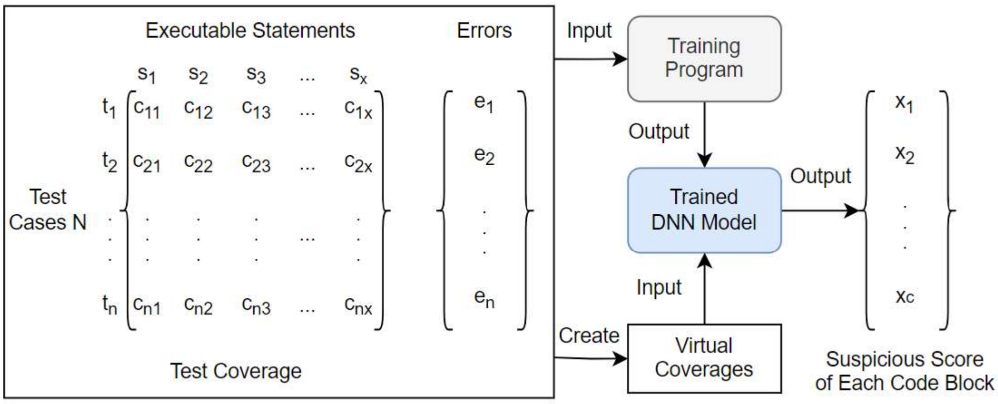


Figure 1: Existing Approach of DNN-based SFL

$$
\text{Virtual Coverage}\quad
\begin{array}{c}
 \\ v_1 \\ v_2 \\ v_3
\end{array}
\begin{array}{c}
\text{Statements} \\
\begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\
\left\{\begin{array}{ccccc}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}\right\}
\end{array}
$$

Figure 2: Virtual Coverage at Code-Block Granularity

2.2 Deep Learning-based SFL Approaches

The most relevant researches to this paper are approaches[14][16][7] that use virtual coverage to locate faults. In these approaches, a DNN model is first trained that takes test execution coverages as input and classifies whether the test result of the input coverage is Pass or Fail. Next, virtual coverage is constructed that indicates that only certain a statement or a code block is executed. Several approaches have been proposed for constructing virtual coverage, including the approach[16] at the statement granularity and the approach[7] at the code block level. The statement granularity virtual coverage is shown in Figure 4 and the code block granularity virtual coverage is shown in Figure 2. By inputting the virtual coverage in Figures 4 and 2 into a trained DNN model, a failure suspiciousness score is given for each virtual coverage. Since each virtual coverage is a coverage that represents only certain a statement or a code block executed, output values of a DNN model are treated as suspiciousness scores for each statement or code block. It is known experimentally that using virtual coverage at the code block granularity is more accurate for SFL than virtual coverage at the statement granularity[7].

Existing virtual coverage treats statements that are executed at least once in common in all test cases as code blocks. In this approach, the source code is divided into code blocks, which improves the accuracy to the limit of coverage-based SFL. Since the existing approach constructed virtual coverage on a coverage basis, there are possible cases where statements are aggregated in some blocks and the sizes of the blocks are too large. Since our approach constructs virtual coverage based on the number of executions, it is expected to improve the accuracy compared to the existing approach.

3 OUR APPROACH

An overview of the fault localization in our approach is shown in Figure 5 and Figure 6. Fault localization in our approach consists of the following steps.

1. Input test input values to System Under Test (SUT) and execute tests.

2. Collect a Coverage Report for each test case at test run time and generate an execution count report.

3. Label the test results (Pass or Fail) in the generated execution count report.

4. A DNN model is trained using code coverage informations as input data and test result labels as supervisor data.

5. Create virtual coverage from the execution count report.

6. Virtual coverage is input to a trained DNN model and a suspiciousness score is given to each code block.

7. Treat the descending order of the suspiciousness score of each code block as the rank of suspiciousness.

Figure 5 shows Steps 3 and 4; Figure 6 shows Steps 6 and 7. Figure 7 shows the DNN architectures. The coverage report in Step 2 describes information such as how many times each statement of the SUT is executed during each test case run. An overview of the execution count report in Step 2 is shown in Figure 8. The execution count report shown in Figure 8 indicates how many times each statement of the SUT is executed in each test case. Code coverage is represented as 1 if a statement is executed at least once and 0 if it is not. On the other hand, in the execution count report, each statement is represented by the number of execution at test runtime.

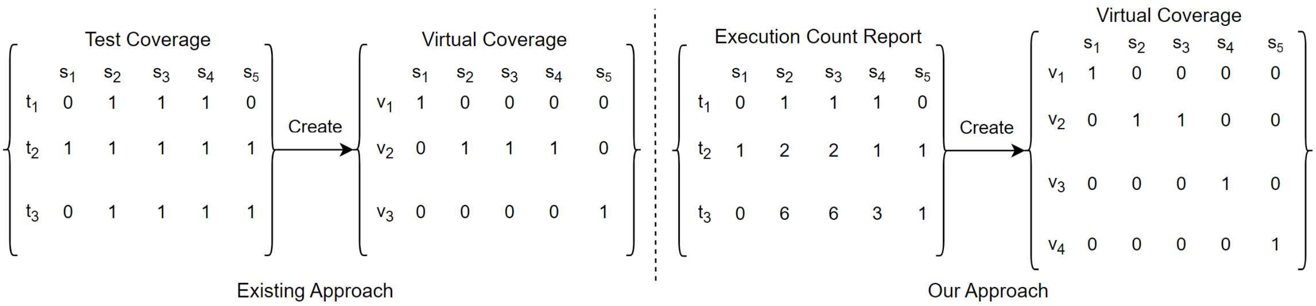


Figure 3: Virtual Coverage Creation in Each Approach

$$
\text{Virtual Coverage}\quad
\begin{array}{c}
 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5
\end{array}
\begin{array}{c}
\text{Statements} \\
\begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 4: Virtual Coverage at Statement Granularity

In step 4, the developer can use a DNN model that takes the code coverage informations as input and classifies the test results of the input reports. The output value of this DNN model is a float value between 0 and 1, which is regarded as the probability that the classification result is a Fail. A DNN model learns the SUT's code coverage informations and learns the difference in pattern between Pass test results and Fail test results.

The next step is to create virtual coverages that will be input to a trained DNN model. Virtual coverage is code coverage that virtually represents that only a certain code block is executed. The developer creates a virtual coverage and inputs it to a DNN model. Since the output of a DNN model is considered to be the probability that the classification result is Fail, the output value of a DNN model when a virtual coverage containing the faulty statement is input is expected to be higher than other input values. Therefore, the output values of a DNN model in descending order are treated as the rank of the final suspiciousness score, and a suspiciousness rank is assigned to each code block.

The basic idea of SFL (Steps 4, 6 and 7) is the same as the existing approach[7] and is not a new contribution of this paper. We propose a DNN-based virtual coverage fault localization approach based on execution count information. Our proposed approach is used in Steps 5. The details of our proposed approach are described below.

3.1 Create Virtual Coverage with Execution Count Reports

The following is a step-by-step description of how to create virtual coverage using execution count reports.

1. In each test case, statements that are adjacent and have the same number of executions shall be temporary blocks.
2. In all test cases, statements contained in a common temporary block are defined as code block.
3. Based on the code blocks defined, create virtual coverage that represents that only certain code block is executed.

An overview of each Step is shown in Figure 9. In Figure 9, the dotted squares are temporary code blocks and the blue squares are the final code blocks to be defined. In Step 1, temporary code blocks are defined in each test case. In test case 1 (t_1), three temporary code blocks are defined, and in t_2, four temporary code blocks are defined. Each temporary code block is assigned a block id. The color of the dotted line in Step 1 indicates the id of the temporary code block. For example, statement 1 (s_1) belongs to the same temporary code block with the same id in all test cases.

Next, Step 2 defines the final code block from the temporary code block. In Step 2, statements that belong to the temporary code block with the same id in all test cases are defined as the final code block. Since s_1 belongs to the temporary code block with the same id (color) in all test cases, we define it as the final code block (blue square). Next, since s_2, and s_3 belong to the same id (color) in all test cases, they are defined as the final code block. At this time, the temporary code block id of s_4 in t_1 is changed to the same id (color) as the next statement, s_5. Next, in Step 2-2, s_4 is defined as the final code block, and the temporary code block id of s_5 in t_1, t_3 is changed to the next temporary code block id (yellow). Finally, in Step 2-3, s_5 is defined as the final code block, and four code blocks are defined in the execution count report shown in Figure 9.

Our proposed virtual coverage treats statements that are common execution count patterns in all test cases as code blocks. Since code blocks can be created according to the actual execution patterns, source code can be divided into more code blocks than the existing approach. The difference between our proposed code block and the code block of the existing approach[7] is shown in Figure 10.

In Figure 10, our approach creates four code blocks, while the coverage-based existing approach creates three code blocks. By creating code blocks according to the pattern of execution counts, code blocks can be created with a finer granularity than the existing approach.

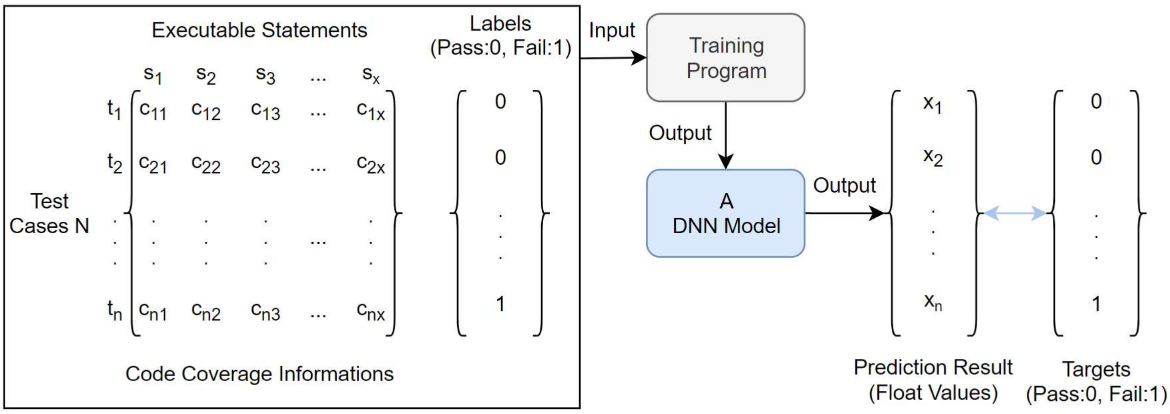


Figure 5: Training A DNN Model in Our Approach

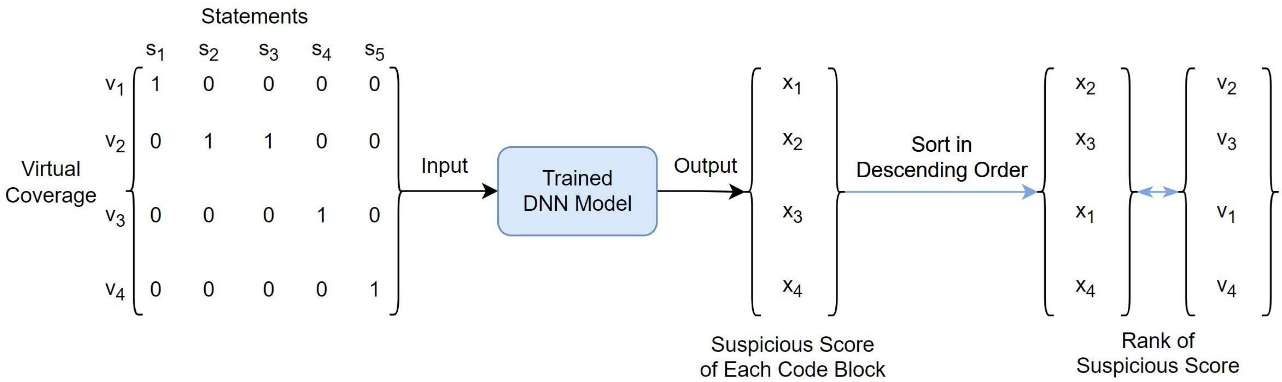


Figure 6: Calculating Suspicious Scores in Our Approach

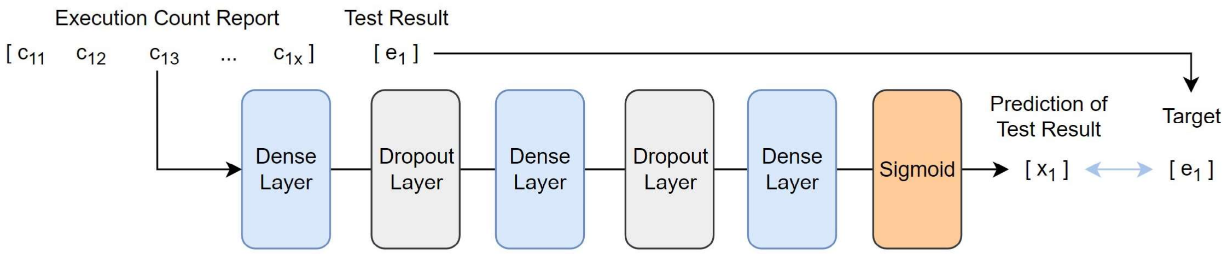


Figure 7: DNN Architecture in Our Approach

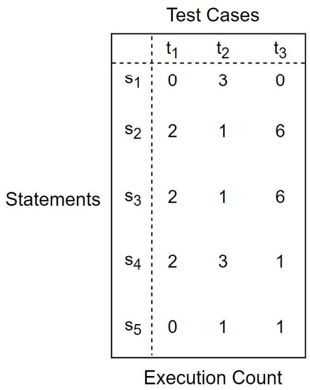


Figure 8: Execution Count Report

4 EVALUATION EXPERIMENT

4.1 Research Questions

In this paper, the following research question is investigated in the evaluation experiment.

RQ1. Comparison of fault localization accuracy with existing approach[7] when virtual coverages created based on the number of executions are input to a DNN model.

In this paper, we propose a new approach for creating a new virtual coverage. We evaluate the effectiveness of our proposed virtual coverage by comparing its fault localization accuracy with existing coverage-based virtual coverage.

The TopN % is used as a measure of the accuracy of fault localization, where the TopN % represents that a bug is classified into the top N % of the total, and a smaller value indicates a better fault localization performance. In the case of multiple bugs, the largest TopN % is used.

4.2 Subject Programs

We conducted an evaluation experiment using bugs and their fixes provided by Defects4j[5]. Defects4J is a database of actual bugs in Java projects. Lang, Math, and Chart from Defects4j are selected as the projects for the experiments. We use bugs that meet the following conditions for our experiments.

