IWIN2014

International Workshop on Informatics

Proceedings of

International Workshop on Informatics

September 10-12, 2014

Prague, Czech Republic

Informatics Society

Sponsored by Informatics Society

URL: http://www.infsoc.org/

Publication office:
Informatics Laboratory
3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan
Publisher:
Tadanori Mizuno, President of Informatics Society
ISBN:
978-4-902523-35-5
Printed in Japan

Table of Contents

Session 1: Intelligent Transportation Systems

(Chair: Tomoya Kitani) (10:00 - 12:00, Sept. 10)

Session 2: Databases and Systems

(Chair: Takuya Yoshihiro) (13:00 - 15:30, Sept. 10)

Session 3: Social System

(Chair: Kazo Okano) (15:45-18:15, Sept. 10)

Session 4: Intelligent Systems and Applications

(Chair: Yoshitaka Nakamura) (09:30-12:00, Sept. 11)

Keynote Speech 1

(13:00-14:00, Sept. 11)

Keynote Speech 2

(14:00-15:00, Sept. 11)

Session 5: Networks, Applications and Web

(Chair: Yoh Shiraishi)(15:15-17:15, Sept. 11)

Keynote Speech 3

(17:30-18:30, Sept. 11)

Panel Session: Toward Future ICT Systems

- Requirements and Solutions –

(09:30-12:00, Sept. 12)

Chair

- Prof. Norio Shiratori, Waseda University, Japan

Panelists

- Masashi Saito, Mitsubishi Electric Corporation, Japan
- Prof. Yoh Shiraishi, Future University Hakodate, Japan
- Prof. Tomoya Kitani, Shizuoka University, Japan

A Message from the General Co-Chairs

It is our great pleasure to welcome all of you to Prague, Czech Republic, for the Eighth International Workshop on Informatics (IWIN 2014). This workshop has been held annually and sponsored by the Informatics Society. The first, second, third, fourth, fifth, sixth, and seventh workshops were held in Napoli, Italy, Wien, Austria, Hawaii, USA, Edinburgh, Scotland, Venice, Italy, Chamonix France, and Stockholm, Sweden, respectively. The first workshop was held in 2007. All of workshops were held in September.

In IWIN 2014, 24 papers have been accepted and 13 papers have been further selected as excellent papers that are considered having significant contributions in terms of the quality, significance, current interest among the professionals, and conference scope through the peer reviews by the program committees. Based on the papers, five technical sessions have been organized in a single track format, which highlight the latest results in research areas such as mobile computing, networking, information system, and groupware and education systems. In addition, IWIN 2014 has three invited sessions from *Mr. Atsushi Takeshita* of *DOCOMO Communications Laboratories Europe, from Dr. Ichiro Iida of Fujitsu Laboratories Ltd*, and from *Dr. Masashi Saitoh, Information Technology R&D Center, Mitsubishi Electric Corp*. We really appreciate the participation of the three invited speakers in this workshop.

We would like to thank all of participants and contributors who made the workshop possible. It is indeed an honor to work with a large group of professionals around the world for making the workshop a great success.

We are looking forward to seeing you all in the workshop. We hope you all will experience a great and enjoyable meeting in Prague.

Yoshimi
Teshigawara, and
Kozo Okano

General Co-Chairs
The International Workshop on Informatics 2014

Organizing Committee

General Co-Chairs

Yoshimi Teshigawara (Tokyo Denki University, Japan)

Kozo Okano (Osaka University, Japan)

Steering Committee

Toru Hasegawa (Osaka University, Japan)

Teruo Higashino (Osaka University, Japan)

Tadanori Mizuno (Aichi Institute of Technology, Japan)

Jun Munemori (Wakayama University, Japan)

Yuko Murayama (Iwate Prefectural University, Japan)

Ken-ichi Okada (Keio University, Japan)

Norio Shiratori (Tohoku University, Japan)

Osamu Takahashi (Future University-Hakodate, Japan)

Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

Program Chair

Takuya Yosihiro (Wakayama University, Japan)

Program Committee

Behzad Bordbar (University of Birmingham, UK)

Naoya Chujo (Aichi Institute of Technology, Japan)
Satoru Fujii (Matsue College of Technology, Japan)
Hisao Fukuoka (Tokyo Denki University, Japan)
Teruyuki Hasegawa (KDDI R&D Laboratories, Japan)
Haruo Hayami (Kanagawa Institute of Technology, Japan)
Takaaki Hishida (Aichi Institute of Technology, Japan)
Tomoo Inoue (University of Tsukuba , Japan)
Masahiko Ishino (Fukui University of Technology, Japan)
Yoshinobu Kawabe (Aichi Institute of Technology, Japan)
Gen Kitagata (Tohoku University, Japan)
Tomoya Kitani (Shizuoka University, Japan)
Ryozo Kiyohara (Kanagawa Institute of Technology, Japan)
Tsukasa Kudo
(Shizuoka Institute of Science and Technology, Japan)
Hiroshi Mineno (Shizuoka University, Japan)
Yasue Mitsukura (Keio University, Japan)
Shinichiro Mori (Fujitsu Laboratories, Japan)
Jun Munemori (Wakayama University, Japan)
Yuko Murayama (Iwate Prefectural University, Japan)
Yoshitaka Nakamura (Future University-Hakodate, Japan)
Masakatsu Nishigaki (Shizuoka University, Japan)
Masashi Saito (Mitsubishi Electric Corporation, Japan)
Yoshia Saito (Iwate Prefectural University, Japan)
Fumiaki Sato (Toho University, Japan)
Jun Sawamoto (Iwate Prefectural University, Japan)
Toshifusa Sekizawa (Osaka Gakuin University, Japan
Hiroshi Shigeno (Keio University, Japan)
Toshihiro Shikama (Fukui University of Technology, Japan)
Yoh Shiraishi (Future University-Hakodate, Japan)
Hideyuki Takahashi (Tohoku University, Japan)
Osamu Takahashi (Future University-Hakodate, Japan)
Noriko Takaya (NTT Corporation, Japan)
Yoshiaki Terashima (Mitsubishi Electric Corporation, Japan)
Takaaki Umedu (Osaka University, Japan)
Hirozumi Yamaguchi (Osaka University, Japan)
Koji Yoshida (Shonan Institute of Technology, Japan)
Tomoki Yoshihisa (Osaka University, Japan)
Takaya Yuizono
(Japan Advanced Institute of Science and Technology,)
Yuji Wada (Tokyo Denki University, Japan)

Session 1:
Intelligent Transportation Systems

(Chair : Tomoya Kitani)

A Method for Estimating Road Surface Conditions with a Smartphone

Tomohiro Nomura[*] and Yoh Shiraishi[**]

[*] Graduate School of Systems Information Science, Future University Hakodate, Japan
[**] School of Systems Information Science, Future University Hakodate, Japan
{g2113026, siraisi}@fun.ac.jp

Abstract - In recent years, GPS (Global Positioning System) sensors, acceleration sensors and so on have been embedded in smartphones and become popular. We can gather various kinds of information simply and on a massive scale by using such smartphones. A system that creates new information from various kinds of information and shares such information through a network is called a “probe information system”. Recently, such probe information systems are focused and used to share traffic information. In this study, we focus on road surface conditions concerning comfort driving and ride quality. We need to share facts, for example when ruts appear in a road because road conditions are changeable. Therefore, this study proposes a method for estimating and detecting changes in road surface conditions by using a smartphone. The proposed method uses acceleration sensors embedded in smartphones and estimates road surface conditions. Then, the method detects changes in the conditions by comparing the latest estimation results with past results. The proposed method can confirm that road conditions changed in winter, even in the same segment.

Keywords: probe information system, smartphone, log data, sensing, estimation of road surface condition

1 INTRODUCTION

Recently sensors with small size and high performance have become widespread due to the development of MEMS (Micro Electric Mechanical Systems), and are embedded in various kinds of objects in our living environment such as personal computers, beacons, cars and smartphones. Many researchers tackle advanced studies in the field of mobile sensing. Such mobile sensing uses sensors embedded in moving objects such as cars, bikes and smartphones, and regards cars and humans as sensors. Moreover the penetration rate of smartphones is increasing and will continue to do so. Many sensors such as acceleration sensors, gyro sensors and so on are embedded in smartphones. Consequently, we can develop convenient and efficient systems with low cost and gather various kinds of information simply and on a massive scale [1]. The gathered information is called “probe information.” A system that generates new information from probe information and shares such information through a network is called a “probe information system” [2]. A conventional sensory system can only gather information on a road that has stationary devices, and it is necessary to increase the number of devices in order to extend the range over which information can be gathered. On the other hand, a probe information system can gather various kinds of information because it gathers information by using cars and humans as sensors without locating stationary devices such as beacon devices [3-5]. This study focuses specifically on traffic information.

There are road bumps and road surface conditions affecting comfort driving and ride quality. In winter, uneven surface form on roads in snowy regions due to snow and ice. In addition, ruts form on a road due to the temperature rising in the daytime and cooling at night. It is difficult for drivers to drive because road surface conditions change depending on the season or time, even on the same road. Accordingly, it is necessary not only to grasp present road surface conditions accurately but also to detect the changes in road surface conditions. It will become possible to navigate roads that are comfortable for driving and do not change if we can gather the road surface conditions as probe information and detect the changes in road surface conditions.

However, some related works on estimating road surface conditions have a problem regarding cutting costs because they need to introduce stationary devices on roads and in-vehicle cameras. Moreover, as mentioned above, we have to consider the change of road surface conditions to grasp them accurately because they change depending on the season or time.

Our study proposes a method for detecting changes in road surface conditions, and solves the problems of introductory costs and robustness by using a smartphone and its acceleration sensors. The proposed method calculates the variance of the vertical component of acceleration values when a driver drives a car. The method classifies road surface conditions into three levels: rough road level 0, rough road level 1 and rough road level 2. Moreover, this method partitions a road from intersection to intersection into multiple segments, and estimates road surface conditions on each partitioned segment. In addition, this method achieves detection of changes in road surface conditions through comparing the result of the latest estimation and past estimations.

2 RELATED WORK

There are some related works on estimating road surface conditions. One is an approach that uses the polarization property of fixed cameras. Another uses in-vehicle cameras. A further work uses acceleration sensors. We explain the details of these works in this section.

2.1 An approach using fixed cameras

There is a study that uses fixed cameras on poles on a road for estimating road surface conditions [6]. This method estimates road surface conditions based on the polarization property of images by using fixed cameras. It locates a stationary device at each pole on a road. Then, it irradiates a light onto a road using stroboscopic illumination and captures images continuously with CCD cameras, then estimates road surface conditions by analyzing the captured images. It is possible to gather information regardless of brightness because the method takes images alternately while irradiating light and while not irradiating light. Moreover, this study classifies road surface conditions into four states: dry, wet, snowy and freezing. The characteristics of the four states are shown in Table 1.

Table 1: Characteristics of each road surface condition (Quoted from [6]).

State	Degree of reflectivity	Brightness	Road temperature	Characteristics
Dry	Low	Low	-	Even and a little dark.
Wet	High	Low – Medium	Above -3°C.	Degree of reflectivity is high.
Snowy	High	Low – Medium	Below -3°C.	Degree of reflectivity is high.
Freezing	Low	High	-	Even and bright.

This method needs to introduce fixed cameras, and the introductory cost of the devices is high. In addition, it has a problem with granularity of the system structure because the estimation targets are limited to roads that have stationary devices.

2.2 An approach using in-vehicle cameras

Recently, cars mounted with cameras are increasing due to the development of image processing techniques. Cameras used on cars are called in-vehicle cameras, and many researchers consider various kinds of applications using these cameras [7-10]. The method [7] estimates road surface conditions by using the characteristic of image brightness acquired from in-vehicle cameras. This study assumes that a coefficient of friction is low if the road is bright, and proposes a method for estimating road surface conditions based on the degree of brightness of the road surface. The brightness of the road is derived from images taken by in-vehicle cameras which monitor the area in front of the car.

When reflected sunlight spreads across the surface of a dry road, the brightness signal from in-vehicle cameras becomes constant. On the other hand, when a wet road surface becomes like a mirror because it is covered by water, the brightness signal is non-constant, because the reflective areas of wet road surface are not evenly ranged.

This study estimates whether the road surface is dry or wet by using these brightness signals. The introductory cost of this method is lower than that of the method that uses fixed cameras. This method also solves the problem of granularity. However, the method cannot estimate the conditions in bad weather and at night because the method uses sunlight. Accordingly, this study has a problem regarding robustness.

2.3 An approach using acceleration sensors

There are some studies of detecting road bumps using acceleration sensors [11-16]. The method [11] is based on iterating multibody analysis and uses multibody vehicle model. The method [12-14] uses three-axis acceleration sensors and GPS sensor embedded in a vehicle. The method [15-16] involves placing a smartphone on the dashboard of a car, and can detect road bumps only during driving. The IRI (International Roughness Index), an index of flatness of road surfaces, has a relationship with the RMS (Root Mean Square) of the vertical component of acceleration values. Therefore, the study [16] proposes a method for estimating the height and length of bumps using acceleration sensors. This method estimates the amount of vertical displacement using the double integral of the vertical component of acceleration values, and defines it as the height of a bump. In addition, this method estimates distance travelled forward using GPS sensor, and defines it as the length of a bump. In Figure 1, results of estimation of road bumps and actual visual check of road bumps are shown. The results of estimation of road bumps are shown as blue circles, and visual check of road bumps is shown as orange circles.

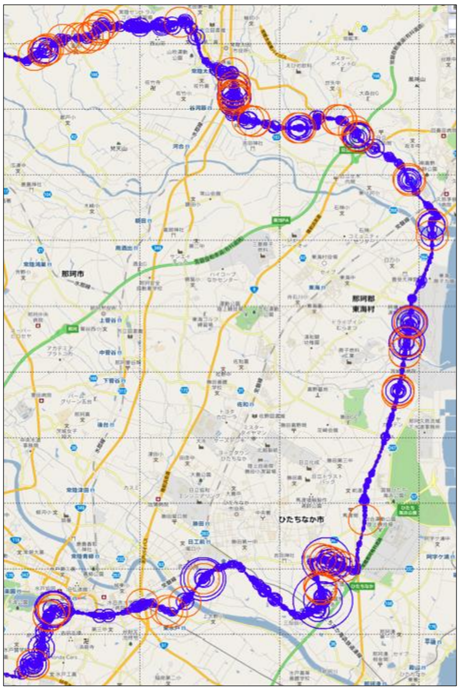


Figure 1: The relationship between estimation result and visual check (Quoted from [16]).

This method can estimate road surface conditions rapidly at low cost by using smartphones. However, road bumps that do not cause fellow passengers to feel vibrations are detected, and road bumps that do cause fellow passengers to feel vibrations are not detected. Therefore, this method has a problem with accuracy. Moreover, a map like the above does not present changes in road surface conditions depending on season or time. In winter, information on whether the condition of rough roads is constant or changeable is important to help drivers select the best route. Accordingly, this study has a problem inadequate detection of changes in road surface conditions.

2.4 Summary of related works

In Table 2, we show the advantages and disadvantages of the related works mentioned in this section.

Table 2: Advantages and disadvantages of related works.

	Estimation target	Estimation accuracy	Introductory cost and estimation granularity	Robustness	Inadequate detection of changes in road surface conditions
Fixed cameras [6]	Wet, dry, snowy and freezing roads	Y	N	Y	-
In-vehicle cameras [7]	Wet and dry roads	Y	Y	N	-
Acceleration sensors [16]	Road bumps	N	Y	Y	N

The approach using fixed cameras has problems such as introductory cost and estimation granularity. The approach using in-vehicle cameras solves these problems. However, this method has a problem of robustness. The approach using acceleration sensors is able to estimate the height and length of road bumps. This method solves the problems of introductory cost, granularity and robustness because it uses a smartphone. However, it has a problem with accuracy, and inadequate detection of changes in road surface conditions.

For these reasons, in this study we propose a method for solving problems such as introductory cost, estimation granularity and robustness by using smartphones. This method can detect changes in road surface conditions from hour to hour. The proposed method estimates road surface conditions and gathers results of estimation, and compares the latest results with past results.

3 PROPOSED METHOD

In this section, we explain an approach for solving problems in the related works and the purpose of this study. We explain an overview of the proposed method and the details.

3.1 Purpose and approach

The related works have problems such as introductory cost, estimation granularity, scale of robustness and inadequate detection of changes in road surface conditions. We propose a system that can gather driving log data at low cost and is robust, and which compares the latest estimation and past estimations to solve the problems with the related works.

Therefore, in this study, our method gathers driving log data by using the sensors of smartphones, and estimates road surface conditions using only the gathered log data. Moreover, the method manages the estimated results in a database, and detects changes in road surface conditions by comparing the latest estimation with past estimations. Recently various kinds of sensors are embedded in smartphones, and driving log data gathered by the sensors of smartphones have a high degree of usability. Smartphones are often used on the dashboard of a car because it has features such as audio and navigation applications. Therefore, smartphones can reduce the burden for drivers when we gather driving log data. In addition, sensors embedded in smartphones can robustly gather the driving log data because they can be used in all weathers and at all times. For these reasons, we think that it is possible to gather robust sensor data at low cost by using smartphones. This study aims to estimate road surface conditions and to detect changes in road surface conditions by using driving log data gathered by smartphones.

3.2 An overview of the proposed system

Smartphones on car dashboards gather vertical components of acceleration values, location information such as latitude, longitude and time stamps during driving. The gathered driving log data are managed in a database, and we estimate road surface conditions and detect changes of them by using the driving log data. An overview of the proposed system is shown in Figure 2.

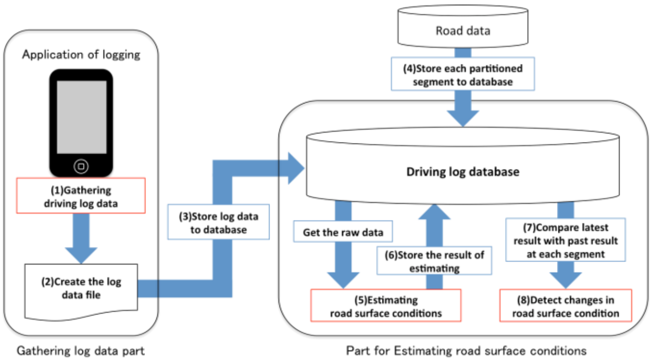

Road data
(4)Store each partitioned segment to database
Application of logging
(1)Gathering driving log data
(2)Create the log data file
(3)Store log data to database
Driving log database
Get the raw data
(6)Store the result of estimating
(7)Compare latest result with past result at each segment
(5)Estimating road surface conditions
(8)Detect changes in road surface condition
Gathering log data part
Part for Estimating road surface conditions

Figure 2: An overview of the proposed system.

The proposed system consists of two parts: gathering log data part and estimating road surface conditions part. First, the gathering log data part gathers the vertical components of acceleration values, location information, dates and time stamps by using the acceleration sensor and GPS sensor of a smartphone (1). Next, the part generates the log data file, and stores log data into the database (2), (3). Our method partitions a road from intersection to intersection into multiple segments, and creates a road segment table that it then stores into the database as a preliminary preparation (4). Then, the estimating part on road surface conditions estimates road surface conditions using gathered log data, and manages the estimated results as attribute data of segments with dates and time stamps (5), (6). Finally, the part compares the latest result of estimation with past results of estimation at each segment, and detects changes in road surface conditions (7), (8).

3.3 Estimating road surface conditions

In this section, we explain how to gather driving log data, estimate road surface conditions and detect changes in road surface conditions.

3.3.1. Gathering driving log data

Our method gathers driving log data by using smartphones on the dashboards of cars. The log data to be gathered are date, time stamp, 3-axis acceleration values, 3-axis gyro values, speed of car, latitude, longitude and direction. Smartphones gather this information every 100 [Hz]. The driving log data are stored into a raw data table of the database through the mail function of smartphones. The structure of log data table is shown in Table 3.

Table 3: Structure of log data table.

Attribute	Detail
Id	Id of log data
date	Value of date
time	Time stamp
accx	3-axis acceleration values
accy	
accz	
pitch	3-axis gyro values
roll	
yaw	
speed	Speed of car
lat	Latitude
lon	Longitude
direction	Direction

This study focuses on vertical component of acceleration values because road bumps have most influence on vertical component of acceleration values in these log data. Vertical component of acceleration values change widely when cars bounce over road bumps. Accordingly, we estimate road surface conditions using changes in the vertical component of acceleration values.

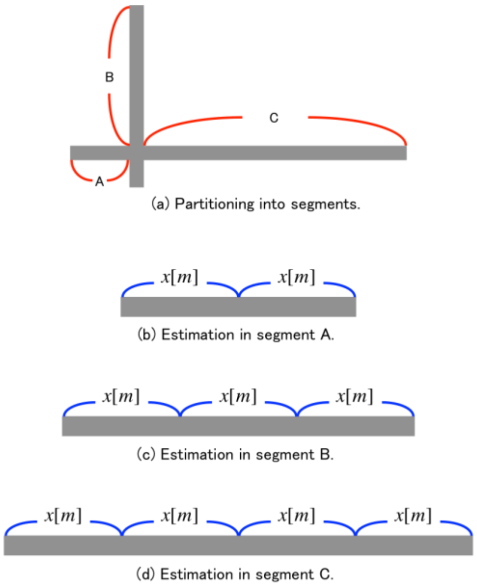
3.3.2. Estimating of road surface conditions

There are various kinds of road surface conditions. They are classified as point information and line information if we focus on comfort driving and ride quality. The point information is expected partial change when cars pass over a manhole or road bump. The method for detecting them is explained in a paper by Koichi Yagi [15,16]. On the other hand, the line information is the states of a segment, such as a minor bounce segment and bad ride quality segment, when we focus on units from intersection to intersection. This estimation method is not touched in any related works. We can use the navigation system to avoid a bad ride quality segment in advance if we grasp the road surface conditions of the segment. This study defines the conditions of the segment using the three levels shown in Table 4.

Table 4: Definition and classification of rough road levels.

	Feature	Measure of continuous bounce
Rough road level 0	A flat road on which no bounce is felt.	Small
Rough road level 1	A road on which bounce is felt in certain spots due to asphalt damage.	Medium
Rough road level 2	A road on which bounce is felt continuously, such as a dirt road.	Large

The differences appear when we observe driving on a single segment of road. We can expect that variability of acceleration values is low when we drive on a road classed as rough road level 0. We can expect that variability of acceleration values is a little higher when we drive on a road classed as the rough road level 1, and very high when we drive on a road classified as rough road level 2. The proposed method classifies roads into three levels by calculating the variance of vertical component of acceleration values in a constant segment and setting thresholds. The thresholds are explained in Section 4.2. We need to estimate every fixed distance because we have to classify same segment when we use thresholds. However, the lengths of actual roads and driving routes vary. In addition, we need to compare estimation results for each same segment to detect changes in road surface conditions. For these reasons, the proposed method partitions a road from intersection to intersection into multiple segments. Moreover, the proposed method further partitions segments into sub-segments of constant length, and estimates the road surface conditions of each partitioned segment, in order to solve the problem of the varying length of segments. Figure 3 show an illustration of partitioning into segments and estimating in the partitioned segment, and Table 5 show a sample data of segment table.



(a) Partitioning into segments.

(b) Estimation in segment A.

(c) Estimation in segment B.

(d) Estimation in segment C.

Figure 3: An illustration of partitioning into segments and estimating in the partitioned segment.

Table 5: A structure of segment table and example data.

Attribute	seg_id	str_lon	str_lat	end_lon	end_lat
Detail	Id of the segment	Longitude of start point	Latitude of start point	Longitude of end point	Latitude of end point
Example	1	41.843252	140.768283	41.840409	140.767791
-	…	…	…	…	…

Actual lengths from intersection to intersection are different, as (a) in Figure 3. Therefore, the proposed method continues to calculate every x $[m]$ until it reaches the end of the segment, as shown in Figure 3(b), (c) and (d). 'x $[m]$ ', the variable used for calculating the interval of variance, is explained in Section 3.4. The results of estimating every x $[m]$ are output as three integers (0: rough road level 0, 1: rough road level 1, 2: rough road level 2). We define the calculation order as i, the result of estimating every x $[m]$ as $Var(i)$, the calculation count of variance as n and the ID of each segment as seg_id. Then, the result of estimating each segment, $Est(seg\_id)$, is calculated as follows (after the decimal point is rounded):

$$Est(seg\_id) = \frac{\sum_{i=1}^{n} Var(i)}{n} \qquad \cdots(1)$$

This $Est(seg\_id)$ is stored into the database table of estimation results with the estimation result ID, date and time stamp. Table 6 shows the structure of the table of estimation results.