- Bugs are not fixed by code addition only: If the bug is fixed by code addition only, the original buggy source code does not have any defects to be pointed out.

- The fix (the defect to be pointed out) is executed at least once in each of the fail and pass tests: because the SFL approaches require the bug to be executed in both the fail and the pass tests.

Execution coverage and execution count reports are collected using OpenClover[1]. Because execution is not recorded for class member variable definitions, etc., due to OpenClover's specifications, we excluded from the fault set the parts of the program that are not recorded as execution coverage. The number of lines (LOC) and number of tests for the experimental program are shown in Table 1.

Table 1: Details of Target Programs

Project	Number of Versions	LOC	Number of Tests
Lang	30	58389	54987
Math	29	23623	83364
Chart	15	10094	27036

4.3 Setup for a DNN Model

In performing supervised training, the weight parameters of Dense Layers are initialized with random values. The size of the hidden layer of each Dense Layer is changed according to the size of the coverage size to be trained. The first Dense Layer is the minimum size of 100, and the second Dense Layer is the minimum size of 20 hidden layers. The third Dense Layer is the Output Layer, which is a hidden layer of size 1. The ReLu and Sigmoid functions are used as activation functions. In addition, a Dropout Layer is used to suppress over-learning. Adam optimizer[6] is used, and the learning rate is set to $1.0e-3$. TensorFlow (ver. 2.12.0) is used as the learning framework.

5 RESULTS AND DISCUSSIONS

5.1 RQ1: Result

Figure 11 shows the experimental results for RQ1. The horizontal axis in Figure 11 represents the TopN % and indicates the amount of source code examined by the developer. The vertical axis shows the percentage of identified faults, where 100 % on the vertical axis means that all faults are identified. For example, a vertical axis plot with a Top 50% is more than 90%, indicating that more than 90% of the faults are identified by investigating half of the source code. There are two plots in Figure 11: the gray plot shows the accuracy of the existing approach, and the orange plot shows the accuracy of the proposed virtual coverage.

As a result, except for the 25% and 35% TopN % plots, the proposed virtual coverage has a larger value in the vertical axis plot. This means that for each TopN %, the number of faults that can be identified is higher with the proposed virtual coverage than with the existing approach.

Table 2 shows the results of adapting the Wilcoxon Signed-Rank Test to the experimental results. The Wilcoxon Signed-Rank Test is a nonparametric test that tests for differences between two corresponding groups. The null hypothesis indicates that there is no significant difference between the two groups, while the alternative hypothesis indicates that there is a significant difference between the two groups. The 1-tailed test and the alternative hypotheses are listed below.

- **1-tailed (left):** The proposed approach has a smaller TopN % value than the existing approach.

- **1-tailed (right):** The proposed approach has a larger TopN % value than the existing approach.

Since the value of TopN % is the amount of code investigated by the developer to identify the fault location, a smaller value indicates a high accuracy. Therefore, if the left-tailed

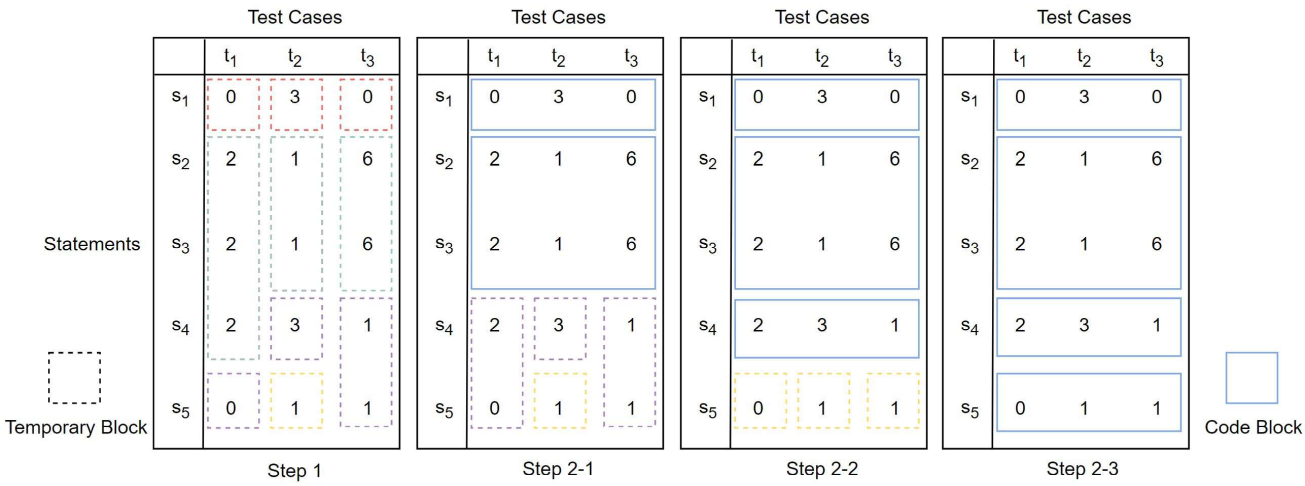


Figure 9: Steps in The Creation of Our Proposed Code Block

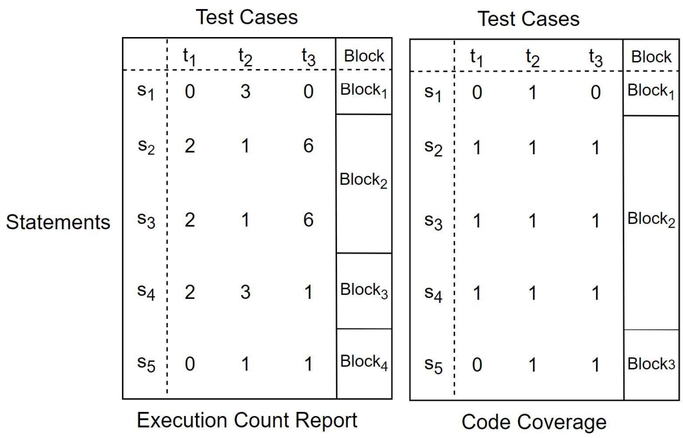


Figure 10: Differences in Code Blocks

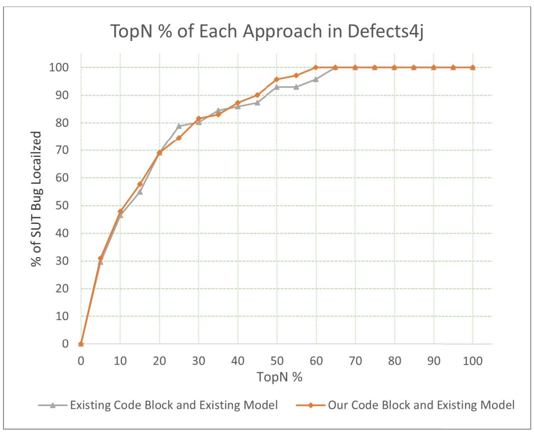


Figure 11: Results in RQ1

Table 2: Test Results in RQ1

Subject	1-tailed (left)	1-tailed (right)
Defects4J	3.534E-02	9.647E-01

Wilcoxon Signed-Rank test is accepted, the proposed approach is significantly better than the existing approach in fault localization performance.

The test results of Defects4J show that $p = 0.035 < 0.05 = \alpha$, so the proposed virtual coverage is significantly more accurate than the existing approach.

5.2 RQ1: Discussions

Our proposed virtual coverage is created based on the number of executions. In our approach, the source code is divided into many more code blocks than the existing approach. We consider that defining code blocks with a finer granularity than the existing approach according to the actual execution pattern (number of executions) contributed to the improved accuracy shown in Figure 11.

6 CONCLUSION

In this paper, we propose a new virtual coverage to be used for existing DNN-based SFL approaches. Our proposed virtual coverage is created based on the number of executions and is able to divide the source code with finer granularity than the existing virtual coverage.

Our approach is evaluated with three projects available on Defects4j. Experimental results show that our proposed virtual coverage is more accurate than existing virtual coverage.

Further evaluation of our approach in broader and larger SUTs is needed and is a topic for future work. In the future, we intend to improve a DNN model in our approach to identify faults that cannot be addressed (e.g., faults for which the only bug fix is to add code, performance bugs, etc.).

ACKNOWLEDGMENT

Part of this work is supported by fund from Mitsubishi Electric Corp. The research is also being partially conducted as Grant-in-Aid for Scientific Research C (21K11826).

REFERENCES

[1] OpenClover. https://openclover.org/

[2] Abreu, R., Zoeteweij, P., van Gemund, A.J.: On the accuracy of spectrum-based fault localization. In: Testing: Academic and Industrial Conference Practice and Research Techniques - MUTATION (TAICPART-MUTATION 2007). pp. 89–98 (2007). https://doi.org/10.1109/TAIC.PART.2007.13

[3] Abreu, R., Zoeteweij, P., Van Gemund, A.J.: An evaluation of similarity coefficients for software fault localization. In: 2006 12th Pacific Rim International Symposium on Dependable Computing (PRDC'06). pp. 39–46 (2006). https://doi.org/10.1109/PRDC.2006.18

[4] Jones, J., Harrold, M., Stasko, J.: Visualization of test information to assist fault localization. In: Proceedings of the 24th International Conference on Software Engineering. ICSE 2002. pp. 467–477 (2002). https://doi.org/10.1145/581396.581397

[5] Just, R., Jalali, D., Ernst, M.D.: Defects4j: A database of existing faults to enable controlled testing studies for java programs. In: Proceedings of the 2014 International Symposium on Software Testing and Analysis. pp. 437–440. ISSTA 2014, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2610384.2628055, `https://doi.org/10.1145/2610384.2628055`

[6] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), `http://arxiv.org/abs/1412.6980`

[7] Kiryu, H., Ogata, S., Okano, K.: Improve measuring suspiciousness of bugs in spectrum-based fault localization with deep learning. In: Proceedings of International Workshop on Informatics. pp. 3–8. IWIN '22, Informatics Laboratory, Kii-Katsuura, Japan (2022)

[8] Masri, W.: Fault localization based on information flow coverage. Software Testing, Verification and Reliability **20**(2), 121–147 (2010). https://doi.org/https://doi.org/10.1002/stvr.409, `https://onlinelibrary.wiley.com/doi/abs/10.1002/stvr.409`

[9] Perez, A., Abreu, R.: A qualitative reasoning approach to spectrum-based fault localization. In: Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings. pp. 372–373. ICSE '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3183440.3195015, `https://doi.org/10.1145/3183440.3195015`

[10] Sarhan, Q.I., Beszédes, A.: A survey of challenges in spectrum-based software fault localization. IEEE Access **10**, 10618–10639 (2022). https://doi.org/10.1109/ACCESS.2022.3144079

[11] de Souza, H.A., Chaim, M.L., Kon, F.: Spectrum-based software fault localization: A survey of techniques, advances, and challenges (2016), `http://arxiv.org/abs/1607.04347`

[12] Tassey, G.: The economic impacts of inadequate infrastructure for software testing (2002)

[13] Wong, W.E., Debroy, V., Gao, R., Li, Y.: The dstar method for effective software fault localization. IEEE Transactions on Reliability **63**(1), 290–308 (2014). https://doi.org/10.1109/TR.2013.2285319

[14] Wong, W.E., Debroy, V., Golden, R., Xu, X., Thuraisingham, B.: Effective software fault localization using an rbf neural network. IEEE Transactions on Reliability **61**(1), 149–169 (2012). https://doi.org/10.1109/TR.2011.2172031

[15] Zhang, Z., Lei, Y., Mao, X., Yan, M., Xu, L., Wen, J.: Improving deep-learning-based fault localization with resampling. J. Softw. Evol. Process **33**(3) (mar 2021). https://doi.org/10.1002/smr.2312, `https://doi.org/10.1002/smr.2312`

[16] Zhang, Z., Lei, Y., Mao, X., Yan, M., Xu, L., Zhang, X.: A study of effectiveness of deep learning in locating real faults. Information and Software Technology **131**, 106486 (2021). https://doi.org/https://doi.org/10.1016/j.infsof.2020.106486, `https://www.sciencedirect.com/science/article/pii/S0950584920302287`

A Spherical POV Heatmap using Mixed Reality in 360-degree Internet Live Broadcasting

Yoshia Saito* and Junichiro Suto*

*Graduate School of Software and Information Science, Iwate Prefectural University, Japan
y-saito@iwate-pu.ac.jp

Abstract - 360-degree internet live broadcasting enables viewers to change their point of view (POV) while watching the 360-degree live video and has an issue that the broadcaster cannot check the viewers' POV. To solve the issue, we have proposed a spherical POV heatmap using Augmented Reality (AR) on a smartphone so that the broadcaster can be aware of the viewers' POV. Although the spherical POV heatmap reduced communication errors, the response time increased due to the need to check the heatmap on a smartphone. In this paper, we propose a new spherical POV heatmap using Mixed Reality (MR) through an MR headset. The proposed system displays the heatmap in real space through the MR headset and reduces the response time by eliminating the need to check the smartphone. We implemented a prototype system and found it could reduce the response time through the evaluation.

Keywords: 360-degree Internet Live Broadcasting, Viewers' POV, Mixed Reality.

1 INTRODUCTION

360-degree video frees viewers from the constraints of viewing direction. The viewers can change their point of view (POV) and enjoy all directions of the video. Recently, 360-degree video has also been introduced in internet live broadcasting, where the viewers can typically communicate with the broadcaster via text chat. However, there is an issue that the broadcaster cannot check the viewers' POV in the 360-degree internet live broadcasting. Since the broadcaster does not know what the viewers saw and commented on, it can cause communication errors between the broadcaster and the viewers.

To solve the issue, we have proposed a spherical POV heatmap using Augmented Reality (AR) technology [1-2] (hereinafter called AR spherical POV heatmap). The AR spherical POV heatmap displays a spherical heatmap of the viewers' POV on a QR code using AR technology through a smartphone screen. The broadcaster checks the AR spherical POV heatmap and can know which direction the viewers are watching. Although it can reduce communication errors between the broadcaster and the viewers, it increases the response time to viewer comments due to the need to check the heatmap on the smartphone. The response time should be short to realize smooth communication between the broadcaster and viewers.

To overcome the disadvantage of the AR spherical POV heatmap, we propose a new spherical POV heatmap using Mixed Reality (MR) technology (hereinafter called MR spherical POV heatmap). The MR spherical POV heatmap displays a spherical heatmap of the viewers' POV on top of the omnidirectional camera in real space using the MR technology through an MR headset. Since the broadcaster can check the MR spherical POV heatmap in real space, it is expected that the response time to viewer comments can be shortened.

The contributions of this paper are summarized as follows:

- We developed and evaluated a prototype system of the MR spherical POV heatmap using HoloLens 2.
- We clarified that the MR spherical POV heatmap enabled the broadcaster to reduce the response time to viewer comments.

The rest of this paper is organized as follows. Section 2 describes the AR spherical POV heatmap in the 360-degree internet live broadcasting as our previous work. Section 3 describes an overview of our proposed system. Section 4 describes the implementation of the prototype system. Section 5 describes an evaluation experiment to clarify the effects of the proposed system. Section 6 summarizes this study.

2 AR SPHERICAL POV HEATMAP

The AR spherical POV heatmap shows a sphere that represents the broadcasting space on an AR marker through a smartphone as shown in Figure 1. The spherical heatmap visualizes the viewers' POV by displaying the angular coordinate vector on the spherical surface as a heatmap. By synchronizing the direction of the image taken by the omnidirectional camera with the vector of the spherical heatmap, the broadcaster can grasp the viewers' POV in real space simply by checking the heatmap. In addition, the heatmap is easy to visualize multiple data at the same time. It is also possible to check the POV of multiple viewers. By checking the density pattern of the viewers' POV on the heatmap, the broadcaster can grasp where many of the viewers are interested, and where the other viewers of the minority are watching. Therefore, it gives some hints about the viewers' interests and achieves smooth communication between the broadcaster and viewers.

In the evaluation experiments, the AR spherical POV heatmap could reduce communication errors by visualizing the viewers' POV and the speaking timing is easier for the broadcaster to grasp when waiting for the viewers' attention. On the other hand, the time it takes for the broadcaster to respond to the viewer's comment was increased by

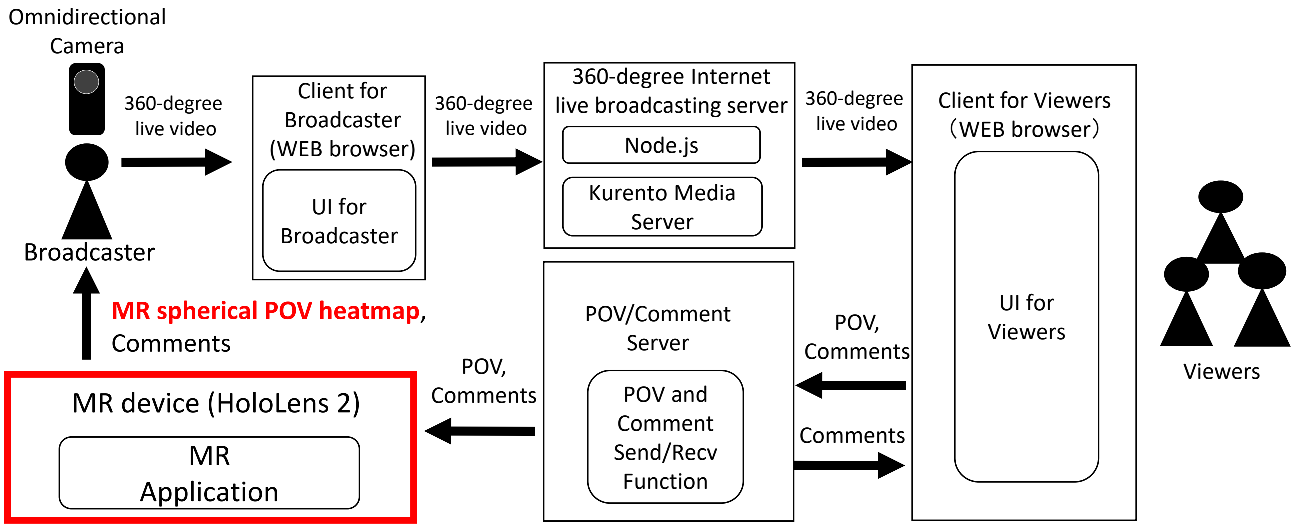


Figure 3: System architecture of the prototype system

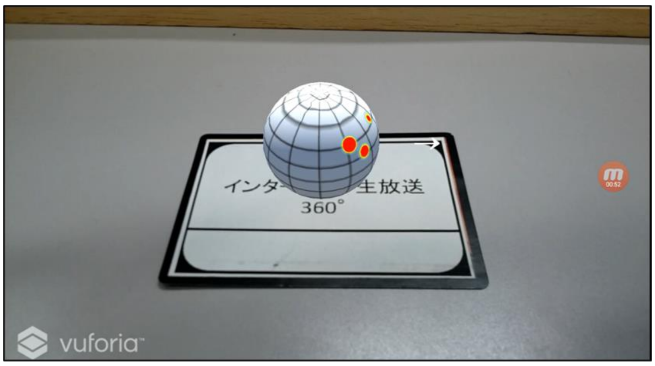


Figure 1: AR spherical POV heatmap through smartphone display

approximately 10 seconds compared with the broadcasting without the AR spherical POV heatmap. This result was due to the time required to check the heatmap through a smartphone. If the wasted time can be eliminated, the effectiveness of the spherical POV heatmap can be improved.

3 PROPOSED SYSTEM

In our previous study [1-2], the issue was that checking the viewers' POV was a burden for the broadcaster. This is because the spherical POV heatmap was implemented using AR and forced broadcasters to check their smartphone to see the heatmap. To solve this issue, we propose a system that displays the spherical POV heatmap using MR technology through an MR headset. The proposed system is expected to eliminate the need to confirm the POV heatmap using a smartphone in the previous study. The use of a head-mounted type MR device also solves the issue of both hands being occupied and reduces the burden on the broadcaster in 360-degree internet live broadcasting.

Figure 2 shows a system model of the proposed system. The proposed system consists of a 360-degree internet live broadcasting system and an MR spherical POV heatmap function. The broadcaster sends a 360-degree live video to the proposed system and the viewers watch the live video. The viewers send comments to the proposed system, which are sent to the broadcaster and the viewers. The proposed system receives viewers' POV in real-time and makes an MR spherical POV heatmap. The MR spherical POV heatmap is provided to an MR headset of the broadcaster.

The proposed system is expected to have several advantages by reducing the burden on the broadcaster as follows: (1) the time it takes for the broadcaster to respond to the viewer's comment is expected to be shorter than the previous study with few communication errors, (2) the speaking timing is easier for the broadcaster to grasp than the previous study when waiting for the viewers' attention. The advantages of the previous study will be further enhanced, and the disadvantage is solved by the proposed system.

4 IMPLEMENTATION

We implemented a prototype system of the MR spherical POV heatmap using HoloLens 2. In this section, we describe the architecture of the prototype system and its main application.

4.1 System Architecture

The prototype system was implemented by replacing the AR spherical POV heatmap in the previous study with an MR application for the MR spherical POV heatmap. Figure 3 shows the system architecture of the prototype system. The red square in the figure shows new implementation in this study and the other parts are diverted from the previous study. In the prototype system, we used Microsoft HoloLens 2 as an MR headset for the broadcaster and THETA V as an omnidirectional camera.

A broadcaster can start 360-degree internet live broadcasting using the broadcaster client on a web browser.

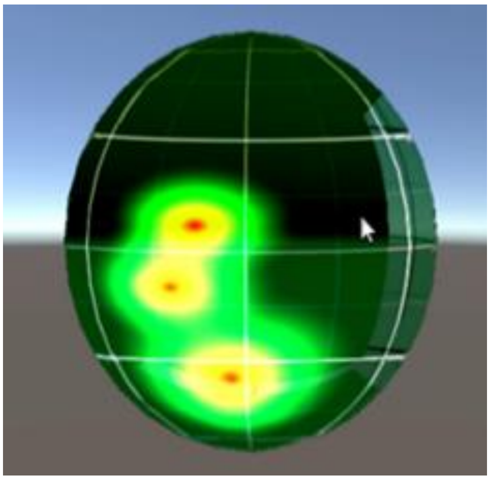


Figure 4: The heatmap when multiple viewers are watching from different POV

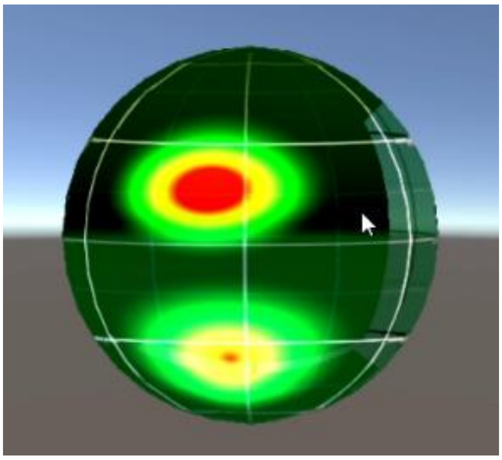


Figure 5: The heatmap when several viewers are watching in a particular POV

Figure 6: The user interface for the broadcaster through the HoloLens 2

The 360-degree internet live broadcasting server distributes it to viewers. The viewer can watch the 360-degree live video using the viewer client on a web browser and send text comments to the comment server. The comment server forwards the received comments to the viewers and the broadcaster. The viewers can read the comments on their web browsers and the broadcaster can read the comments on the HoloLens 2. The viewer client also automatically sends the viewer's POV to the POV server every second. The POV is represented by spherical coordinates (r, θ, φ) where r denotes the radial distance, θ denotes the polar angle, and φ denotes the azimuthal angle. The POV server forwards the received viewers' POV to the MR application on the HoloLens 2.

4.2 MR Spherical POV Heatmap

The MR application on the HoloLens 2 is implemented by Unity and it shows a spherical POV heatmap and comments to the broadcaster. The MR application receives the viewers' POV and displays their heatmap on a sphere object which is shown above the omnidirectional camera. Based on a list of the viewers' POV received within a certain period, it creates circles for the heatmap. In this circle, the closer to the center, the higher the heat value, and the farther from the center, the lower the heat value. If a circle overlaps another circle, the heat values of the two circles are added and the overlapping area has a higher heat value.

Figure 4 shows an example of the heatmap when multiple viewers are watching the 360-degree live video from different POV. The broadcaster can see that the viewers' interests are dispersed in different directions. Figure 5 shows an example of the heatmap when several viewers' POV are concentrated. The broadcaster can see that several viewers are interested in a particular direction of the size of the red area.

Figure 6 shows the user interface for the broadcaster through the HoloLens 2. The MR application gets a coordinate and direction of the omnidirectional camera, and it displays the MR spherical POV heatmap which is fixed above the omnidirectional camera. The comment window tracks the broadcaster's sight so that the broadcaster can read the viewers' comments. The MR spherical POV heatmap is updated every second. The broadcasters can walk freely around the room wearing hololens2.

5 EVALUATION

We conducted an evaluation experiment using the prototype system compared with the AR spherical POV heatmap. In this section, we describe the evaluation environment and the results.

5.1 Environment

The experiment was conducted under the same conditions as in the previous study [2]. The purpose of the experiment was to confirm whether the MR spherical POV heatmap reduced the burden on the broadcaster compared to the AR spherical POV heatmap.

Each experiment was conducted with one broadcaster and three viewers, for a total of four times. The broadcaster and viewers were in different rooms in the experiment. The equipment used in the experiment was one notebook PC for the broadcaster, three laptop PCs for the viewers, one PC for the server, Ricoh Theta V for the omnidirectional camera, and HoloLens 2 for the MR headset.

The content of the broadcast was a chat about objects in the broadcaster’s room. In the room, there were various objects. The viewers commented on objects in the broadcaster’s room freely changing their POV. The procedure of the experiment is as follows:

- [Practice] The broadcaster and viewer practice the operation of the prototype system for 5 minutes.
- [Task 1] One of the viewers sends a designated question comment to the broadcaster 4 times at 3-minute intervals.
- [Task 2] The broadcaster directs the viewer's attention to a specified object and chats about it 3 times at 2-minute intervals.

In Task 1, the viewer communicates to the broadcaster by sending a question comment about an object in the room, such as "What kind of animal is this stuffed animal?". The text to be sent, the timing of the comment, and the object to which the viewer is directed are specified in the procedure manual. The broadcaster determines which object the viewer is commenting on and responds to the viewer's comment. If the broadcaster understands what the viewer is commenting on, the time it takes for the broadcaster to respond correctly to the viewer’s comment is expected to be shorter. We measure the time and define it as the *response time*. We also count the *number of communication errors* if the broadcaster makes a mistake with the object the viewers are talking about.

In Task 2, the broadcaster points to a specified object and instructs the viewers "Please look at this". The broadcaster starts chatting with the viewers about the object when their POV are gathered. If the broadcaster is aware of the viewers’ POV, it is expected to be able to start chatting at the same time when the viewer's POV is gathered. We measure the time between the broadcaster's attention instruction to the viewers and when the broadcaster begins to speak. It is defined as the *wait time*.

5.2 Results

In Task 1, the mean response time was 21.47 seconds for the MR spherical POV heatmap, compared to 30.88 seconds for the AR spherical POV heatmap. In comparison, the MR spherical POV heatmap reduced the response time by about 9 seconds. In terms of the number of communication errors, there were 4 communication errors in the MR spherical POV heatmap, while there was no communication error in the AR spherical POV heatmap. This is because some broadcasters selected the target object without checking the MR spherical POV heatmap at the first time. The number of communication errors could be reduced to zero if checking the MR spherical POV heatmap.

In Task 2, the mean wait time was 14.78 seconds for the MR spherical POV heatmap, compared to 16.67 seconds for the AR spherical POV heatmap. In comparison, the MR spherical POV heatmap shortened the wait time by approximately 2 seconds. Meanwhile, we conducted a Mann-Whitney U test for the results and there was no difference between the mean wait time (p = 0.20402 > 0.05). The MR spherical POV heatmap was not able to reduce the wait time although we expected to be able to reduce the time by eliminating the need to check the smartphone as with the previous result of the response time. One possible reason for this result is that when several heatmap circles were gathered into one heatmap circle, it was difficult to determine how many viewers were in the heatmap.

Furthermore, the broadcaster was asked to freely describe what they found advantages and what they found disadvantages about the prototype system. As an advantage point, many participants said that "it is easy to visually understand where the viewer is looking". As a disadvantage point, many participants said that "when the viewers’ POV is focused on one place, the heatmap becomes one circle, and it is difficult to grasp how many viewers are there". There was also an opinion that "the conversation is delayed by one step because we have to look at the POV heatmap before speaking". This is because it is necessary to check the MR spherical POV heatmap above the omnidirectional camera before finding the target object. To solve this problem, instead of displaying the heatmap on a sphere, a method may be effective in which the heatmap is displayed directly on the target object.