Table 6: Structure of the table of estimation results and example data.

Attribute	est_id	seg_id	date	time	estimation
Detail	ID of estimation result	ID of a segment	Date of estimating result	Time stamp of estimation result	Output integer of estimation result
Example	1	1	2014/03/20	12:09:40:666	0
-	…	…	…	…	…

In this way, we classify roads of various lengths into three levels. Moreover, comparing the results of the same segment is enabled by recording estimation results in units of segments.

3.3.3. Detecting changes in road surface conditions

We detect changes in road surface conditions by using a table of estimation results. In the table of estimation results, we compare the result of estimation with any past data that have the same segment ID and time. We divide time into three periods: 0:00-8:00, 8:00-16:00 and 16:00-24:00.

We compare the results of estimation using the following process. We define the number of estimation results which matching segment ID and time as n, the order of managed data as i and the result of estimation as $Est(i)$. Then, the average of past results $Past\_Est(seg\_id)$ is calculated as follows (after the decimal point is rounded):

$$Past\_Est(seg\_id) = \frac{\sum_{i=1}^{n} Est(i)}{n} \qquad \cdots(2)$$

The reliability of estimation results is high when there is a wide range of past data to refer to. However, this changes depending on the season. In summer, road surface conditions do not change frequently. For this reason, it is appropriate to compare the average of the past month's results with the latest result to detect the latest changes. In contrast, in winter in snowy regions road surface conditions change frequently, meaning that comparing the average of the past month's results with the latest result is not effective. Therefore, this study detects changes in winter road surface conditions by comparing the result from the same time on the previous day and the average of the past week's results with the latest result. In this way, the range of reference data varies according to the season and purpose. Therefore, we define that the range is set based on season and purpose, and is set by the user for detecting changes in road surface conditions.

Next, we compare the latest estimation $Latest\_Est(seg\_id)$ with $Past\_Est(seg\_id)$. The $Past\_Est(seg\_id)$ is equal to the $Latest\_Est(seg\_id)$ if the road are constantly good or bad, and we can assume that the road surface conditions do not change. However, we can consider that $Past\_Est(seg\_id)$ is not equal to the $Latest\_Est(seg\_id)$ if the road surface conditions have changed within the past few hours. Accordingly, we can assume that road surface conditions of a segment where $Past\_Est(seg\_id)$ is not equal to $Latest\_Est(seg\_id)$ have changed.

4 EXPERIMENTS AND DISCUSSIONS

In this section, we explain the experiments we conducted for confirming the effectiveness of the proposed method and discuss the results. We fixed the vehicle speed in the following experiments.

4.1 Implementation

We implemented a system that gathers log data, estimates road surface conditions by using the log data of a database and visualizes the results. We implemented the processing of estimation of road surface conditions and detection of changes in these conditions by using Java and JDBC (Java Database Connection), and we also implemented an application for visualizing the results of estimation by using JavaScript.

4.2 Preliminary experiment

We conducted a preliminary experiment to set the calculation interval of variance and the thresholds that are explained in Section 3.3.2. To gather log data we implemented a logging application on iOS. We set the smartphone horizontally on the dashboard, as in Figure 4. In this case, the y-axis of acceleration becomes the vertical component of acceleration values, as shown in Figure 5.

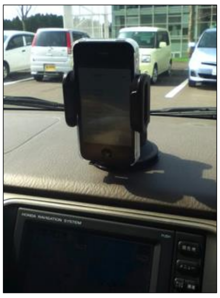


Figure 4: Image of setting a smartphone.

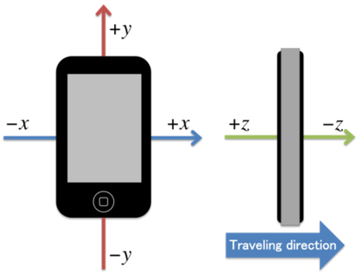


Figure 5: Acceleration axis.

We drove on three levels (rough road level 0, level 1 and level 2) of road while running the logging application, and examined the suitable interval of calculation of variance, and then we set the thresholds. We examined changes in the variance when we changed the interval of calculation of variance. The intervals of calculation of variance are every 5 meters, every 10 meters, every 20 meters and every 40 meters. The variances of vertical component of acceleration values when we drove on the three levels of road in each calculation interval are shown in Figures 6 and 7.

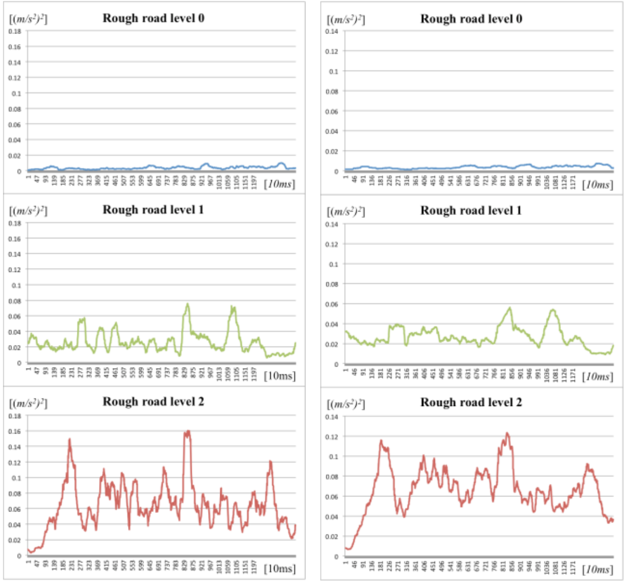


Figure 6: Variance in every 5 and 10 meters.

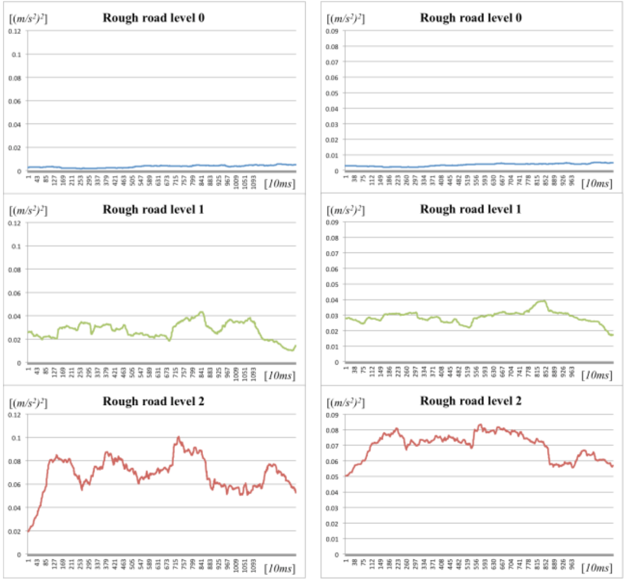


Figure 7: Variance in every 20 and 40 meters.

These results show that variances when we drove on the three levels of road were different, and the magnitude relation of variance of each level is non-constant in the cases where the calculation intervals of variance are every 5 meters, every 10 meters and every 20 meters. Therefore, we cannot define the thresholds that classify the rough road level distinctively. On the other hand, the magnitude relation of variance of each level is constant in the case where the variance calculation interval is every 40 meters. Accordingly, we can classify roads into three levels if we set thresholds. For these reasons, we set the variance calculation interval at every 40 meters to classify roads into three rough road levels by threshold. In addition, we repeated this preliminary experiment, and set the threshold of discriminating rough road level 0 and 1 at 0.0190$[(m/s^2)^2]$, and discriminating level 1 and 2 at 0.0428$[(m/s^2)^2]$.

4.3 Experiment to evaluate accuracy

We conducted an experiment to evaluate the accuracy of estimation of road surface conditions by using the calculation interval and thresholds established in Section 4.2. We analyzed the estimation result of each segment that uses the proposed method by using a visualization application. Then, we defined a video captured by an in-vehicle camera tracking on an actual road as correct data, and compared the estimation results with the correct data. The results of this experiment are shown in Table 7.

Table 7: Detail of estimation results.

		Estimation result of rough road level		
		Level 0	Level 1	Level 2
Correct data of rough road level	Level 0	31	0	0
	Level 1	3	10	0
	Level 2	0	0	5

The shaded areas of Table 7 show the number of segments for which estimation was accurate, and the remaining squares show the number of segments for which estimation was not accurate. These results show that rough road level 0 and rough road level 2 were successfully estimated but there were false estimations when a rough road level 1 was estimated as rough road level 0. We analyzed the log data of the false estimations to find the cause. The estimation result of a segment in which false estimation occurred is shown in Figure 8(a). The estimation result of every 40 meter interval in the segment is shown in Figure 8(b). A rough road level 0 is indicated by a blue line, and level 1 is indicated by a green line in the figures.

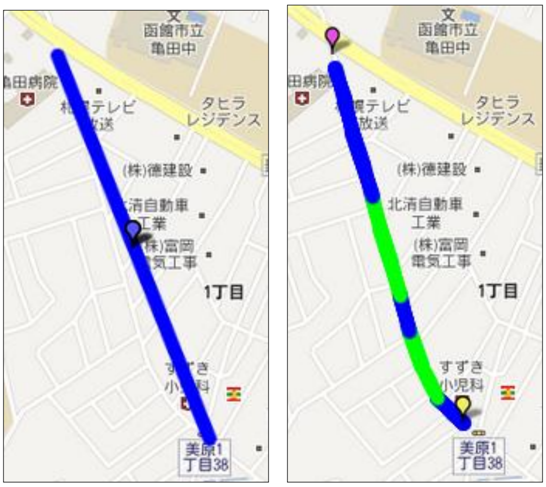

(a) Segment (b) Every 40 meters

Figure 8: Estimation result of a segment and of every 40 meter interval within the segment.

The segments in these figures are segments of a rough road level 1, as defined using in-vehicle video capture. In fact, the estimation results of every 40 meter interval show that rough road level 1 was estimated at multiple sites, as shown in Figure 8(b). However, the estimation result of the segment using the proposed method shows it as a rough road level 0.

4.4 Elementary experiment for detecting changes

We conducted an elementary experiment to determine whether or not road surface conditions changes depending on date and time. We analyzed driving log data in winter when road surface conditions often change by using a visualization application. The estimation result of driving log data collected when we drove in a snowy segment at night is shown in Figure 9(c), and the estimation result of driving log data collected when we drove in the same segment in good weather on the following day is shown in Figure 9(d). The rough road level 0 is indicated by a blue line, level 1 by a green line and level 2 by a red line in these figures.

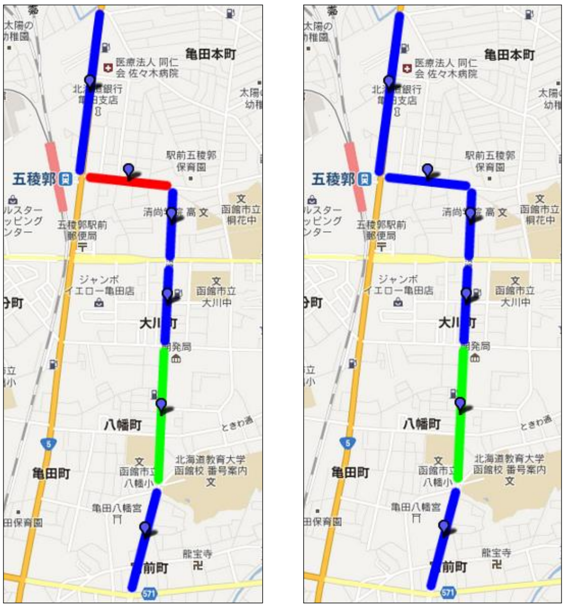

(c) Snowy day (January 23th) (d) Later fine day (January 29th)

Figure 9: Estimation result of a snowy day and later fine day.