6 CONCLUSION

We proposed the MR spherical POV heatmap to reduce the burden on the broadcaster in 360-degree internet live broadcasting, which displayed a spherical POV heatmap on the MR space. We compared it with our previous study, the AR spherical POV heatmap to clarify its advantages and issues. The results of evaluation experiments showed that the MR spherical POV heatmap reduced the response time compared to the AR spherical POV heatmap, and the objects could be found quickly. In addition, the use of the MR headset eliminated the need to check the smartphone and reduced the burden on the broadcaster. On the other hand, the wait time could not be reduced. One of the reasons for this result was that it was difficult to grasp how many viewers were there on the MR spherical POV heatmap. In future work, we will improve the way the heatmap circles are displayed so that it is more intuitive for the broadcaster to know approximately how many viewers are on the heatmap. Moreover, we will implement a method in which the heatmap is displayed directly on the target object.

REFERENCES

[1] Masaya Takada, Dai Nishioka and Yoshia Saito: Proposal of a Spherical Heat Map in 360-Degree Internet Live Broadcasting Using Viewers' POV, The 2nd International Workshop on Mobile Ubiquitous Systems, Infrastructures, Communications and Applications (MUSICAL2019), pp.596-600 (2019).

[2] Masaya Takada, Dai Nishioka and Yoshia Saito: Proposal of the Viewer’s “POV Heat Map” in 360-degree Internet Live Broadcasting, IPSJ Journal, Vol.62, No.1, pp. 145–159 (2021) (in Japanese).

Were you speaking to me?: A trial to use physical avatars to establish gaze awareness in hybrid meetings[1]

Kenta Ohnaka*, Taai Tsukidate*, Yo Kuwamiya*, Kazuyuki Iso**, and Minoru Kobayashi***

*Frontier Media Science Program, Graduate School of Advanced Mathematical Sciences, Meiji University, Japan
**Tokyo Information Design Professional University, Japan
***Department of Frontier Media Science, School of Interdisciplinary Mathematical Sciences, Meiji University, Japan
{Kenta.Ohnaka, Taai.Tsukidate, Yo.Kuwamiya}@koblab.org
iso@tid.ac.jp, minoru@acm.org

Abstract -Hybrid meetings involving co-located and remote participants have been facing problems such as presence asymmetry between participants at each site, and difficulty in participants understanding what is happening at each site This is problematic because if the remote participants are not well represented, the meeting may proceed only with the local participants. To overcome these limitations, we propose a meeting environment that includes a bird's-eye view video and physical avatars. This paper provides visual representations of meeting environments using two types of physical avatar, a simple human-like object, and a telepresence robot, kubi, combined with the use of a bird's-eye view video. Through the trials we conducted, we gained an insight that the combination of the physical avatars and the bird's-eye view has the potential to help the remote participants understand the spatial arrangement of the main meeting site including who is being spoken to in the main site. The trials also demonstrated that the design of the avatar was important for participants at both sites.

Keywords: Video Communication, Telepresence, Gaze A~~a~~wareness, Physical A~~a~~vatar

1 INTRODUCTION

Advancements in technology have allowed remote communication on a daily basis. The last few years in particular have seen remote working promoted around the world, and we increasingly communicate with friends and colleagues at a distance.

Decades of research have attempted to address problems in telecommunications. In particular, difficulties in "hybrid meetings" involving both co-located and remote participants include the asymmetric presence between participants, and turn-taking issues when participants lack an understanding of what is happening in each site.

In order to solve problems with these types of hybrid meeting environment, we consider it important to reduce the asymmetry of presence between co-located and remote participants, and to convey information that would generally be available to participants in face-to-face meetings. Such information includes the atmosphere of the meeting room, the way participants look at the screen, and the way they speak to other participants. To this end, we propose a meeting design that involves using physical avatars at the main meeting site as a surrogate for the remote participants, and conveying a bird's-eye view video of the main site, including the avatar, to remote participants. We seek answers to the following question: Does a bird's-eye view video that includes avatars of remote participants facilitate hybrid meetings? This paper describes trials for this purpose in two types of environments and discusses the findings, which suggest that a bird's-eye view video can help the remote participants understand the spatial arrangement of the main site, and the physical avatar can act as a surrogate for the remote participant.

2 RELATED WORKS

Researchers have explored the embodiment of remote participants in various ways in remote meetings, especially in hybrid meetings. For example, Kuzuoka et al. [1] proposed a telepresence robot that gestured and pointed to remote participants, and Sakamoto et al. [2] proposed a humanoid robot (i.e., android) with a realistic presence. Making a physical embodiment of remote participants in this way made it possible to improve their experience of the meeting. Sellen et al. [3] proposed a tabletop video conferencing system to enable cross-referencing between participants. Equally, today's remote meetings make effective use of telepresence robots such as Beam [4], Double [5], and kubi [6].

The physical embodiment of remote participants makes both remote and co-located participants feel that those attending remotely are physically participating at the main site. In addition, telepresence and humanoid robots are excellent at conveying the status of remote participants to those at the main site. However, users of these systems are usually limited in their field of view, and it is difficult to grasp the situation at the main site when avatars are used as surrogates.

[1] This work was supported by JSPS KAKENHI Grant Numbers JP18K11410, JP22H03635.

The telepresence robot proposed by Jouppi et al. [7] and Morita et al. [8] has a camera with a wide field of view, which makes it easy for remote participants to understand the situation at the main site. These studies show that providing remote users with a wide view facilitates interaction, but remote users must operate the robots with a camera to obtain information.

In remote communication using Mixed Reality (MR), which includes the increasingly popular Virtual Reality and Augmented Reality (AR), the use of virtual avatars makes the presence of remote and co-located participants more symmetrical.

Piumsomboon et al. proposed a system called Mini-Me [9] and reported that by displaying virtual avatars in an MR environment and transmitting the gaze and gestures of remote participants, the presence of remote participants could be noticeably improved. In the VROOM [10] system proposed by Brennan et al., a life-sized avatar of the remote participants is presented to the local participants through AR, and remote users have an immersive 360° view of the main site through a Head Mounted Display

In terms of MR, immersive room-sized communication systems such as t-Room [11], proposed by Hirata et al., and Room2Room [12], proposed by Pejsa et al., visualize each participant's non-verbal information, including their gaze, and make participation more symmetrical.

The above examples therefore suggest that using avatars as surrogates for remote participants, whether physical or virtual, can resolve asymmetry in hybrid meetings.

In face-to-face communication, people obtain verbal and non-verbal information from others in a natural way. In particular, the way participants look at each other is important in terms of taking turns, because it conveys the intention of an utterance and the object of attention. The same is true for video or audio-mediated communication, and a number of studies have helped raise awareness of gaze in remote meetings.

GAZE-2 [13], a system proposed by Vertegaal et al., establishes accurate eye contact between participants by conveying their gaze, which is measured by an eye-tracker in a virtual-space video conference. The Mona Lisa effect in video communication is also regarded as a problem. It is caused by the combination of a single camera and a 2D display in the video link, so that participants find it difficult to understand who is talking to whom, and who is looking at whom. To resolve this problem, Iso et al. [14] proposed a 3D display to represent the gaze direction of a remote user. These attempts to establish gaze awareness are approaches common to systems such as Hydra by Sellen et al. [3], as well as methods that use multiple viewpoints and share vision and space remotely, as described below.

A number of researchers have shown how an approach that transmits the camera images from multiple viewpoints is also effective for interaction using non-verbal information, and for sharing the atmosphere of a meeting with remote participants.

In addition to the face-to-face view which is usual in video communication, Gaver et al. [15] designed multiple views showing the whole room, and showing materials related to the task. This enabled the participants to switch freely between views. They showed that the availability of multiple viewpoints was effective in carrying out collaborative tasks remotely. Referring to this example, Yamaashi et al. [16] found that it was important to support peripheral awareness by offering remote participants a fixed global view. In addition, "ESP" [17], a system designed by Venolia et al., contains multiple cameras that can be used in hybrid meetings. Camera images in ESP show co-located participants, their reactions, and the objects they are referring to. The authors reported that the ESP camera helped remote participants understand the context of the conversation and take turns smoothly.

These examples suggest the importance of the views conveyed between participants in remote meetings, and the way views are designed can help develop successful remote communication.

In remote collaboration using MR, more methods have recently been made available for conveying the situation at the main site to remote participants using a 360° live image.

Kasahara et al. [18] enabled remote users to view the full meeting environment by transmitting panoramic images in real time from cameras worn by co-located participants. Lee et al. [19] developed a system to display a virtual hand superimposed on a panoramic image. This allowed all participants to supplement non-verbal communication cues by visualizing where remote participants were pointing and by seeing other participants' views. Teo et al. built on this by [20] proposing a system that combined a 360° view and 3D virtual space. They investigated the various views, suggesting that these helped carry out tasks efficiently and improve the social presence of participants.

These methods give remote participants a very natural view of what is happening at the main site and enable participants to share non-verbal communication cues between sites. However, as these methods demand high-spec devices, they may take longer to spread for common use.

3 METHOD: A BIRD'S-EYE VIEW VIDEO & PHYSICAL AVATARS

As described in Chapter 2, many methods have been studied to facilitate remote meetings, including methods that physically embody remote participants and methods that use multiple viewpoints. However, there are still difficulties for remote participants with regard to understanding the situation, what is happening at the main site, and feeling that the remote participants are being spoken to by the main site participants. To solve these two problems, we used a bird's-eye view video and a physical avatar in combination to make a hybrid meeting environment. We designed the meeting environments shown in Figure 1. In this environment, we expect that using a bird's-eye view video will help a remote participant to understand the situation at the main site. Additionally, we set up the physical avatars to represent the remote participants in the environment with bird's-eye view video. Since we are still in the process of considering what kind of physical avatar would be appropriate, we have prepared two types of avatar.

3.1 Bird's-eye view video

Today, people are very familiar with picture-in-picture devices and the face-to-face view offered by video-meeting services such as Zoom [21] and Microsoft Teams [22]. The effect has already been discussed by pioneers such as [15] [16] [17], but we wish to return to a discussion on the viewpoint of images transmitted to remote participants.

In a face-to-face meeting, people unconsciously perceive the situation in the meeting room surrounding them, as well as peripheral information. The information helps them to sense the atmosphere, understand the context, and take turns smoothly in conversation. This leads to the idea that conveying information involving the whole meeting room could enable remote participants to do these things in hybrid meetings.

We believe that it is important for remote participants to be able to understand that other people are talking to them. The whole view helps with this as it contains information such as what the room looks like, the gaze of the participants co-located at the main site, and the location of objects including avatars. A bird's-eye view video which covers the whole site including the remote participant's avatar enables the remote participant to see that someone is addressing him. In addition, it makes it possible to view both the remote and main site participants simultaneously. We consider that this effect allows co-located participants at the main site to carry out the meeting in a natural way, involving avatars in turn-taking and feeling that they can be seen.

Methods of transmitting a panoramic view from the center

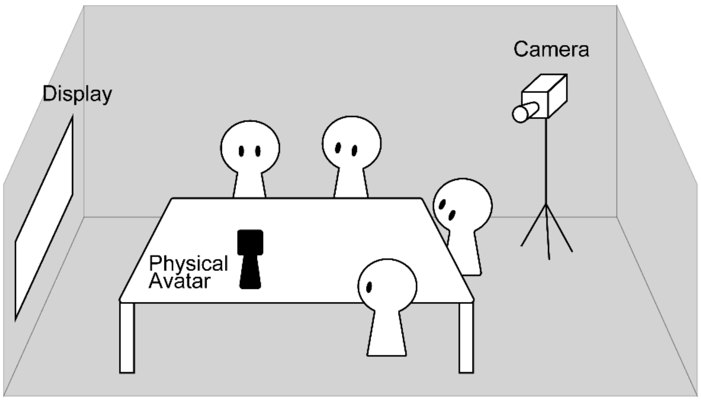


Figure 1: A hybrid meeting environment that includes a physical avatar of a remote participant and a bird's-eye camera to capture the environment.

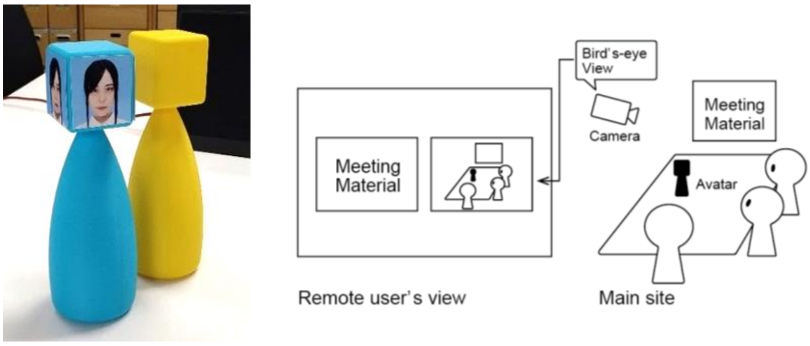


Figure 2: Simple human-like objects (left) and these objects in a meeting environment (right).

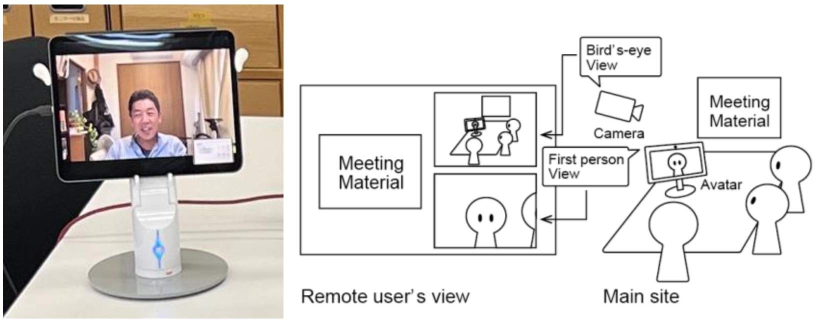


Figure 3: 'kubi' as a telepresence robot (left) and kubi within a meeting environment (right).

of the meeting room, which have become increasingly available recently, may satisfy these requirements. However, it is not always easy to understand the gaze direction or who is looking at whom with a panoramic view image because of the distortion of the image.

3.2 Physical avatar

As research has revealed, the physical embodiment of remote participants in telecommunications has significant implications in terms of making all participants feel that remote participants are present at the main site. In hybrid meetings in particular, it could help make the presence of participants more symmetric.

An important point here is that it is necessary to make the avatar physical. This approach allows participants to handle objects in real space and preserves the atmosphere of a real meeting room. These features make the use of physical avatars very different from situations where avatars are used for all participants in a virtual space, or where virtual avatars of remote participants are shown in a real-space image. Above all, avatars should exist physically if people are to interact naturally with each other.

To design the meeting environments, we prepared two kinds of physical avatar. To represent the remote participants, we first used a simple human-like object (shown in Figure 2), and then kubi [6] as a telepresence robot (shown in Figure 3). As the human-like object did not have any input/output functions such as cameras, microphones, speakers, displays, it acted only as a physical surrogate for the remote participant. Telepresence robots, however, can transmit the voices and images of the remote users to each other. A telepresence robot functions as the eyes, ears, mouth and living portrait of the remote participants.

In the trials described below, we observed the influence of these two kinds of avatar on hybrid meetings when they were used with a bird's-eye view camera.

4 TRIAL USE OF PROPOSED SYSTEM

We tried using the bird`s eye view video and the physical avatar in a hybrid meeting. In this chapter, we report our reflections on this trial use.

4.1 Avatar Type 1: A simple human-like object

We first tested a simple human-like object as an avatar of the remote participant (shown in Figure 4). As mentioned above, this object had no means of communication and did not resemble the remote participant. In this trial, it was difficult for the main site participant to communicate with the remote participant, or to feel that this avatar was a surrogate for him. As seen in Figure 4, it was more natural for main site participants to speak to the remote partner in the video on the screen, rather than the physical avatar that had no visual or audio representation of the remote participant.

On the other hand, the remote participant said that the bird's-eye view video made it easy to understand the situation at the main site. Though it was unnatural for main site participants, we experimented with participants in the main site talking to the physical avatar to see what a bird's-eye view video of talking to the physical avatar would look like. When the remote participant saw the main site participants talking to his physical avatar, the remote participant could temporarily perceive the avatar as if it were his surrogate.

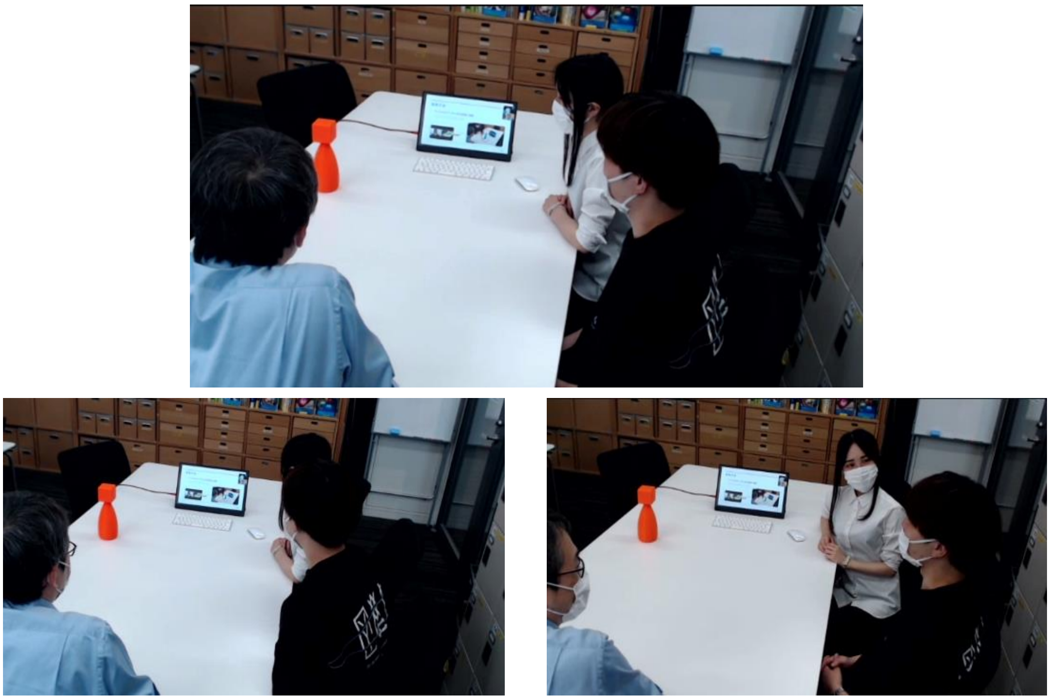


Figure 4: The image transmitted to the remote participant in the meeting experiment using a simple human-like object as the physical avatar. Top: the co-located participants gazing at the avatar. Lower left: the co-located participants referring to the display. Lower right: the co-located participants talking to each other.

4.2 Avatar TYPE2: Telepresence robots

We then tested with an environment in which a telepresence robot was used as an avatar of the remote participant along with a bird's-eye view video (shown in Figure 5). This avatar was different from the simple human-like object in that the device itself had a video phone function with which the remote participant's image and his/her voice were presented on the device, and the remote participant had a first-person view of the main site from the location of the device (shown in Figure6). In addition, the remote user could control its orientation. In this way, the remote participant had two views of the situation at the main site: a fixed wide view and a controllable narrow view.

In this trial, the participants at the main site looked at the avatar of the remote participants as they talked to him. This behavior suggests that the participants at the main site instinctively understood that the avatar was the surrogate of the remote participant. Such behavior was not observed in the previous trial in which the remote participant's status was not presented by the avatar. The phenomenon is therefore considered to have been caused by showing the remote participant's live image on the display of the avatar. The display of the avatar was highly effective for transmitting the status of the remote participant to the main site. Though either the kubi display, the kubi camera, or the display on the table could be the object that the main site participants regarded as the remote participant, the participants mostly faced the kubi in this trial. It could be assumed that the main site participants paid less attention to the kubi camera because they knew that the remote participant was mostly referring to the bird's-eye view video to see the main site, as the remote participant mentioned in the conversation. Therefore, it is suggested that the kubi display, rather than the kubi camera, attracted more attention from the main participants. This indicates that the main site participants recognized the kubi display as representing the remote participants.

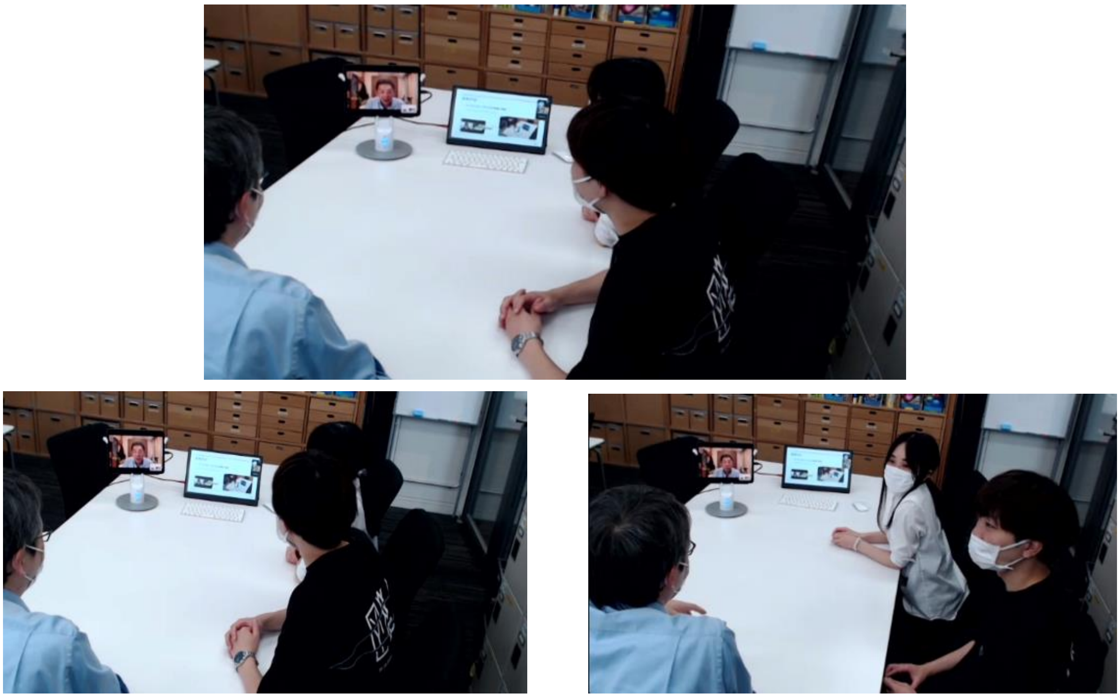


Figure 5: The image transmitted to the remote participant in the meeting experiment using a telepresence robot (kubi) as the physical avatar. Top: the co-located participants gazing at the avatar. Lower left: the co-located participants referring to the display. Lower right: the co-located participants talking to each other.

Figure 6: First-person image captured by the telepresence robot's camera and transmitted to the remote participant.

The display of the telepresence robot simultaneously helped the remote participant to understand that it was his surrogate. The remote participant could confirm visually that his surrogate was physically participating in the meeting at the main site by viewing the avatar through the bird's-eye view. Furthermore, the remote participant referred to the bird's-eye view video more often than the first-person view, remarking that he felt comfortable using the bird's-eye view in a hybrid meeting. The differences between two kinds of views are shown in Figure 5 with Figure 6.

Incidentally, the remote participants could control the orientation of kubi, but did not do so very much. The reason for this may be the large latency of the control. There was a large latency between the remote participant's control operation and the actual change in orientation of the kubi. Therefore, it was difficult to change the orientation of the kubi in response to the change of conversation.

5 DISCUSSION

5.1 Implication

In the hybrid meetings, in order to convey the spatial arrangement of the main site to the remote participants and help them understand the situation at the main site, we set up a bird's-eye view camera. In addition, the physical avatar of the remote participant was placed at the main site so that the remote participant could recognize that he was being spoken to. We prepared two types of physical avatar, a human-like object and a telepresence robot, kubi. We used each physical avatar and a bird's-eye view video in combination to make a hybrid meeting environment.

Through the trials, the remote participant felt that the bird's-eye view video helped them understand the spatial arrangement of the main site. In the Type 1 trial in which the simple human-like-object was used as the physical avatar, it was difficult for the main site participants to treat the physical avatar as a conversation partner, and, as a result, the remote participant did not recognize that he was being spoken to. On the other hand, in the Type 2 trial in which a telepresence robot was used as the physical avatar, the remote participant felt that he was being spoken to.

To properly understand the difference, we need to detail how the telepresence robot was used in the trial. The remote participant could change its viewpoint, but he rarely did so, mainly because of the large latency in the control operation. In addition, the remote participant mostly focused on the bird's-eye view video rather than the first-person view video from the telepresence robot. Though these things are just what we experienced in a short-term trial, these facts imply that the device functioned more to show the remote participant's facial image than as a means to enable the remote participant to look at the main site freely by changing the orientation of the device.

Based on the experiences in the trials, we have come to think that using the combination of the bird's-eye view video and the physical avatar is promising for realizing a hybrid meeting environment where people easily feel they are being spoken to by remote partners. We also found that a design of the avatar that acted as a good surrogate for the remote participant, displaying the facial image of the user, was essential to realize such an environment.

5.2 FUTURE WORKS

In section 5.1, we showed our expectation that the combination of the bird's-eye view video and the physical avatar could create a hybrid meeting environment in which the remote participant easily understands the spatial arrangement of the meeting space and has the sense of being spoken to.

We are planning two experiments to further examine the proposed approach. The first experiment is to observe the usage of the bird's-eye view video. In our trials, the remote participant tended to focus on the bird's-eye view video, and this helped him understand the situation at the main site. To see if this is the case for everyone, we will use a gaze-tracking system to record the gaze movement path of the participants in hybrid remote meetings. While providing both the bird's-eye view video and the first-person perspective video, we will see which video receives more attention from the participants.

The second experiment is to examine the design of the physical avatars. The physical avatar needs to be perceived as a surrogate for the remote participant by both the main site participants and the remote site participant. A physical avatar with a video screen that displays the remote participant's facial image would possibly be a good design. Or, even with simpler devices, there might be designs that work as good surrogates. We will build and compare several avatars, including a simple human-like object, a telepresence robot, and a physical avatar with a video screen, to explore the design principles for building better physical avatars. In this experiment, again, we plan to use a gaze-tracking system to evaluate how often and how long the participants looked at the avatar.

6 CONCLUSION

Hybrid meetings involving co-located and remote participants have faced problems such as asymmetry in terms of the presence of participants at each site, and difficulties in participants understanding what is happening at each site. To overcome this limitation, we used a physical avatar and a bird's-eye view video in combination to create a hybrid meeting environment. By conveying a wide view of the main site to a remote participant through a bird's-eye view camera, we expected to help the remote participants understand the spatial arrangement at the main site. We also expected the physical avatar to help the remote participants feel that they were being spoken to when the remote participants, through the bird's-eye view video, watched the main site participants speak to the physical avatar.

To examine the possibilities of the approach, we have conducted a short-term trial of having meetings by ourselves using the bird's-eye view video and the physical avatar. Through the trials, we have come to think the bird's-eye view video has the potential to help the remote participants understand the spatial arrangement of the main meeting site.

In addition, the use of a telepresence robot made it easy for the main site participants to speak to the physical avatar, which made the remote participant feel as if he was being spoken to.

We are planning two experiments to further explore (1) the use of the bird's-eye view video in hybrid remote meetings, and (2) a design of a physical avatar that acts as a better surrogate for the remote participant. There are many things to do to establish the meeting environment based on the proposed approach, including finding the best bird's-eye viewpoint and the best audio environment. Much work has been done to create remote communication media that conveys the participants' gaze direction, including on eye-contact displays and gaze-aware displays [23]. For the multiparty remote meetings, which are popular in these days, other approach that fit better is necessary. By carefully designing the environment, we believe we can create a hybrid meeting environment in which we can feel other participants' gaze without having “eye contact” displays.

REFERENCES

[1] H. Kuzuoka, J. Kosaka, K. Yamazaki, Y. Suga, A. Yamazaki, P. Luff, and C. Heath, Mediating dual ecologies, Proceedings of the 2004 ACM conference on Computer-Supported Cooperative Work (CSCW 2004), pp. 477–486 (2004).

[2] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, Android as a telecommunication medium with a human-like presence, Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI 2007), pp. 193–200 (2007).

[3] A. Sellen, B. Buxton, and J. Arnott, Using spatial cues to improve videoconferencing, Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI 1992), pp. 651–652 (1992).

[4] Beam is moving to GoBe Robots, https://gobe.blue-ocean-robotics.com/beam-to-gobe, last accessed 2023/06/11.

[5] Double Robotics - Telepresence Robot for the Hybrid Office, https://www.doublerobotics.com/, last accessed 2023/06/11.

[6] kubi, https://kubi-robot.com/, last accessed 2023/06/11.

[7] N. P. Jouppi, First steps towards mutually-immersive mobile telepresence, Proceedings of the 2002 ACM conference on Computer-Supported Cooperative Work (CSCW 2002), pp. 354–363 (2002).

[8] T. Morita, K. Mase, Y. Hirano, and S. Kajita, Reciprocal attentive communication in remote meeting with a humanoid robot, Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI 2007), pp. 228–235 (2007).