On the day in which the data used in Figure 9(c) was collected, some roads were covered in snow due to snowing the previous day. The red line in Figure 9(c) indicates a bumpy road surface caused by snow. Accordingly, there was a high degree of bounce throughout the segment. However, a change in the estimation result of this segment can be seen in Figure 9(d) because the temperature increased and the snow melted on daytime.

4.5 Discussion of experiment results

In this section, we discuss the experiment results. The experiment to evaluate accuracy revealed that the accuracy rate of estimating road surface conditions using the proposed method was about 94 percent. For this reason, we could confirm that we can effectively classify a road into three rough road levels by using the variance of vertical component of acceleration values. Moreover, we were also able to confirm that the rough road level 0 and 2 can be estimated accurately. However, there was some false estimation of rough road level 1. This is thought to be due to the fact that in level 1 bounce occurs only in certain spots. The length of some segments is long because the proposed method partitions a road from intersection to intersection. In addition, it is not always true that bounce is felt continuously, even in a road that is a rough road level 1. Therefore, the proposed method confuses level 1 with level 0 when it estimates the road as segments. Accordingly, we need to consider an improved method such as adding new parameters and further partitioning longer segments.

From the elementary experiment for detecting changes, we were able to confirm that the rough road levels changed in winter, even in the same segment.

5 CONCLUSION

In this study, we proposed a method for estimating and detecting changes in road surface conditions using a smartphone. In addition, we implemented the system of the proposed method and conducted experiments to confirm the effectiveness of the proposed method. In the future, we will consider a method to improve estimation accuracy and implement a system for detecting and visualizing changes in road surface conditions. We will also consider the influence of vehicle speed.

REFERENCES

[1] Jeffrey Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy and Mani B Srivastava, “Participatory sensing,” World Sensor Web Workshop, ACM Conference on Embedded Networked Sensor Systems 2006.

[2] Keisuke Uehara, “Probe vehicle system: getting environmental information using vehicle sensors,” IPSJ Magazine, Vol.51, No.9, pp.1144-1149, 2010 *(in Japanese).*

[3] Hitoshi Yamauchi , Akira Tomono and Akihiro Kanagawa, “A road map making probe system by integration of road shapes with roadside hue information,” IPSJ Journal, Vol.52, No.1, pp.257-268, 2011 *(in Japanese).*

[4] Hitoshi Yamauchi, Fumiaki Oka and Hiromitsu Takahashi, “Unification of road map informations generated by multiple probe cars,” IEICE Technical Report ITS, Vol.106, No.534, pp.109-114, 2007 *(in Japanese).*

[5] Akihito Kubota, Norio Kitajima, Yuki Kobayashi and Satoshi Ichimura, “A safe driving support system for bicycles which can share road conditions and traffic,” IPSJ SIG Technical Reports, Vol.72, No.18, pp.1-6, 2009 *(in Japanese).*

[6] Yoshiyuki Miyazaki, Hiroshi Okabe, Kazuhiro Tanji, Masaya Otsuki and Shinya Fujiwara, “Image type road surface recognition sensor using reflection characteristic of the light,” IEICE Technical Report ITS, Vol.102, No.223, pp.43-46, 2002 *(in Japanese).*

[7] Tetsuya Kuno, Hiroaki Sugiura and Junichi Yoshida, “Detection of road conditions with CCD cameras mounted on a vehicle,” The IEICE Transactions D-II, J81-D-2, No.10, pp.2301-2310, 1998 *(in Japanese).*

[8] Michio Tanaka, Takashi Morie, Satoru Matsuoka, Koji Iwase and Yasunori Yamamoto, “Dry/Wet judgment of road surface using Gabor filtering of vehicle camera images,” ITE Technical Report, Vol.35, No.9, pp263-267, 2011 *(in Japanese).*

[9] Tatsuya Furukane, Keiji Shibata and Yuukou Horita, “Distinction of road surface condition in road image illuminated by car headlight at night-time,” ITE Technical Report, Vol.35, No.7, pp.11-14, 2011 *(in Japanese).*

[10] Yusuke Hinagata, Mikio Bando and Yukihiro Kawamata, “In-vehicle stereo camera system for off-road environments,” Multimedia, Distributed, Cooperative, and Mobile (DICIMO2012) Symposium, pp.1375-1381, 2012 *(in Japanese).*

[11] Naoki Takahashi, Taichi Shiba, Keisuke Morita and Yoshimitsu Endo, “Estimation of road profile with multibody vehicle model,” Proceedings of the 19th JSME Transportation and Logistics, pp.105-108, 2010 *(in Japanese).*

[12] Kongyang Chen, Mingming Lu, Xiaopeng Fan, Mingming Wei and Jinwu Wu, “Road condition monitoring using on-board three-axis accelerometer and GPS sensor,” Proceeding of the 6th International Conference on Communications and Networking in China, pp.1032-1037, 2011.

[13] Jakob Eriksson, Lewis Girod, Bret Hull, Ryan Newton, Samuel Madden and Hari Balakrishnan, “The Pothole Patrol: Using a mobile sensor network for road surface monitoring,” Proceeding of the 6th International Conference on Mobile Systems, Applications, and Services, pp.29-39, 2008.

[14] Kasun De Zoysa, Chamath Keppitiyagama, Gihan P.Seneviratne and W.W.A.T Shihan, “A public transport system based sensor network for road surface condition monitoring,” Proceedings of the 2007 Workshop on Networked Systems for Developing Regions, 2007.

[15] Koichi Yagi, “Extensional smartphone probe for road bump detection,” 17th ITS World Congress, pp.1-10, 2010.

[16] Koichi Yagi, “Road bump detection method using accelerometer on smartphone and an application result of TOHOKU earthquake,” Proceedings of the 31th JSTE Workshop, pp.249-252, 2011 *(in Japanese).*

Method for Gathering Road Surface Conditions on Bikes with Smartphones

Natsumi Takahashi*, Seiji Matsuyama* , Hirohito Kakizawa** , Ryozo Kiyohara*

*Dept. of Information and Computer Science, Kanagawa Institute of Technology, Japan
** Graduate School of Information and Science, Kanagawa Institute of Technology, Japan
{s1121141, s1121032, s1485010@cce, kiyohara@ic}.kanagawa-it.ac.jp

Abstract - Smartphones are popular telematics terminals having several sensors and wireless communication functions. Smartphones can transmit a small amount of probe data from vehicles frequently. However, in developing countries and in Asia, we think a lot of people do not use telematics services because of the costs. Moreover, many people use bikes (motorbikes or bicycles). We therefore propose a method for gathering road surface conditions with bikes using smartphones. We experimented with a simple case to show that we can reduce the probe data used by the proposed system.

Keywords: Road surface management, probe data, smartphones, motorbikes and bicycles, sensing

1 INTRODUCTION

Recent years have witnessed the emergence of many telematics services. For example, vehicle information devices assess traffic conditions, weather information, emails, stolen vehicle information, and display the fastest route to a destination [1].

Typical services include OnStar [2] and G-book [3], among others [4] [5]. These telematics services gather a lot of information from many vehicles or smartphones having drivers as sensors. They analyze the data and return information to drivers simultaneously, as shown in Figure 1.

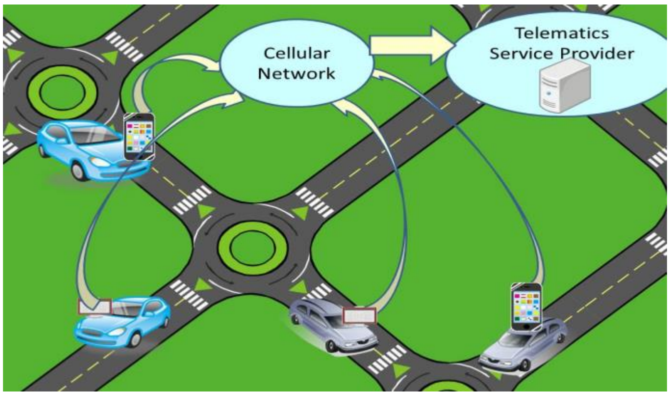


Figure 1. Telematics service

Smartphones are popular telematics terminals having several sensors and wireless communication functions. Each smartphone frequently transmits a small amount of probe data. However, in Asian and developing countries, we think few people opt to use the telematics service because of the costs. Many people, however, use motorbikes and bicycles.

More than ten billion vehicles are used in the world, and more than two billion motorbikes are among them [6]. Moreover, there are countless bicycles.

On the other hand, recent unseasonable weather has been seen throughout the world. Therefore, the costs of road surface management are increasing. However, such management is very important for preventive safety. In developing countries, many people ride motorbikes or bicycles, posing considerable problems to the drivers.

It cannot be expected that motorcycle drivers use telematics terminals on their motorbikes because of costs. However, many such drivers have and carry smartphones. Therefore, we propose a method for gathering road surface conditions with bikes using smartphones.

In Section 2, we introduce the maintenance of road surface conditions. In Section 3, we introduce the classification of data types. In Section 4, we propose a method for gathering the road surface conditions with bikes. In Section 5, we show the basic experiment and results. Then, in Section 6, we discuss the experiment's results before, in Section 7, concluding with a brief summary.

2 MAINTENANCE OF ROAD SURFACE CONDITION

In Japan, the maintenance of road surface conditions is defined as follows: “Road surfaces always need to be kept in good condition and repaired for safety and smooth traffic.” (Japanese Road Act). The subject of maintenance applies not only to road surfaces, but also to bridges, tunnels, and precipices.

There are innumerable roads throughout the world. Therefore, a core problem is how to find locations in need of maintenance. Patrols for the roads are very important for finding the problem areas. However, patrols are costly.

Therefore, in Japan, the bulk of patrols are directed to the roads that many vehicles pass. Government offices mainly provide regular maintenance to national roads. This is because the amount of traffic is large, and the average speed of the vehicles is higher than on other roads.

Table1. Standard number of patrols for national roads

Kind of road	Frequency of patrol
More than 50,000/day vehicles	Everyday
More than 5000/day vehicles	Every two days
Otherwise	Every three days

Table 2. Types of road surface management

Item	Detail
Fallen object	Suddenly appeared Moved by wind or Staying
Pothole	Suddenly appeared Staying
Difference in level	Suddenly or slowly appeared Staying
Rut	Slowly appeared Staying
Crack	Slowly appeared Staying

Table 3. Standard of crack depth to be repaired

Road surface condition	Repair
Crack depth is less than 20mm	Not required
Crack depth is more than 20mm and less than 40mm	The road should be repaired
Crack depth is more than 40mm	The road should be resurfaced

Table 1 shows the number of patrols for national roads in Japan. Table 2 shows the types of road surface management.

Table 3 shows a reference for determining how and when the road should be repaired. These rules are very strict, because many trucks pass by the road at high velocities.

However, the area used by national roads is a small part of the gross area for roads generally. Therefore, we focus on the non-national roads. On these roads, no vehicles pass by at high speeds, because these roads are typically narrow. Therefore, the requirements for road surface management are less strict.

In light of this, we propose a method for gathering the road surface conditions with bikes. There are plenty of motorbikes and bicycles on local roads in developing countries. Therefore, people-centric methods for collecting data, such as with smartphones [7], are effective in these cases. Thus, we propose a method for gathering road surface conditions with bikes using smartphones.

3 CLASSIFICATION OF DATA TYPE

Probe-data from smartphones can be classified as follows:

(1) Real-time data: this type of probe data is used for real-time monitoring, relating to traffic jams, disasters, etc.

(2) Accumulated data: this type of probe data is used for personal logs, average information about road conditions, etc.

(3) Emergency calling when traffic accidents occur, stolen vehicles, etc.

We focus on accumulated data, which is used for the management of road surface conditions. Some studies exist regarding the gathering of probe data generally [8] [9]. However there are no studies concerned with gathering accumulated probe data.

Accumulated probe data does not require gathering the data in real-time. Therefore the probe data is transmitted when the smartphones connect to Wi-Fi or other offloading access points.

4 PROPOSED METHOD

Our idea is very simple. Users gather a lot of information with their smartphones. Almost every smartphone now has GPS, accelerometers, etc. When users arrive at a wireless hot-spot, their smartphones will automatically send the relevant information, as shown in Figure 2. However, there is a smartphones' risk gathering too much data. Therefore, the data needs to be filtered. The filtering algorithm is based on selecting both the peak value and the frequency of the peak.

In many cases, smartphones are placed in the pocket, either at the chest or in slacks. In a few cases, users place smartphone in a handle fixed by a cradle. In the pocket, shockwaves are insulated by the human body. However, from the point of view of road-surface management, the

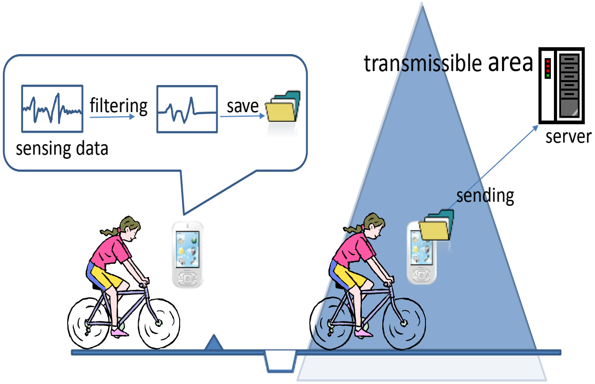


Figure 2. Gathering the probe data

shock sensed by a pocketed smartphone is a useful signal for the maintenance of roads. If the driver's body does not feel the shock, we can assume that any road-surface problems are not outstanding. Considering this, we propose the following method in four steps:

(1) In the smartphone, the accelerometer data, location data by GPS, and time-stamp data are logged.
(2) In suitable fixed-time intervals, logging data is analyzed and filtered.
(3) If the smartphone goes into an area where Wi-Fi is available for the device, the filtered logging data is transmitted to the service provider's server.
(4) The service provider analyzes the logging data and provides the road information to the road management office.

In this paper, we focus on the first and second steps. In the following sections, we demonstrate the feasibility of our idea.

5 EXPERIMENTATION

We experimented using a smartphone and a bicycle. Figure 3 shows the proposed method gathering data on a cracked road. On national roads in Japan, these cracks have to be repaired. However, on local roads, these cracks do not. Figure 4 shows an example of a pothole. Potholes should be repaired, even on local roads. However, motorbikes and bicycles often avoid passing directly over the pothole. Figure 5 and Figure 6 show a bumpy road with different levels. In this case, it is difficult for cyclists to avoid passing over them.

We pedaled a bicycle with a smartphone that sensed the acceleration as it remained in pockets on the chest and waist.

In the chest pocket, acceleration is constant with its user. Therefore, we expect the data gathered from a smartphone placed in a chest pocket to accurately report bumps and potholes. However, cracks are unlikely to be found with this data.

In the waist pocket, acceleration is also constant except for the noise generated by pedaling. We expect that we can distinguish pedaled noise from surface accelerations.

We experimented with a bicycle under the following conditions:

(1) Position of the smartphone: chest pocket, waist pocket

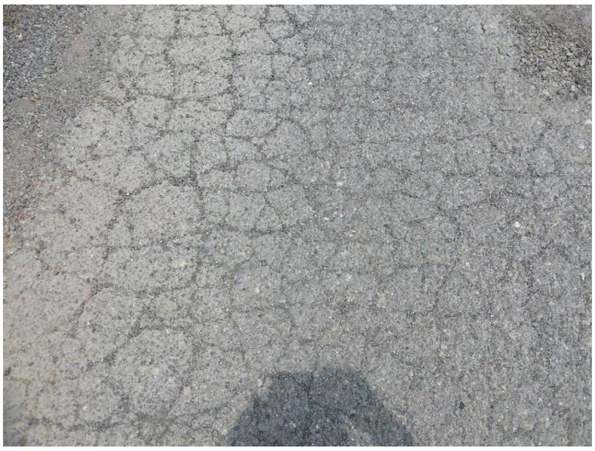


Figure 3. Example of cracked road

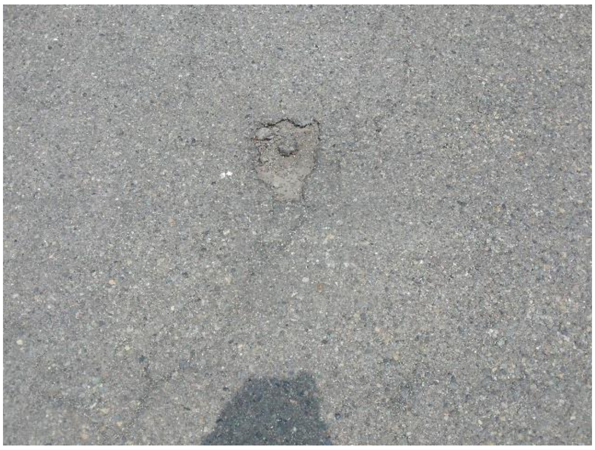


Figure 4. Example of a pothole

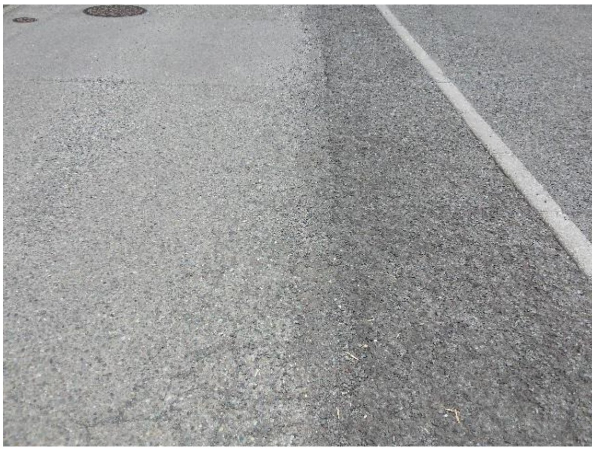


Figure 5. Example of an uphill bump

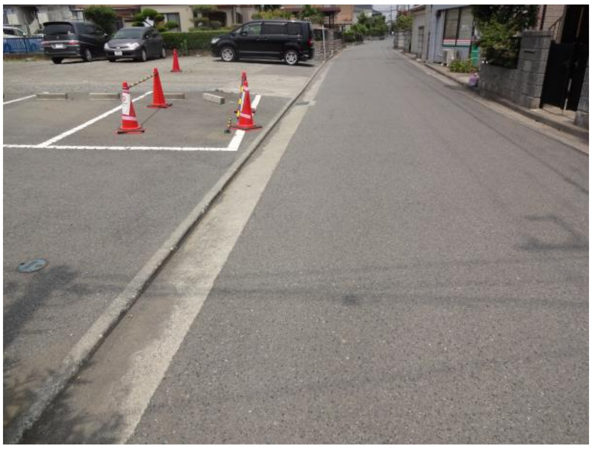


Figure 6. Example of a downhill bump

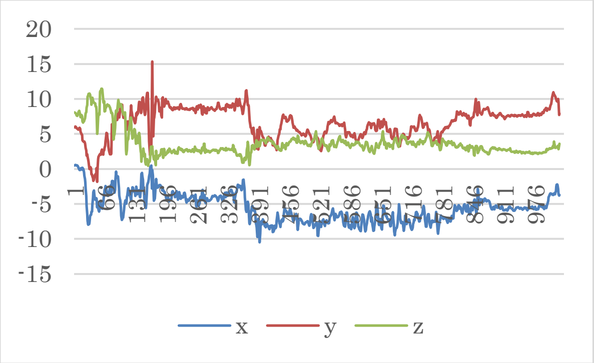


Figure 7. Vertical cracks, waist position, slow speed

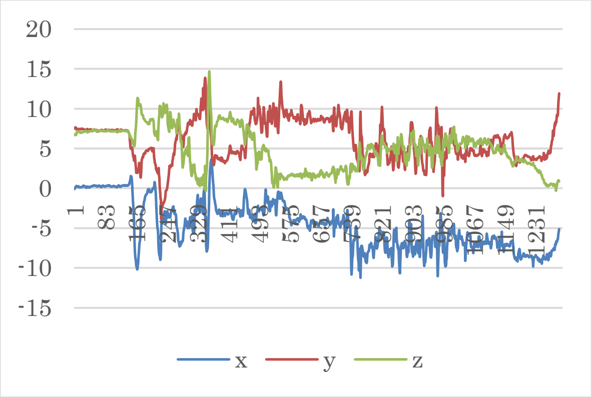


Figure 8. Vertical cracks, waist position, high speed

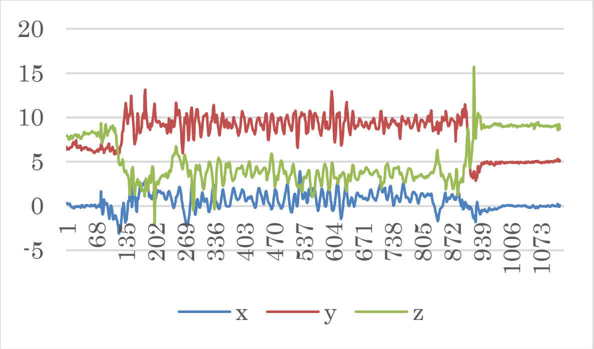
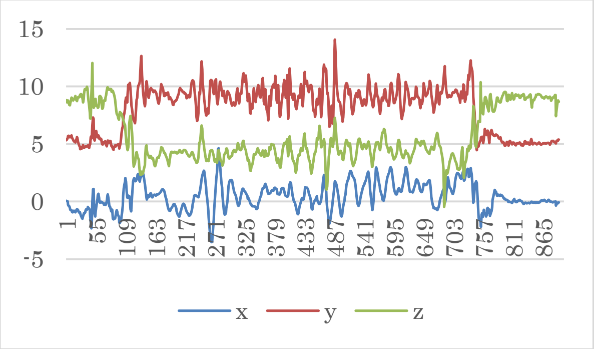


Figure 9. Vertical cracks, chest position, slow speed



Figrue 10 Vertical cracks, chest position, high speed

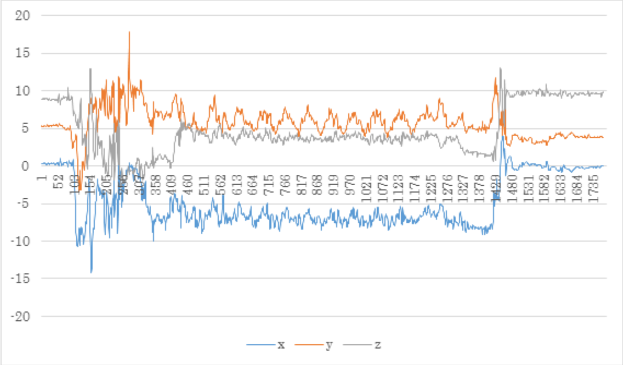


Figure 11. Normal road, waist position, slow speed

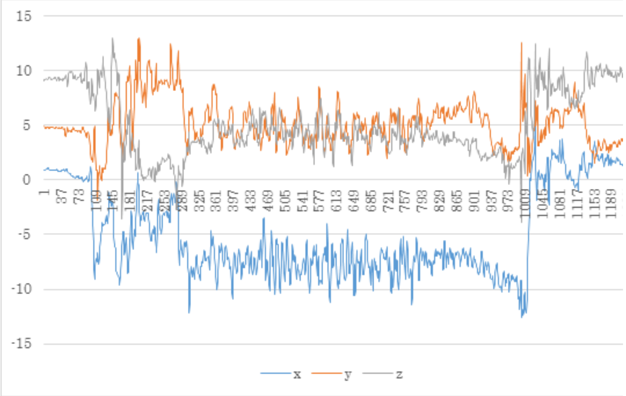


Figure 12. Normal road, waist position, high speed

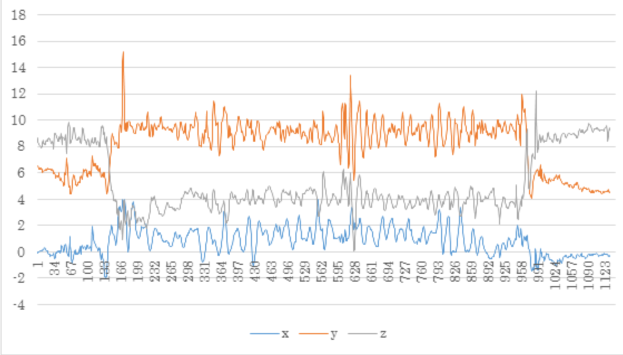
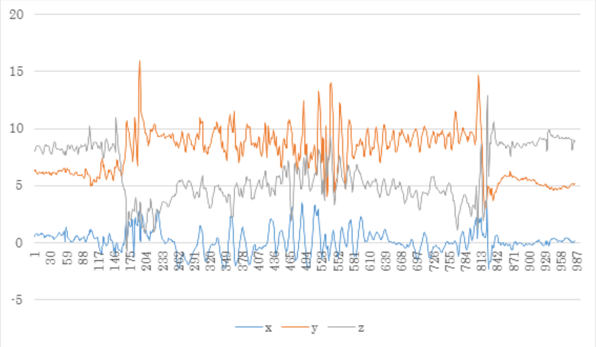