[9] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI 2018), Paper 46, pp. 1–13 (2018).

[10] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel, VROOM: Virtual Robot Overlay for Online Meetings, Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA 2020), pp. 1–10 (2020).

[11] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, K. Kaji, J. Yamato, and K. Nakazawa, t-Room: Next Generation Video Communication System, Proceedings of the Global Telecommunications Conference, pp. 1-4 (2008).

[12] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment, Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW 2016), pp. 1716–1725 (2016).

[13] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003), pp. 521–528 (2003).

[14] K. Iso, S. Ozawa, M. Date, H. Takada, Y. Andoh, and N. Matsuura, Video Conference 3D Display That Fuses Images to Replicate Gaze Direction, Journal of Display Technology, vol. 8, no. 9, pp. 511-520 (2012).

[15] W. W. Gaver, A. Sellen, C. Heath, and P. Luff, One is not enough: multiple views in a media space, Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI 1993), pp. 335–341 (1993).

[16] K. Yamaashi, J. R. Cooperstock, T. Narine, and W. Buxton, Beating the limitations of camera-monitor mediated telepresence with extra eyes, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1996), pp. 50–57 (1996).

[17] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, and K. Inkpen, Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010), pp. 1049–1058 (2010).

[18] S. Kasahara, and J. Rekimoto, JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration, Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST 2015), pp. 217–225 (2015).

[19] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, Mixed reality. collaboration through sharing a live panorama, SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications (SA 2017), Article 14, pp. 1–4 (2017).

[20] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019), Paper 201, pp. 1–14 (2019).

[21] Video Conferencing, Cloud Phone, Webinars, Chat, Virtual. Events | Zoom, https://zoom.us/, last accessed 2023/06/11.
[22] Video Conferencing, Meetings, Calling | Microsoft Teams, https://www.microsoft.com/en/microsoft-teams/group-chat-software, last accessed 2023/06/11.
[23] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin, Integration of interpersonal space and shared workspace: ClearBoard design and experiments, ACM Transaction on Information Systems, vol.11, no.4, pp.349–375(1993).

Exploring Key Concepts in Arts and Mental Health: A Questionnaire-based Analysis of Perception on Arts, Sports, and Therapy

Michi KOMURA*, Hajime KANEKO**, Midori ISHIHARA***, Akihiro HAYASHI****

*Osaka Metropolitan University, Osaka, Japan
**Kobe University of Future Health Sciences, Hyogo, Japan
***ATAS Laboratory, Hyogo, Japan
****Shizuoka Institute of Science and Technology, Shizuoka, Japan
{*mitikomm347, **kaneko68h, ***midorippa2021, ****pixysbrain}@gmail.com

Abstract - The use of arts for healing and empowerment has spread throughout Japan in recent years, a trend broadly related to mental health. Our study defines this social phenomenon as empowerment arts therapy (EAT) or arts empowerment (AE). In this context, the terms “arts,” “therapy,” or a combination of these words have varying intentions and implications due to the range of perceptions these words evoke. A questionnaire survey was conducted on university students to investigate their perceptions of these terms. The questionnaire results indicated that perceptions of the arts and therapy varied widely; in some cases, there was a significant discrepancy between students’ perceptions and dictionary definitions. Specific insights were gathered to establish a shared comprehension of the terms and concepts related to arts and therapy, which are often used ambiguously. EAT and AE practitioners must consider the diversity of perceptions, effectively communicate their activities, and gain people’s understanding of EAT and AE. The survey findings provide useful insights into understanding and addressing the elusive nature of perceptions surrounding arts and therapy.

Keywords: Empowerment Arts Therapy, Arts Empowerment, Mental Health.

1 INTRODUCTION

Interest in arts therapy has recently increased as social awareness of mental health has grown. The therapeutic effects of art have attracted attention as a tangible approach to improving mental health. Beyond its traditional application as a form of psychotherapy, various arts therapy practices are expanding outside clinical settings. Examples include its implementation in elementary school classes [1], welfare facilities [2], adult art classes [3], and a local community atelier for children [4]. These activities apply arts therapy methods and are carried out in everyday life; however, they differ from arts therapy as a form of psychotherapy aimed at medical treatment within clinical settings. In addition, there is interest in the intrinsic therapeutic properties inherent in art and their correlation with well-being [5]. Various activities using art are being developed beyond high and fine art, focusing on fostering healing and empowerment.

Our research project named Arts Therapy Activities Study (ATAS) explores a range of activities that employ arts therapy outside of clinical settings. In the research process, we proposed the concepts of EAT and AE.

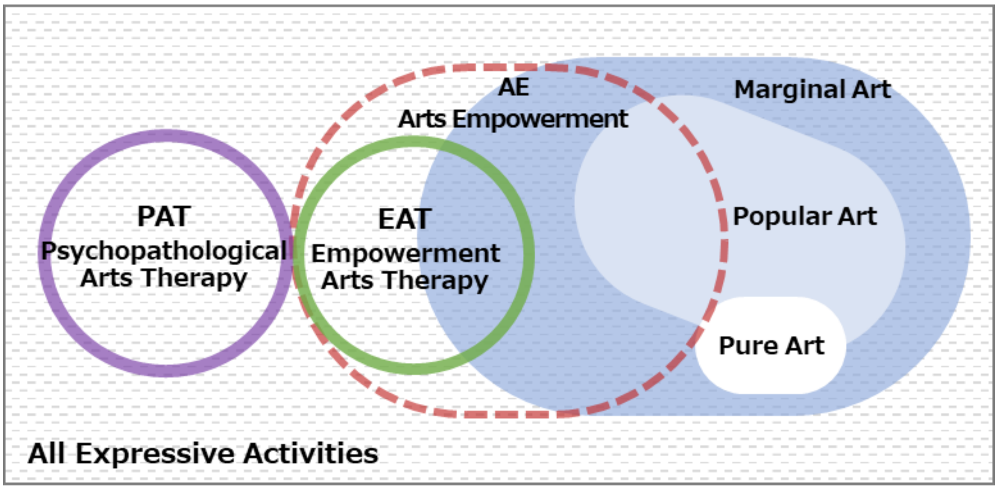
Arts therapy is originally a general term for psychotherapy using various artistic techniques, such as painting, pottery, music, theater, and dance. As a form of psychotherapy, it is conducted by psychiatrists or specialized therapists in clinical settings, such as psychiatric care and mental counseling, as well as in parallel disciplines. In our research of ATAS, we differentiate the form of arts therapy conducted in clinical settings as Psychopathological Arts Therapy (PAT) to distinguish it from EAT.

In contrast to PAT, EAT is defined as an activity that aims to rejuvenate the mind and body, alleviate stress, and improve quality of life (QOL) in everyday settings. EAT is based on concepts and methods of arts therapy, but unlike PAT, it is not intended for medical treatment. Furthermore, AE is defined as an art activity in which the therapeutic nature is inherent in expressive activities, even if they are not intentional.

Both EAT and AE are located at the intersection of art and therapy, with distinctions based on the emphasis on the artistic or therapeutic aspect. EAT is characterized by a stronger therapeutic effect, while AE leans more toward artistic expression. However, some activities fall in between the disciplines, making it challenging to differentiate them. Figure 1 shows the positional relationship between art, therapy, and empowerment in all expressive activities.

Thus, artistic activities, including therapeutic ones, support individuals’ mental well-being in their everyday lives. Despite the increasing prevalence of such activities in Japan, there have been few studies beyond ATAS that comprehensively examine various EAT practices, clarify their unique requirements and significance, and address associated challenges.

Many of these activities share the common problem of resource scarcity. Although EAT is spreading at the grassroots level, the environment surrounding these activities is insufficient.



*Marginal Art, Popular Art, Pure Art: Based on Shunsuke Tsurumi's classification and definition. Marginal art is created by nonexperts and enjoyed by non-specialists. Pure Art is created by experts and enjoyed by connoisseurs familiar with the field. Popular Art is created by experts and companies and enjoyed (consumed) by nonexperts (masses) [6].

Figure 1: Positional Relationship between Art, Therapy, and Empowerment in All Expressive Activities in Japan [7]

Existing issues with EAT can be summarized as follows:

1) Sustainability: Most EAT practitioners work voluntarily. They are working to find and implement training opportunities, but their activities are not economically viable. This is a crucial problem but is difficult to solve in the short term in the present Japanese social system. For now, it is realistic to increase sustainability within the existing constraints of assuming a volunteer basis [7].
2) Quality Assurance: Many EAT practitioners work independently. There are no systematic networks or opportunities for information exchange or training. This means that no standards or indicators have been established to ensure the quality of EAT activities or practitioner capabilities. Self-assessment tools are in development to improve practitioner capabilities [8-11].
3) Awareness / Understanding: Interest in mental health and awareness of arts therapy have recently increased in Japan. However, there is still limited public understanding of arts therapy and the distinction between arts therapy as psychotherapy and EAT activities in daily life. This lack of understanding makes it difficult for EAT practitioners to explain their activities to society.

The first issue needs long-term and large-scale efforts to be solved. It needs to be addressed with a view to revising and reforming the social system.

Relating to the second issue, we have proposed self-assessment tools for EAT practitioners and are working on the investigation of effectiveness [8-11].

This study addresses the third issue: the lack of awareness and understanding of EAT activities in daily life. It is the preliminary study to approach the issue related to people who can be potential clients or users of EAT. We have conducted research and practice mainly related to EAT practitioners so far but have not done the study related to potential clients of EAT yet.

The ATAS has conducted comprehensive research on the current state of EAT and AE using questionnaire surveys, interviews with practitioners, activity observations, and the development of assessment methods to enhance the quality of activities and practitioner capabilities. The survey revealed the diverse perceptions of arts and therapy among individuals, as well as variations in perceptions among different EAT practitioners.

Given these findings, we conducted a questionnaire survey to explore perceptions of arts, sports, and therapy. The survey results were then analyzed to identify key concepts related to EAT and AE.

Chapter 2 summarizes the issues surrounding the basic concepts of EAT and AE, outlining the issues to be solved and the research objectives. Chapter 3 presents an overview of the survey results. Chapter 4 concludes the study and highlights future issues.

2 ESSENTIAL ART AND MENTAL HEALTH CONCEPTS

2.1 Ambiguity of Terms

EAT encompasses elements of both arts and therapy. EAT practitioners often draw upon ideas and methods from PAT, incorporating them into their activities while adding their own innovative approaches. Unlike PAT for medical treatment, EAT activities are practiced to empower people by revitalizing the mind and body, alleviating stress, and improving QOL. However, many EAT practitioners lack precise terms or concepts to convey the nature of their activities accurately.

In Japanese, "art(s)" is a concept introduced and translated from a Western word. The Japanese language has three types of characters: kanji, hiragana, and katakana. Foreign words are translated into Japanese or written in katakana—the Japanese syllabary used for loanwords—based on their pronunciation in the original language. The word "art(s)" is translated as the Japanese term "*geijutsu*" or written as the loanword "*āto*" in katakana.

It is common for words and concepts translated from Western languages to exhibit some differences or gaps in connotation between the original terms and their Japanese translations, whether written in kanji or katakana. There is a distinction between the English "art(s)" and Japanese "*geijutsu*" or "*āto*".

While the English word "art" encompasses techniques and skills as well as fine art, the Japanese term "*geijutsu*" in kanji is limited to aesthetic meanings and may sound academic or unfamiliar to the general public. In contrast, "*āto*" in katakana has a broader connotation and gives a more casual impression.

The situation is similar with the term "therapy". Its Japanese equivalent words, such as "*chiryō*" and "*ryōhō*" in kanji, have a more professional connotation. Similarly, "*serapī*" in katakana tends to be used in a broader or more casual context.

The polysemy of these Japanese terms suggests that people's perceptions of these concepts are diverse and even ambiguous. Loanwords written in katakana possess a broader connotation that includes both the original meanings in English and the Japanese translations in kanji.

Consequently, in Japanese, katakana loanwords have meanings that are not entirely equivalent to the original foreign words or their kanji translations. As a result, foreign words in katakana are not always as clear in meaning as the original terms or kanji translations; their usage and recognition may vary depending on the individual. Therefore, there are gaps in perceptions between EAT practitioners and potential clients.

2.2 Issues to Be Solved

Through our survey so far, the following characteristics and situations with EAT activities have been grasped. These were mainly revealed through the interviews to EAT practitioners who were informants of the nationwide survey [12].

EAT practitioners utilize various methods of arts therapy for their activities. However, they tend to hesitate about or avoid using the term "arts therapy" because their activities do not aim at psychotherapy. On the other hand, EAT practitioners emphasize supporting and accepting individuals' unique forms of expression rather than focusing on technical instructions or improvement in artistic activities, such as painting, music, and dance. They view their activities as distinct from general art activities. Therefore, EAT practitioners recognize the challenge of finding precise words to describe their activities accurately. The unique nature of their approach which is different from both arts therapy as psychotherapy and usual art activities makes it difficult to convey the full scope of their work.

Given these situations and challenges, we conducted a preliminary survey as a first step to grasp how people perceive the terms art(s) and therapy. EAT is a newly proposed academic concept unfamiliar to people, so we investigated concrete items related to EAT instead of EAT itself. Understanding the gaps and commonalities in consciousness between EAT and AE practitioners and their potential clients makes it possible to facilitate the comprehension of EAT and AE.

3 ANALYSIS OF THE CONCEPTS OF ART AND THERAPY

3.1 Outline of Questionnaire Survey

We administered the following questionnaire to investigate perceptions of arts and therapy:

- Target group: Students from two universities in Osaka and Shizuoka.
- Number of respondents: 277
- Valid responses: 277
- Period of the survey: April 11 to 13, 2023
- Format: Selective questionnaire
- How to answer: Respondents answered the questionnaire created with Google Forms on their smartphones or personal computers.

The questionnaire was administered during class hours at the authors' universities. The male-to-female ratio of students was approximately 8:2.

3.2 Contents of the Questionnaire

The questionnaire consists of the following main sections:

[Section 1]

(1) Age of respondents

(2) Place of residence during high school

[Section 2]

(1) Classification of 60 items[1] of arts and sports genres into the following categories:

- Entertainment/Hobby
- Competition
- Sport
- Education/Physical Education
- Art (*geijutsu* in Japanese)
- Art (*āto* in Japanese)
- Performing Art

(2) Choose one word most strongly associated with the term "therapy" ("*serapī*" written in katakana) as a loanword from the given options of events related to the arts or sports.

[Section 3]

(1) Respondents' involvement in art and sports activities (with more than three years of experience).

(2) Whether the activities they focused on were athletic or non-athletic.

[Section 4]

Classify events related to the arts or sports into one of the same categories as in Section 2.