Figure 13. Pothole, chest position, slow speed



Figrue 14. Pothole, chest position, high speed

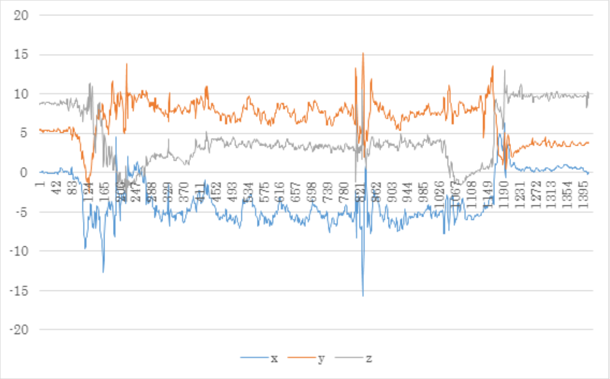


Figure 15 bump (up), waist position, slow speed

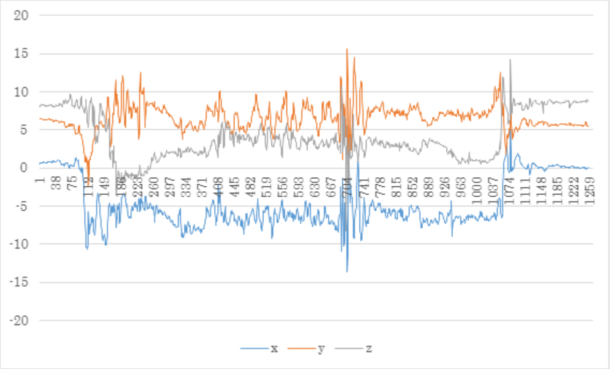


Figure 16 bump (up), waist position, high speed

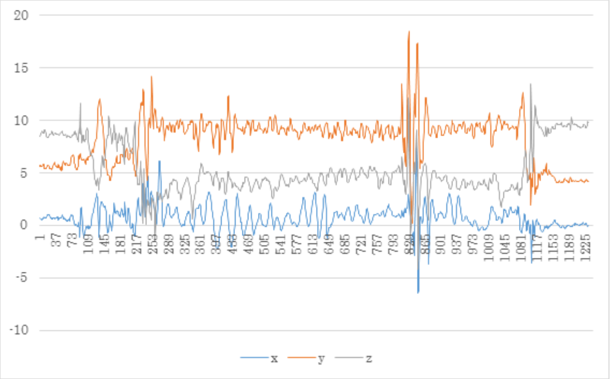


Figure 17 bump (up), chest position, slow speed

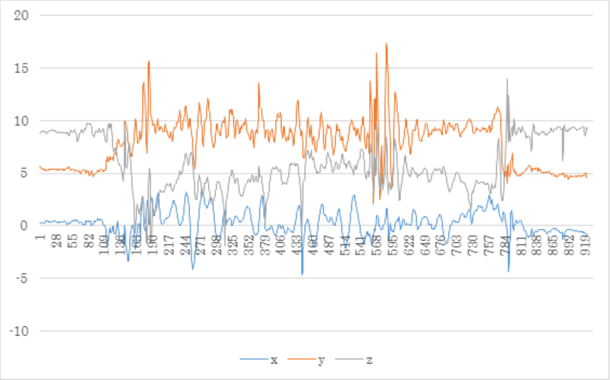


Figure 18 bump (up), chest position, high speed

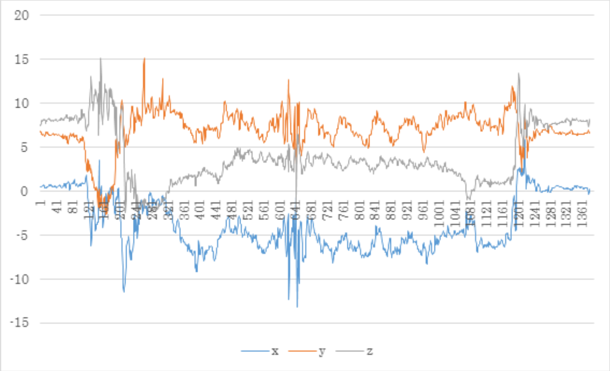


Figure 19 bump (down), waist position, slow speed

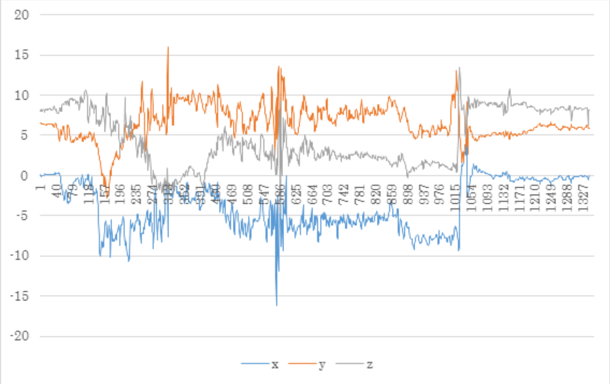


Figure 20. Bump (down), waist position, high speed

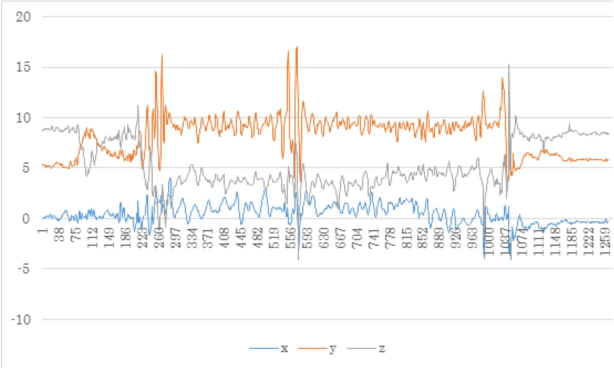


Figure 21. Bump (down), chest position, slow speed

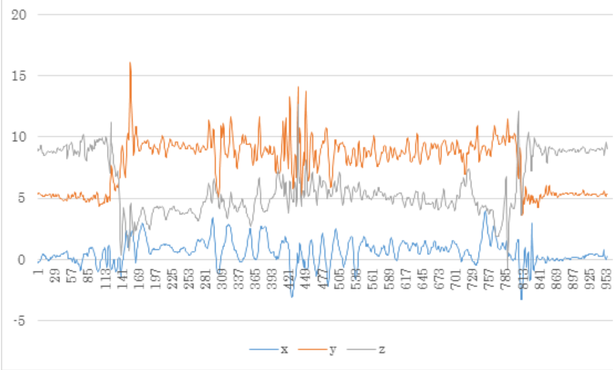


Figure 22. Bump (down), chest position, high speed

(2) Velocity: slow speed (about 4km/h), fast speed (20km/h)

(3) Road surface: cracks (vertical, crosswise), potholes, bumps (up and down)

Figure 7, Figure8, Figure 9, and Figure10 show the results of cracks with vertical.

Figure 11and Figure 12 show the normal road which surfaces are good conditions. In the both figure, the first and the last acceleration means the start and stop. Therefore, we need to focus on the central part of the graph. We think it is difficult to find cracked patterns. At the waist position, we can determine the cyclic noise—that is, the noise generated from pedaling. This noise is not significant, and we can safely ignore it.

Figure 13 and Figure 14 show the results from passing over a pothole with a smartphone placed in the chest pocket. From this figure, we can easily find the pothole if the driver does not avoid them.

Figure 15, Figure 16, Figure 17, and Figure 18 show the results from a road with a up bump. Figure 19, Figure 20, Figure 21, and Figure 22 show the results from a road with a down bump. In all cases, we can easily find the peak of acceleration.

Therefore, when drivers pass over potholes or bumpy areas, we can easily detect the condition of the road.

6 DISCUSSION

Our experiment shows the feasibility of our proposed method. There are several issues that remain regarding this technology, some of which include the following:

(1) Drivers will probably avoid potholes and large bumps in the road. Some pertinent research into bicycle sensing exists. One study [10] shows a method for detecting and avoiding collisions. Using the same technology, we can analyze driver avoidance behavior to determine where potholes and bumps might be located. However, with this behavior we can infer only the possibility of a road surface problem.

(2) There are battery problems. Where there are many bikes, only a few smartphones should be selected to act as sensors by the servers to avoid superfluous battery costs.

The data size presents challenges. Bad road conditions are ubiquitous, and an abundance of data will be collected for transfer. Data compression technologies will be pertinent to address this issue.

7 CONCLUSION

We have successfully gathered information about the road surface conditions with bikes. The acceleration value shows the features of the road surfaces. Therefore, we can easily find locations that require the attention of maintenance crews.

However, several issues remain and require further discussion and research. In future work, we shall implement the real application on the smartphone and gather more information.

ACKNOWLEDGE

This work was supported by JSPS KAKENHI Grant Number 25330119

REFERENCES

[1] M. Maekawa, T. Fujita, A. Satou, and S. Kimura,” Usage of M2M Service Platform in ITS,” NEC Technical Journal, Vol.6. No.4, pp.43-47(2011)

[2] Onstar: <http://www.onstar.com>.

[3] G-Book: <http://www.prepaidmvno.com/capacity-carrier-db-2010-2/company-briefs/company-overview-toyota-g-book-japanese-mvno-service-by-toyota-motor-corporation/>

[4] CARWINGS: <http://www.nissanusa.com/innovations/carwings>

[5] Smartloop: <http://pioneer.jp/press-e/2007/0509-1.html>

[6] http://www.jama-english.jp/

[7] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, R. Peterson, “People-centric urban sensing,” ACM Proceedings of the 2nd annual international workshop on Wireless internet, 2006

[8] R. Kiyohara, H. kakizawa, S. Kitagami, Y. Terashima, and M. Saito, “Reducing Probe Data in Telematics Services Using Space and Time Models,” International Workshop on Informatics (IWIN 2013)

[9] H. Kakizawa, R. Kiyohara, “Reducing the Amount of Small Data Communication for Telematics Services,” The 28th IEEE International Conference on Advanced Information Networking and Applications (AINA-2014)

[10] S. Kaneda, S. Asada, A. Yamamoto, Y. Kawachi, Y. Tabata, “A Hazard Detection Method for Bicycles by using Probe Bicycle,” The 2nd IEEE International Workshop on Consumer Devices and Systems (CDS2014).

A Method for Detection of Traffic Conditions in an Oncoming Lane Using an In-Vehicle Camera

Ryo Shindo[*] , and Yoh Shiraishi[**]

[*] Graduate School of Systems Information Science, Future University Hakodate, Japan
[**] School of Systems Information Science, Future University Hakodate, Japan
{ g2113016, siraisi }@ fun.ac.jp

Abstract - In recent years, we have become able to acquire traffic information about traffic congestion through the VICS (Vehicle Information and Communication System). The VICS is one of the traffic systems that provide drivers with information on the state of traffic congestion. However, it is difficult for drivers to decide appropriately as to whether they should change lanes or make detours because the VICS provides information on the causes of traffic congestion, such as traffic accidents or road works, in the form of icons. Icons are simple representations, but are not intuitive and informative. In contrast, presenting images recorded by an in-vehicle camera to represent the causes of traffic congestion is more effective than presenting icons to help users to understand the causes intuitively. When an in-vehicle camera records the conditions directly in front of a moving vehicle, recording the traffic conditions of an oncoming lane is simpler than trying to record the conditions in the lane in which the user is driving (driving lane), as preceding vehicles may obscure the camera view. If images representing the conditions in front of preceding vehicles are sent to drivers from vehicles in the opposite lane in advance, the drivers can avoid the congestion effectively. Therefore, we propose a method for detecting the traffic conditions of an oncoming lane using in-vehicle cameras. In addition, we conducted some experiments to show the effectiveness of the proposed system.

Keywords: in-vehicle camera, detection of vehicles, traffic congestion, sensing, estimation of vehicle speed

1 INTRODUCTION

Drivers cannot effectively avoid traffic congestion through methods such as changing lanes and making detours if they are not aware of conditions of traffic congestion, such as the causes and ranges of the congestion, in advance. The VICS (Vehicle Information and Communication System) is one of the traffic systems that provide information on the conditions of traffic congestion [1]. In the VICS, traffic information such as the volume of traffic, the speed of vehicles, and so on is acquired by sensors located on roads and sent to the information center. The collected information is converted into traffic information. The center sends the traffic information to car navigation systems and other in-vehicle devices. However, the VICS provides information on the causes of traffic congestion, such as traffic accidents or road works, in the form of icons. Icons are simple representations, but are not intuitive and informative. Therefore, it is difficult for drivers to decide how to avoid traffic congestion effectively.

Currently, Probe Information Systems are in wide-spread usage [2-4]. Probe Information Systems are systems that support aspects of driving, such as navigating and calling for attention, by using information collected by sensors embedded in vehicles. Probe information includes location information, air temperature, engine rotation speed, actuating information of the ABS (Antilock Brake System), and so on. The collected probe information can be shared among vehicles through a network or directly with a wireless connection called “inter-vehicle communication” [5-7].

A driver’s front view is partially obscured by the preceding vehicles in the driving lane when the driver tries to record the causes of congestion using an in-vehicle camera. Consequently, the driver cannot grasp the causes of traffic congestion and cannot avoid traffic congestion in advance unless the driver comes close to the site of the cause. For example, in Figure 1, vehicle A’s front view is partially obscured by the preceding vehicles in the driving lane. The driver of vehicle A cannot grasp the causes of traffic congestion unless the driver comes at points of vehicle C. The cause is in front of vehicle C. On the other hand, vehicles in the oncoming lane (oncoming vehicles) can grasp the causes of traffic congestion in the driving lane. The driver of vehicle A can identify congestion in front of the preceding vehicles and avoid it if the driver gets images representing the causes of traffic congestion from oncoming vehicles in advance. In this figure, vehicle B can grasp the causes of traffic congestion in the opposite lane, and vehicle A can acquire an image representing the causes from vehicle B when vehicle B comes at the point of vehicle B*.

For these reasons, in this study we assume that vehicles can share probe information and we propose a method for detecting traffic congestion in an oncoming lane, by using an in-vehicle camera. This study aims to detect traffic congestion in an oncoming lane from the view point of vehicle B in this figure.

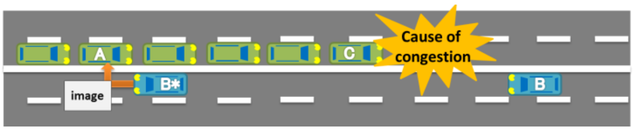


Figure 1: The positional relation between vehicles

This paper is organized as follows. Section 2 mentions research related to our study. Section 3 discusses the requirements of the proposed system. We outline our proposed method in Section 4. Finally, we discuss the effectiveness of our proposed method in Section 5.

2 RELATED WORK

This section introduces research related to our study. First, we discuss research and technologies related to presenting and sharing information on traffic conditions in Section 2.1. In addition, we discuss research on sharing information on traffic conditions by using an in-vehicle camera in Section 2.2. Finally, we discuss and compare the related research and our proposed method.

2.1 Presenting and sharing information on traffic conditions

The VICS is one of the traffic systems that provide information on the conditions of traffic congestion [1]. In the VICS, information is collected by sensors located on roads and sent to information center. The collected information is converted into traffic information, such as the range of traffic congestion, road obstacles and highway regulations. The center sends the converted information to car navigation systems and other in-vehicle devices using microwaves in the ISM band and frequency modulation (FM), similar to the Radio Data System (RDS) or Data Radio Channel (DARC). Thus the VICS can provide traffic information in real time. In the VICS, information displayed on maps of car navigation systems presents the traffic congestion classified into three degrees (sparse, crowded, and congested) based on the VICS's classification of traffic congestion (Table 1). VICS also displays icons representing highway regulations, hazard to moving traffic, and so on (Figure 2). Drivers can grasp the traffic conditions anywhere by observing the displayed information.

However, the VICS cannot necessarily collect and provide this information for every road, because some roads do not have the necessary locating devices. In addition, the VICS provides information on the causes of traffic congestion, such as traffic accidents or road works, as icons. Therefore, drivers must be able to understand the meanings of the icons. However, drivers cannot decide whether or not they will be able to avoid traffic congestion effectively because it is difficult for them to imagine the scale and the influence of the event that is happening in the driving lane from icons. Icons provided by the VICS are not intuitive information for drivers because they are simple information that does not depend on the scale of events.

Presenting camera images representing the causes of traffic congestion is effective for intuitive comprehension of traffic conditions [8, 9]. Intuitive comprehension enables drivers to identify traffic congestion in front of preceding vehicles and to avoid it in advance.

Tamai et al. [8] proposed a system that provides videos recorded at the point of traffic congestion for drivers' intuitive comprehension. A smartphone placed on the dashboard with a cradle records traffic congestion. The system collects and provides the recorded videos effectively, considering the time difference and the degree of congestion in the videos. The time difference means the difference between the time at witch a user receives the video and the time when the video was recorded. Tamai et al. [9] proposed a method that shares short videos representing the traffic conditions on roads with other vehicles. The system grasps the speed of a moving vehicle and determines the ranges of congestion based on the speed. The speed can be calculated based on location information acquired by a GPS sensor embedded in a smartphone placed on the vehicle's dashboard. At the same time, the smartphone records a front view. The system manipulates the video images considering the colors and the shapes, and detects traffic lights when the vehicle is in congestion. In addition, the system generates a video that is about 10 seconds long. The system grasps the speed of the moving vehicle easily by calculating the movement of traffic lights in the video because traffic lights are stationary objects.

Table 1: The VICS's classification of traffic congestion [1]

Degree of congestion (Color)	General road	Inner-city high-speed way	Intercity high-speed way
Congested (Red)	Less than 10km/h	Less than 20km/h	Less than 40km/
Crowded (Orange)	10km/h-20km/h	20km/h-40km/h	40km/h-60km/h
Sparse (Green)	More than 20km/h	More than 40km/h	More than 60km/h

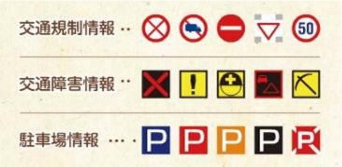


Figure 2: Icons provided by the VICS [1]

2.2 Grasp of traffic conditions by using in-vehicle cameras

We will now introduce some research on grasping the conditions of roads by using in-vehicle cameras. Kutoku et al. [10] proposed a system that detects obstacles on roads by using an in-vehicle camera. An in-vehicle camera is placed on the dashboard of a moving vehicle and records the view in front of the moving vehicle. The system generates subtracted images by using the video currently being recorded and background video. Background video is a video recorded in advance on the same road when it had no obstacles. The system detects obstacles by using subtracted images. Many researchers tackle the detection of objects on roads. However, the objects targeted by such research are assumed objects such as a person, a vehicle, and so on. Kutoku's system can detect unexpected objects by using the subtracted images. To generate subtracted images, the system must examine the time and position of the vehicles in the two videos because the speed and the positions of moving vehicles are different in each video. First, the system considers the time between the two videos using the scale representing the distance between the cameras in the two videos. Second, the system considers the positions of moving vehicles in each video by image processing of the surface of roads. According to this processing, the frames between two videos are selected and subtracted images are generated. The system calculates the recall, the false detection rate and the rate of false detection frames based on the distance between the moving vehicle and obstacles, by using image features of subtracted images. Image features include the brightness, the intensity and the edge. Then, the system detects unexpected objects considering the calculation results.

Hamao et al. [11] proposed a system that detects traffic congestion by using an in-vehicle camera. A smartphone is placed on a moving vehicle and records the view in front of the moving vehicle. The system sets a region of interest (ROI) on images, and calculates the standard deviation of the luminance histogram of the oncoming lane in the ROI. The system detects congestion based on the standard deviation of the luminance histogram between congested roads and uncongested roads.

2.3 Comparing the related works with our method

Providing information on the conditions of traffic congestion using the VICS is not intuitive for drivers because the VICS presents such information as icons. The method proposed by Tamai et al. demonstrates that presenting information on traffic congestion as camera images taken by an in-vehicle camera is effective. However, in the case where preceding vehicles are moving in front of the vehicle with an in-vehicle camera, the camera cannot capture the state of traffic congestion and its causes in the area in front of the preceding vehicles. Therefore, as in the method proposed by Hamao et al., capturing traffic congestion from an oncoming lane is easier than from a driving lane. To grasp the causes of the congestion by using an in-vehicle camera, it is necessary to detect the congestion and its range. In addition, to grasp the range and detect the congestion, it is necessary to detect the speed of oncoming vehicles. Grasping the ranges of the congestion and detecting the congestion are possible by acquiring the speed from oncoming vehicles with inter-vehicle communication. However, the moving vehicle must acquire the speed from a number of oncoming vehicles. On the other hand, detecting congestion is possible with only one moving vehicle with an in-vehicle camera. The method proposed by Kutoku et al. that detects road obstacles can detect congestion, but has difficulty detecting the speed of oncoming vehicles. The method proposed by Hamao et al. cannot detect the speed of oncoming vehicles. In addition, this method cannot discriminate between oncoming vehicles and objects behind them in images.

3 REQUIREMENTS OF THE PROPOSED SYSTEM

For intuitive grasping of the conditions of traffic congestion, presenting camera images is more effective than presenting icons. In addition, recording the conditions of oncoming lanes is easier than recording that of driving lanes when an in-vehicle camera records the view in front of a moving vehicle. Grasping the ranges of the congestion is required in order to detect the causes of the congestion. Grasping the ranges of the congestion is, namely, detecting the beginning point of the congestion and the ending point. Moreover, the speed of oncoming vehicles is required in order to grasp the ranges. From these requirements, images representing the causes of the congestion are generated. In an image, oncoming vehicles and background objects behind them must be distinguished between when image processing is applied to the image. In this study, optical flows generated between two images are calculated in order to grasp the speed of oncoming vehicles. The optical flow is a line that represents the movement of objects between two images as a vector. The length of optical flow (LOF) generated from oncoming vehicles is calculated, and the speed of oncoming vehicles is calculated. In this way, the congestion is detected. LOF depends on the distance between a moving vehicle with an in-vehicle camera and oncoming vehicles, and the relative speed between the vehicles. The distance between the moving vehicle and oncoming vehicles is smaller than the distance between the moving vehicle and the objects behind oncoming vehicles. The movement of oncoming vehicles per a unit of time is different from that of the objects behind oncoming vehicles. In this way, oncoming vehicles and objects behind them are distinguished. In addition, LOF changes depending on not only the change in the speed of a moving vehicle but also the speed of oncoming vehicles. The speed of the moving vehicle can be calculated by using location information acquired by the GPS sensor embedded in the driving recorder and the smartphone.

Therefore, the speed of oncoming vehicles can be estimated by calculating the speed of the moving vehicle and the optical flows on the images from the in-vehicle camera. In addition, traffic congestion can be detected and images representing the causes of traffic congestion can be generated.

4 PROPOSED METHOD

4.1 Summary of the proposed system

On the basis of the considerations as mentioned above, we propose a system to solve these problems. Figure 3 shows the positional relation of a moving vehicle, oncoming vehicles, and a cause of traffic congestion.