*Different questions were given depending on whether the respondent chose athletic or non-athletic in Section 3.

[1] Two of the 60 items were treated as invalid due to incomplete questionnaires. Thus, the final valid number of items was 58.

Section 2(1) presented a wide range of questionnaire items relating to arts and sports, from popular and familiar to classical and unfamiliar. The intention was to observe the uniformity or dispersion of category choices across different items.

The seven category options given in Japanese, including synonymous terms like *geijutsu* and *āto*, were included to analyze the distinction in usage between these two kanji and katakana words based on the respondents' answers. In Japanese, there is an unconscious or implicit usage distinction between terms in kanji and katakana, even when both meanings are identical.

The questionnaire included sports because some activities are considered at the intersection of art and sports. Moreover, there are several concepts such as "competition (*kyōgi*)," "sport (*supōtsu*)," and "physical education (*taiiku*)" which are considered similar concepts in Japanese. Differences in the perception and usage of these terms should also be analyzed.

The purpose of the questionnaire was not to rigorously define or organize the concepts and categories of art, sports, or therapy. It was to grasp how people perceive these terms and the activities they envision and to confirm the state in which the relevance between actual events and the categories and terms representing them was fluctuating. It was significant to estimate how EAT and AE were likely to be accepted by people by viewing what deviations occurred between people with various perceptions. As a preliminary study, the questionnaire survey[2] was conducted on university students. They are considered to have more freedom in club activities and less sense of belonging to their clubs than high school students even if they join any clubs.

Respondents were expected to rely on the words and descriptions commonly used in society to choose the option reflecting their perceptions, regardless of whether they were familiar with the specific content of each item. It should be noted that the questionnaire did not aim to rigidly define or organize the concepts and categories of art, sports, or therapy. Therefore, no categories were defined or explained in detail. The purpose of the questionnaire was to highlight the diversity and fluctuations in the perceptions of these words and concepts, with a focus on art-related questions.

3.3 Results of the Survey

3.3.1 Perception of Art

This chapter examines the results of the responses to Section 2(1) of the questionnaire regarding item categories. This study addresses the items related to visual art, music, theater, and dance (refer to the appendix) out of all questionnaires including the athletic and miscellaneous items. These items are major components of EAT and AE. The following describes the tendency grasped.

■ Visual Art

Regarding paintings (Figure 2) and sculptures/objects (Figure 3), most of the respondents chose "art as *geijutsu*" or "art as *āto*." The selection of "*geijutsu*" surpassed "*āto*" for both items. For paintings, 144 respondents chose "art as *geijutsu*," while 78 did "art as *āto*." For sculptures/objects, 163 chose "art as *geijutsu*" and 61 did "art as *āto*."

On the other hand, "entertainment/hobby" was the most common category for illustrations (133 respondents) (Figure 4), followed by "art as *āto*" (99). "Art as *geijutsu*" (31) accounted for less than one-third of "art as *āto*" for illustrations.

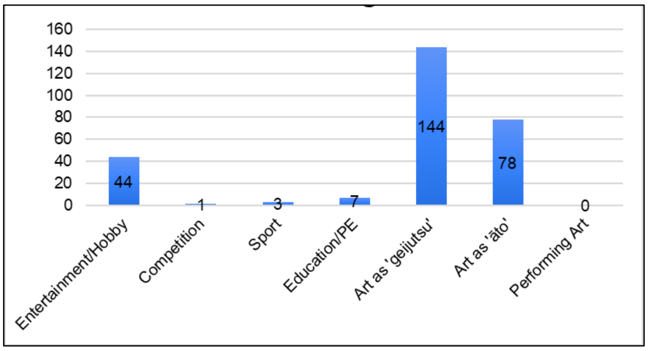


Figure 2: Painting

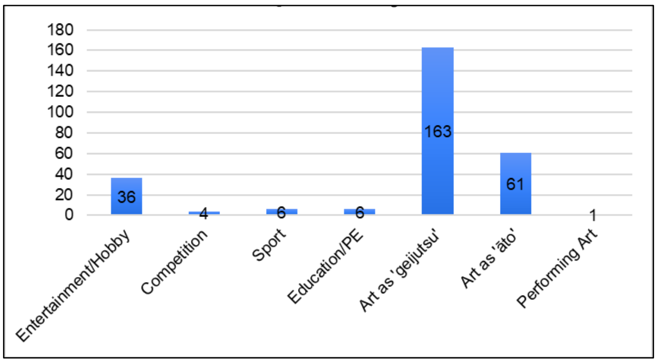


Figure 3: Sculpture/Object

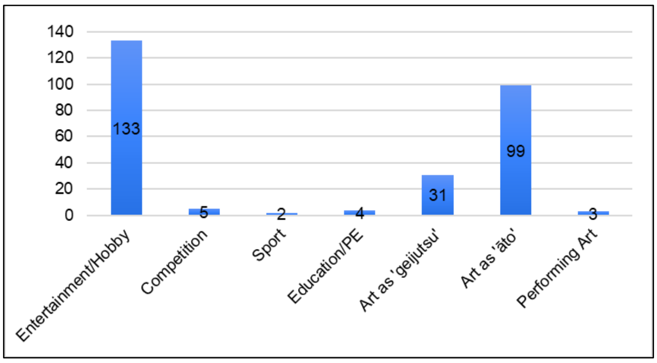


Figure 4: Illustration

Paintings and illustrations are two-dimensional expressions of art, but there is a difference in the category reflected in respondents' perceptions. Respondents selected "art as *geijutsu*" more than "art as *āto*" for paintings and sculptures/objects; however, the converse was true for "art as *āto*" for illustrations. This suggests that "art as *āto*" is considered more familiar and commercial than "art as

[2] The questionnaire was prepared based on the perspective of phenomenological sociology proposed by A. Schutz [13].

geijutsu,” while “art as *geijutsu*” evokes a purer form of fine art, emphasizing “art for art’s sake.” This observation highlights a difference in perception and usage between “art as *geijutsu*” and “art as *āto*” as Japanese terms.

“Entertainment/hobby” was the most common category for photography (Figure 5), animations (Figure 6), movies, and Manga (cartoons). These items can be regarded as visual art in a broad sense. However, they are considered more familiar than paintings and sculptures/objects. It is suggested by the fact that the majority (60-80%) of respondents chose “entertainment/hobby” for each item.

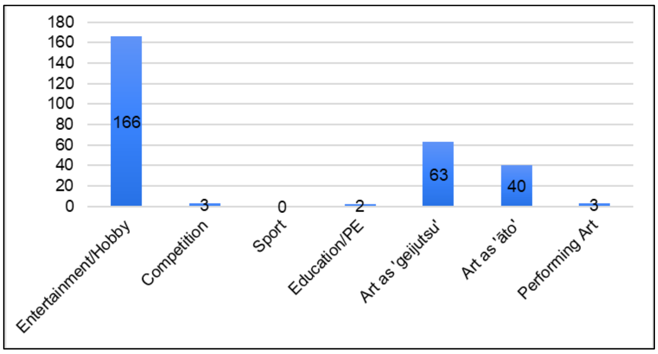


Figure 5: Photography

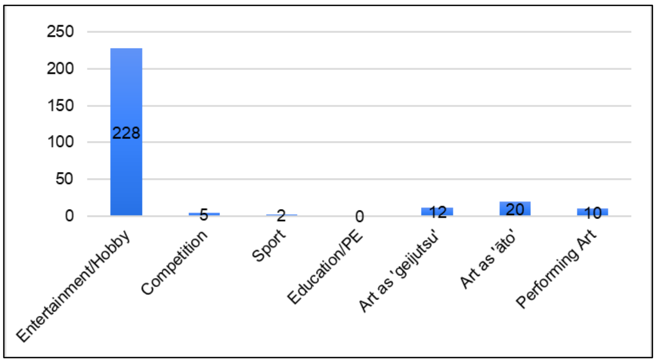


Figure 6: Animation

■ Music

“Art as *geijutsu*” was the most common choice for classical music (126) (Figure 7) and opera (134). On the other hand, “entertainment/hobby” was the most common category for pop music (152) (Figure 8), and other forms of popular music such as music videos (125), jazz (126), and rock (138).

Unlike the visual art field, the “entertainment/hobby” ratio is significantly high in music; the ratio of “art as *āto*” is low. Even in classical music and opera, where “art as *geijutsu*” was the most common, the second most common category was “entertainment/hobby.” It may be inferred that musical and stage performances are considered both high art and entertainment.

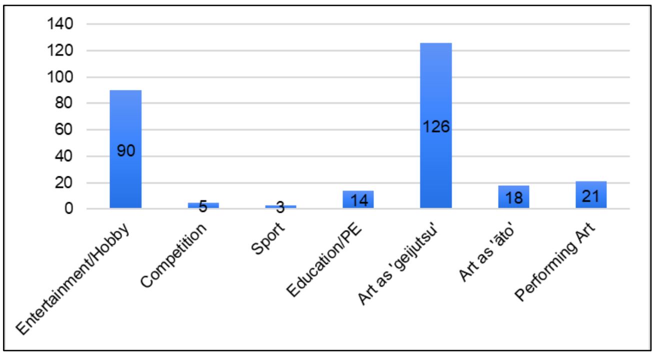


Figure 7: Classical Music

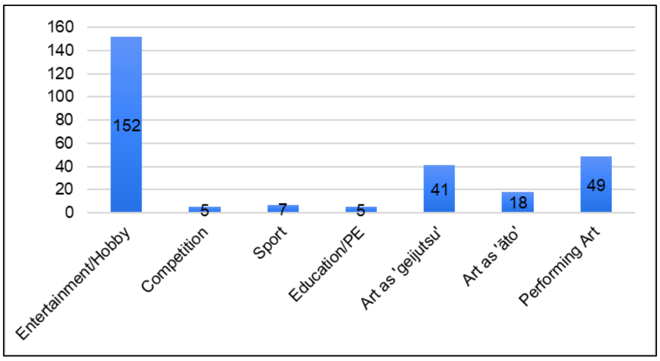


Figure 8: Pop Music

■ Theater

In the domain of theater, “performing arts” was the most common category for all items: dramas (124) (Figure 9), Kabuki (150) (Figure 10), Noh play (140), Bunraku (Japanese puppet theater) (136), and Manzai (comic dialogue) (140) (Figure 11).

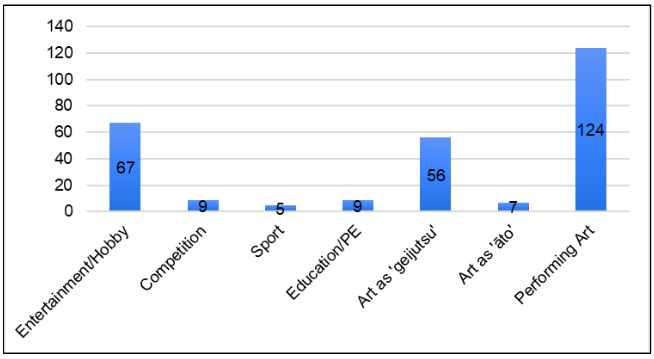


Figure 9: Drama

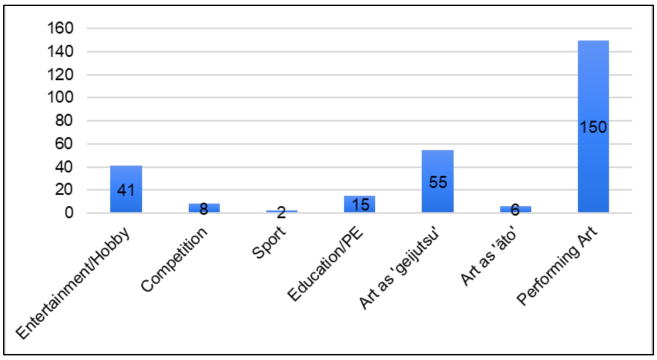


Figure 10: Kabuki

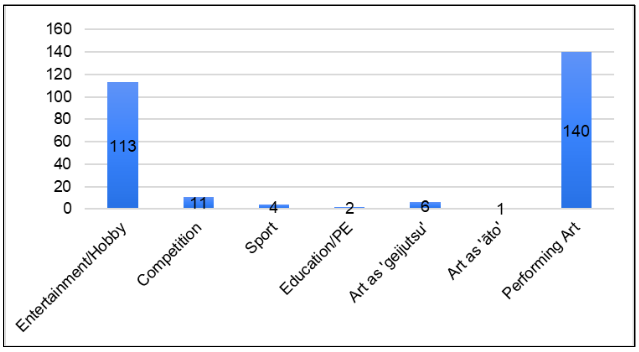


Figure 11: Manzai

The Japanese word "*geinō*" means performing arts and refers to both popular show business and classical performing arts. Kabuki, Noh, and Bunraku are the foremost Japanese classical performing arts registered as World Intangible Cultural Heritage. It is unlikely that they are completely equated with general theatrical performances. Notably, the major category chosen for both popular and classical performances is the word "*geinō*" in Japanese, indicating the wide-ranging scope of this term in Japanese.

■ Dance

In the domain of dance, there was notable recognition of it as a physical activity across various items. Dance was perceived as closely related to sports or existing at the intersection of art and sports. Moreover, many dance items were evenly spread across categories compared to other areas.

"Entertainment/hobby" was the most common category for almost all individual dance items. However, the "art as *geijutsu*" category was not commonly chosen. Even in ballet (Figure 12), a representative dance art, "sport" was the most common choice (74), followed by "art as *geijutsu*" (54). In ice dancing (Figure 13) "sport" and "entertainment/hobby" ranked highest (both 66).

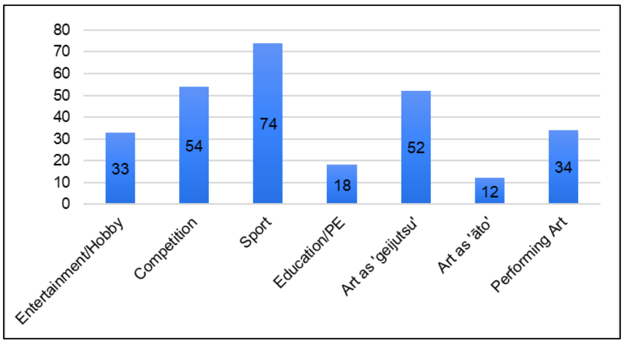


Figure 12: Ballet

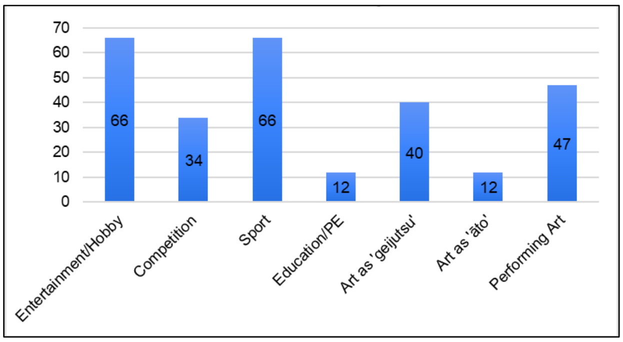


Figure 13: Ice Dancing

Ballroom (Figure 14) and creative dancing were commonly chosen as "entertainment/hobby" (67 and 79, respectively). The second most common category for ballroom dancing was "competition" (59), and for creative dance was "sport" (65).

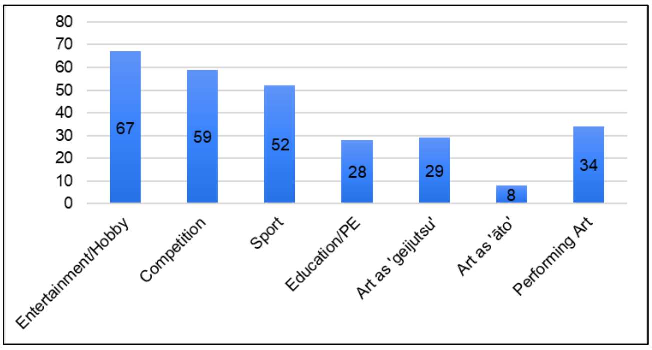


Figure 14: Ballroom Dancing

In hip-hop (Figure 15) and break dancing, which are both familiar to the younger generation, "entertainment/hobby" was the most common category (126 and 97, respectively), while "sport" ranked second (50 and 63, respectively).

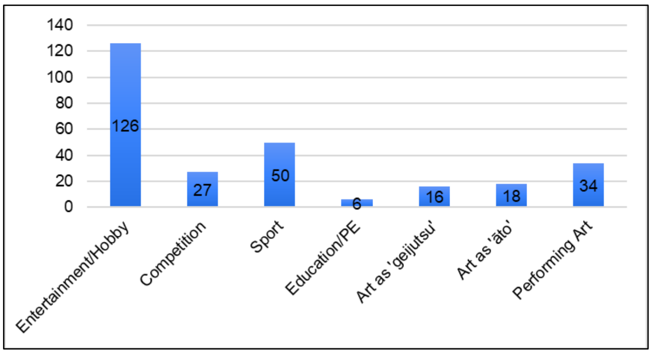


Figure 15: Hip-Hop

Traditional Japanese dance (Figure 16) had the largest number of "art as *geijutsu*" (73) in all kinds of dance. However, its most common category was "performing art (*geinō*)" (99). It is presumed that this is treated as a classical performing art, similar to Kabuki, Noh, and Bunraku.

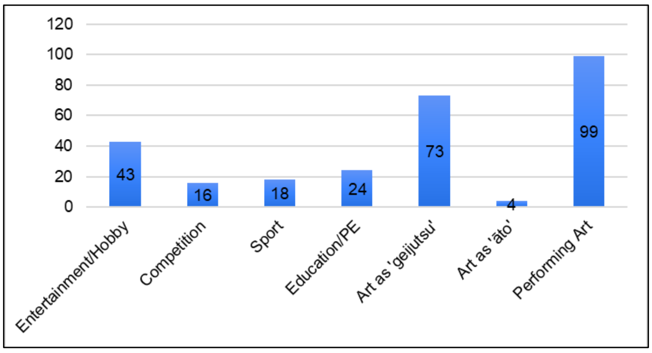


Figure 16: Japanese Dance

The university students surveyed were not majoring in arts or sports and did not have specialized knowledge or experience in these areas. Hence, the choices noted in the results can predict the cognitive tendencies of the general students' perception of the arts and sports.

3.3.2 Perception of Therapy

The respondents were asked to select one of the options that most strongly reminded them of the word "therapy." The results are shown in Figure 17.

The English word therapy refers to treatment or cure, which refers to medical practice or related fields. However, the word most strongly associated with "therapy" as a Japanese word written in katakana (*serapī*) was "relaxation" (87), followed by "healing" (66). While this perception might differ among medical professionals and those studying medicine, it suggests that Japanese people generally perceive the loanword *serapī* in a broader sense of relaxation and healing rather than strictly within the context of medical practice in the Japanese language.

***[Question]** Therapies utilizing arts and sports have recently been increasing. Please choose one word that you most strongly associate with the term "therapy".*

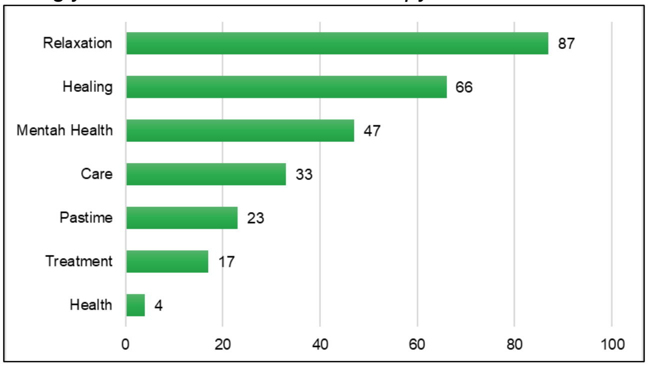


Figure 17: Word Most Strongly Associated with "Therapy"

Some EAT and AE practitioners perceive "therapy" as healing and hope others will perceive it similarly. EAT involves art activities incorporating therapeutic elements to refresh the mind and body and relieve stress. However, since "therapy" is often associated with medical treatment or cure, many EAT practitioners are cautious about using that term, as they are not engaged in psychotherapy.

The associations that Japanese people have with loanwords written in katakana may not always align with the word's original meaning or its Japanese translation. Arts therapy as a loanword tends to evoke a more casual and accessible impression than its Japanese translation *geijutsu ryōhō*. This allows for a broader connotation and recognition among Japanese people.

However, it is necessary to carefully consider whether it is appropriate for EAT practitioners to represent their activities as "arts therapy" because it may be confused with arts therapy as psychotherapy. This becomes a significant challenge for EAT practitioners as they try to explain their activities concisely and accurately. Most EAT practitioners are aware of these difficulties in describing their activities.

There is often a gap between loanwords from foreign languages and their Japanese translations. Linguistic discussions were beyond the scope of this study; therefore, we did not discuss them in detail. While this study did not extensively delve into linguistic discussions, it is commonly understood that there are discrepancies between the literal definitions and people's perceptions of "arts therapy." This is an important aspect to carefully examine when discussing and implementing EAT and AE approaches.

4 DISCUSSION

4.1 Outcomes

Using this questionnaire, we confirmed the diversity and ambiguity of arts and therapy recognition. Overall, "art as *geijutsu*" tends to be associated with things of a historical and traditional nature (e.g., classical music and traditional Japanese dance). "Art as *āto*" is linked to more modern or commercialized items (e.g., illustrations).

"Entertainment/hobby" is commonly chosen for activities related to self-expression and daily enjoyment (e.g., animation, pops). In the field of dance, there is a notable emphasis on the physicality of the expression and the broad range of categories encompassing it.

Regarding "therapy," respondents primarily associated it with healing and relaxation rather than medical treatment, regardless of strict definitions of the term.

EAT and AE practitioners should recognize the diversity and tendencies of perceptions of arts and therapy in order to accurately explain their activities without causing confusion or misunderstanding. To address this issue and increase public awareness and social recognition of their activities, EAT and AE practitioners should recognize everyone has their own unique perception of arts and therapy.

For many practitioners conducting EAT or AE activities independently, ICT tools such as websites and social media are useful for public relations. However, recipients of information do not necessarily have the same perceptions of the arts and therapy as practitioners. Therefore, it is important to establish the basic policies and spread the information considering people's wide-ranged perception of arts and therapy. These actions will assist EAT and AE practitioners in achieving a better public understanding of their activities. The survey results will provide clues of insights for a deeper understanding of diversity of these terms and concepts.

4.2 Future Issues

This survey has several limitations. The respondents were limited to predominantly male university students (80%). Detailed questions were not asked about the respondents' attributes, such as their social, economic, and cultural backgrounds, which could have influenced their responses.

To promote the practice and dissemination of AE and EAT, it is important to understand the circumstances of a more diverse population, including individuals of different ages and genders. This understanding can inform the development of targeted public relations and marketing strategies. It would also be beneficial to analyze the correlation between respondents and their economic and cultural capital.

Despite these limitations, this preliminary survey shed light on the diverse and ambiguous perceptions of arts and therapy, providing useful insights for future research. Future studies will address the unresolved issues that require further examination.

REFERENCES

[1] T. Okada, Psycho-educational Approach with Application of Psychotherapy in Elementary School - Trial and Results of the Technique: ‘Picture Drawing Play’, Japanese Bulletin of Arts Therapy, Vol.40, No.1, pp.43-51 (2009) (in Japanese).

[2] T. Fujii, A Postcard Artist’s Report on Cloth Collage Workshops, Japanese Bulletin of Arts Therapy, Vol.41, No.1, pp.70-80 (2010) (in Japanese).

[3] N. Okada, Mosaic, Reflections in Art Therapy, Japanese Bulletin of Arts Therapy, Vol.50, No.1, pp.90-97 (2019), (in Japanese).

[4] M. Ishihara, “Nichijo ni yorisou art therapy - Kosodate-shien toshite no katsudo-jirei ni art therapy no hattenkei wo miru -”, K. Tokiko, and K. Nishi (eds.), Art Therapy Saiko - Geijutsugaku to rinsho no genba kara -, Heibonsha, Tokyo, pp.232-253 (2013) (in Japanese).

[5] Headquarters for Vitalizing Regional Cultures, Agency for Cultural Affair, Bunka ni kansuru yoron chosa -Well-being to bunka geijutsu katsudo no kanren hokokusho - (2022) (in Japanese).

https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/pdf/93858401_01.pdf

[6] Shunsuke Tsurumi, Genkai-geijutsu-ron, Chikuma Shobo, pp.14-15 (1999) (in Japanese).

[7] H. Kaneko, and M. Ishihara, A View of ‘Art Therapy for Empowerment’, Journal of the Kobe University of Welfare, Vol.19, No.1, pp.1-22 (2018) (in Japanese).

[8] M. Komura, and A. Hayashi, A Method of Assessment for Empowerment Arts Therapy, IEICE Technical Report, Vol.121, No.372, pp. 1-7 (2022) (in Japanese).

[9] M. Komura, M. Ishihara, and A. Hayashi, Proposal for Continuous Improvement of Empowerment Arts Therapy - A Method of Self-Assessment by Applying Rubric -, Proc. International Workshop of Informatics (IWIN 2022), pp.121-126, Nachi-Katsuura, Japan (2022).

[10] M. Ishihara, M. Komura, A. Hayashi, and H. Kaneko, Empowerment gata arts therapy no self-assessment no jitsuyoka ni mukete - Empower suru hito wo empower suru tame ni -, The Crisis of the Mind and Clinical Knowledge (Bulletin of Konan Institute of Human Sciences, Konan University), Vol. 24, pp.1-21 (2023) (in Japanese).

[11] M. Komura, M. Ishihara, H. Kaneko, and A. Hayashi, Effectiveness Applying a Rubric in Self-Assessment of Empowerment Arts Therapy Practitioners, IEICE Technical Report, Vol. 122, No. 282, pp. 33-39 (2022) (in Japanese).

[12] M. Ishihara, and H. Kaneko, Empowerment toshite no shisei no art therapy, The Crisis of the Mind and Clinical Knowledge (Bulletin of Konan Institute of Human Sciences, Konan University), Vol. 16, pp.105-130 (2015) (in Japanese).

[13] A. Schutz, edited by M. Natanson, translated by H. Watabe, H. Nasu, and, K. Nishihara, Alfred Schutz chosaku-shu 2: Shakai-teki genjitsu no mondai II, Maruju-sha (1985) (Japanese translation of A. Schutz, edited & introduced by M. Natanson, Collected Papers I: The problem of the Social Reality - Part III, The Hague (1973)).

Appendix: Questionnaire Results of Items Relating to Arts

	Items	Category Selection							Total	Graph
		Entertain-ment/Hobby	Competi-tion	Sport	Education /PE	Art as *'geijutsu'*	Art as *'āto'*	Perform -ing Art		
Visual Art	Painting	44	1	3	7	144	78	0	277	Fig.2
	Illustration	133	5	2	4	31	99	3	277	Fig.4
	Sculpture/Object	36	4	6	6	163	61	1	277	Fig.3
	Movie	190	0	1	4	32	13	37	277	
	Photography	166	3	0	2	63	40	3	277	
	Animation	228	5	2	0	12	20	10	277	
	Manga	226	2	1	3	17	22	6	277	Fig.5
	Calligraphy	39	17	4	66	98	34	19	277	Fig.6
Music	Karaoke	256	4	3	1	3	2	8	277	
	Classical Music	90	5	3	14	126	18	21	277	Fig.7
	Pop Music	152	5	7	5	41	18	49	277	Fig.8
	Japanese Popular Song (Kayokyoku)	122	3	3	12	44	11	82	277	
	Music Video	125	3	2	5	46	31	65	277	
	Opera	58	8	6	14	134	12	45	277	
	Falk Song	62	8	3	47	83	15	59	277	
	Japanese Enka Song	96	5	4	9	56	11	96	277	
	Jazz	126	7	5	1	61	26	51	277	
	Rock	138	7	4	4	46	26	52	277	
Drama	Drama	67	9	5	9	56	7	124	277	Fig.9
	Kabuki	41	8	2	15	55	6	150	277	Fig.10
	Noh play	43	4	3	20	54	5	148	277	
	Bunraku	45	6	2	15	64	9	136	277	
	Manzai	113	11	4	2	6	1	140	277	Fig.11
Dance	Ballet	33	54	74	18	52	12	34	277	Fig.12
	Creative Dance	79	37	65	19	27	21	29	277	
	Ballroom Dancing	67	59	52	28	29	8	34	277	Fig.14
	Japanese Dance	43	16	18	24	73	4	99	277	Fig.16
	Folk Dance	69	38	49	22	38	12	49	277	
	Ice Dancing	66	34	66	12	40	12	47	277	Fig.15
	Japanese Bon Dance	96	5	11	29	48	2	86	277	
	Hip-Hop	126	27	50	6	16	18	34	277	
	Breakdance	97	41	63	10	16	14	36	277	
	Modern Dance	102	29	59	7	22	19	39	277	
	Jazz Dance	99	34	51	11	27	19	36	277	
	Cheer Dance	76	46	85	18	12	9	31	277	
	Hula	106	32	48	14	22	14	41	277	