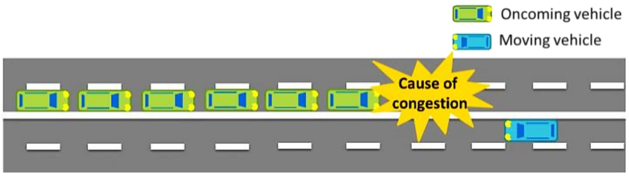


Figure 3: The positional relation between vehicles and the cause of congestion

The proposed system needs to perform the following functions.

A) Detect vehicles in an oncoming lane
B) Estimate the speed of oncoming vehicles
C) Detect traffic congestion
D) Find images representing the causes of traffic congestion
E) Estimate the range of traffic congestion

Figure 4 shows an overview of the proposed system.

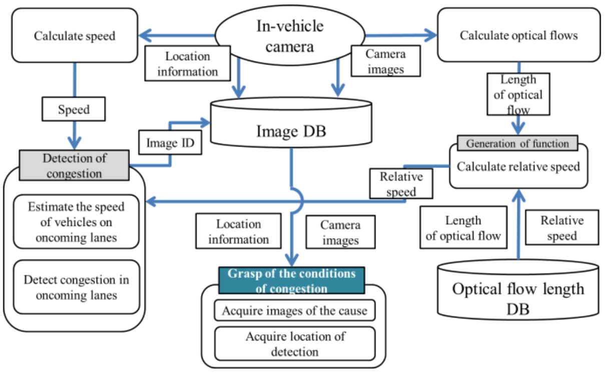


Figure 4: An overview of the proposed system

First, a driver mounts a smartphone on the dashboard and the smartphone records the front view of an oncoming lane. At the same time, the speed of the moving vehicle is acquired by a GPS sensor. Second, the system generates the optical flows between two images recorded by the smartphone. In addition, the system calculates the LOF of each relative speed and stores the dataset of LOF and relative speed in the Optical Flow Length Database. Third, the system defines an interpolation function by using the dataset in the database to calculate the relative speed from LOFs that are not in the database. Fourth, the system estimates the speed of oncoming vehicles by using the newly calculated LOF and the function. The system decides that congestion is occurring in an oncoming lane if the estimated speed falls below the specified threshold. At the same time, the system generates an image representing the cause of congestion by searching for an image recorded at the beginning of the congestion. Finally, the system generates the range of congestion by using location information from the beginning and ending point of the congestion, and presents the image and the range on a map application.

4.2 The way to calculate optical flows

In this section, we explain how to calculate optical flows of oncoming vehicles in in-vehicle camera images. There are two general ways to calculate optical flows called Phase Correlation and Block Matching Method [12, 13]. Phase Correlation is a method that calculates optical flows using a contrast equation of luminance gradient with constraint conditions. Phase Correlation can calculate optical flows, but it makes errors and is especially affected by rapid luminance changes. Block Matching Method is a method that uses a particular part of an image as a template, and calculates optical flows by exploring the parts that fit the template in the next time image. It can calculate optical flows steadily, but it is more computationally expensive than Phase Correlation. In addition, Block Matching Method depends on the size and the features of the block in an image when optical flows are calculated considering the rotation and scaling of an image. In this study, vehicles and other objects in an oncoming lane are enlarged in the image because they are recorded by a moving vehicle. Therefore, in this study, Block Matching Method is not appropriate to calculate optical flows. Our system uses the LK (Lucas-Kanade) Method that is classified into Phase Correlation and calculates optical flows by detecting feature points of an image in order to reduce the errors of rapid luminance changes (Figure 5).

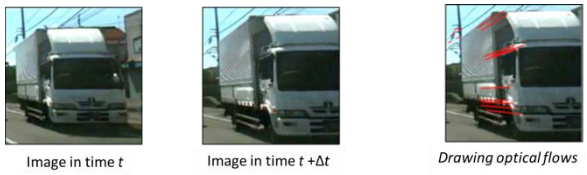


Figure 5: Optical flows drawn by LK method

4.3 Grasping the conditions of congestion

In order to grasp the conditions of congestion, the system uses two databases. One is an Image Database (Image DB) that stores recorded images and location information, and the other is an Optical Flow Length Database (Optical Flow Length DB) that stores LOF and the corresponding relative speed. Optical Flow Length DB is used to detect oncoming vehicles and to estimate the speed of oncoming vehicles. Table 2 and Table 3 show the structure of each database.

Table 2: The table structure of Image DB

Attribute Name	Detail
ID	Identification number of images
Image	Recorded image
Lat	Latitude of recording location
Lon	Longitude of recording location

Table 3: The table structure of Optical Flow Length DB

Attribute Name	Detail
R_{speed}	Relative speed between a moving vehicle and an oncoming vehicle
Len	LOF generated from oncoming vehicles in the relative speed

4.3.1. Detecting vehicles , estimating the speed of vehicles

When optical flows are generated from objects in an oncoming lane in images, the flows generated outside the zone of an oncoming lane are unnecessary. Therefore, we define a region of interest (ROI) that is placed around the oncoming vehicles in an image (Figure 6), and optical flows are generated from the objects within the ROI.

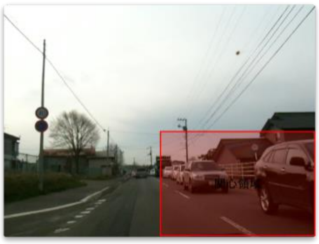


Figure 6: ROI of generating optical flows

The speed of the moving vehicle is calculated by using location information acquired by a GPS sensor of a smartphone when a driver drives the vehicle. We define the value representing the speed of the moving vehicle as M_{speed}, and the speed of oncoming vehicles as O_{speed}. Then the relative speed R_{speed} is calculated using formula (1).

$$R_{speed} = M_{speed} + O_{speed} \qquad (1)$$

As this study considers grasping the speed on general roads, the ranges of M_{speed} and O_{speed} are as follows:

$$0 \leqq M_{speed} \leqq 60 \qquad (2)$$

$$0 \leqq O_{speed} \leqq 60 \qquad (3)$$

Then the range of R_{speed} is as follows:

$$0 \leqq R_{speed} \leqq 120 \qquad (4)$$

The relative speed is acquired from the Optical Flow Length DB by querying the database with the LOF calculated from images. The system calculates the speed of oncoming vehicles by subtracting the speed of the moving vehicle from the relative speed. Then the system detects other oncoming vehicles.

However, the relative speed corresponding to the LOF specified in a query might not be stored in the database because the relative speed is continuous, not discrete. The system provides an interpolation function by using LOFs stored in the database. Then the relative speed corresponding to any LOF can be calculated by using the function. The interpolated value is returned as a relative speed when LOF that is not stored in the database is given to the function as an argument.

As we described previously, LOFs fluctuate according to the distance between an in-vehicle camera and an oncoming vehicle, the relative speed, the speed of the driving vehicle, and the speed of an oncoming vehicle. The distance between the in-vehicle camera and oncoming vehicles is shorter than that between the camera and objects behind oncoming vehicles. Therefore, the LOF generated from oncoming vehicles is longer than that generated from background objects by the vehicle's moving. Consequently, the speed of the moving vehicle is higher than interpolated relative speed. In this way the system can distinguish oncoming vehicles from background objects and detect oncoming vehicles.

4.3.2. Detecting traffic congestion

The causes of traffic congestion are detected considering estimated the speed of oncoming vehicles (SOV). The LOF is calculated, and SOV is estimated for each image when the system detects oncoming vehicles. According to the VICS's classification of traffic congestion, the speed of vehicles in a congested public highway is 10[km/h]. Therefore, the system judges the location of congestion to be a location in which SOV is continually estimated to be less than 10[km/h]. At the same time, the image of where congestion begins is a few images before that of the location where an oncoming vehicle is detected at the beginning. In addition, the system considers a location where the SOV is continually estimated to be less than 0[km/h] or more than 10[km/h] as the end of the congestion. At this time, IDs of images taken at the start and end of traffic congestion are saved. The system queries the Image DB with the saved IDs, and acquires the location information of the beginning and ending of the congestion and an image representing the cause of the congestion.

5 EXPERIMENT AND DISCUSSION

5.1 Experiment environment

We conducted the effectiveness of our proposed method, we conducted two experiments. First, we conducted an experiment for determining the statistical value of optical flows stored in the Optical Flow Length DB. In addition, we conducted an experiment for evaluating the accuracy of SOV estimation for detecting the traffic congestion in an oncoming lane by using videos recorded in the actual environment.

5.2 Experiment for evaluation

To generate optical flows and to estimate SOV, we implemented a program with OpenCV libraries. The program reads images, generates optical flows between two successive images, and draws the optical flows onto output images. To calculate optical flows, we used "cvGoodFeaturesToTrack" method which finds the most prominent corners in the image in OpenCV libraries. To calculate optical flows, we used "cvCalcOpticalFowPryLK" method which is an iterative Lucas-Kanade method using an image pyramid. A vehicle is equipped with an iPhone 3GS placed on the dashboard with a cellular phone cradle, to record video as the vehicle moves. We call this vehicle a 'recording vehicle'. The resolution of videos recorded by the iPhone 3GS is 640×480 pixels, and the frame rate of the videos is 30 [fps]. The rectangular ROI sized 430×210 pixels is placed at the bottom right of images so that oncoming vehicles fit into the ROI, and optical flows are generated within the ROI. To grasp the speed of the recording vehicle, we used the GPS sensor of an iPhone 5 and implemented an iOS application that calculates the speed of the recording vehicle by acquiring location information. To estimate the SOV, we defined three dimensions spline function as an interpolation function by using the dataset of the relative speed and LOF stored in the Optical Flow Length DB. The system gives the spline function with LOF as an argument, and acquires the relative speed. Then the system calculates SOV by subtracting the speed of the recording vehicle from that of the relative speed. In the experiment to determine parameters, four parked oncoming vehicles made congestion on a single lane. In addition, the distance between the vehicles is changed because the distance in actual traffic congestion is non-constant. The driver drove the recording vehicle and past the four parked vehicles five times at different speeds each the inter-vehicular distance, while the iPhone 3GS recorded the oncoming vehicles. Relative speed was equivalent to the speed of the recording vehicle because the oncoming vehicles were parked. We examined parameters such as the time interval between two successive images and the statistical value of LOF for generating optical flows considering the 15 recorded videos. In addition, we defined the interpolation function with parameters derived from the results of the preliminary experiment and the datasets of relative speed and LOF in the database. We considered estimation accuracy with the interpolation function.

5.2.1. Deciding parameters for generating LOF

We describe the result of the determination of parameters for generating LOF. Determined parameters are the time interval between two images in the video (*Interval*), and the statistical values of LOF stored in the Optical Flow Length DB (*Len*). Table 4 shows *Interval* and *Len* considered in this experiment.

Table 4: *Interval* and *Len* considered in this experiment

The statistical values of LOF (*Len*)	The time interval between two successive images (*Interval*)
Average (*ave*) Variance (*var*) Standard deviation (*stddev*) Median (*med*) Maximum (*max*)	2,3,4,8

The LOF generated in an image is counted, and *Len* is calculated. The values of Interval are 2, 3, 4, and 8. Interval between successive images is 1 when a video is divided into multiple images. For example, if Interval is 2, LOF is generated between the nth image and (n+2)th image. LOF is acquired when a recording vehicle passes beside the lead oncoming vehicle. Figure 7 shows the environment of the preliminary experiment.

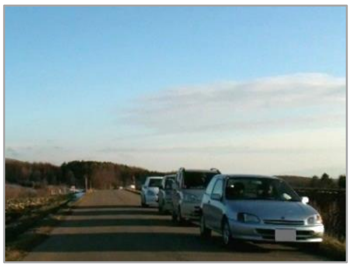


Figure 7: The environment of the preliminary experiment

Tables 5 and 6 show the changes of *Len* for each *Interval*.

Table 5: The changes of LOF in *Interval* = 2

		Len				
the distance between the vehicles [m]	Relative speed [km/h]	*ave*	*var*	*stddev*	*med*	*max*
2	9.25	6.07	0.55	0.74	6.08	7.07
	23.14	11.23	8.62	2.93	10.81	16.12
	30.7	11.87	16.31	4.03	11.04	20.09
	33.76	13.37	40.69	6.37	13.47	26.47
	40.31	12.02	48.48	6.96	8.03	29.42
4	9.9	7.44	1.56	1.24	8.03	9.21
	18.71	11.12	9.42	3.06	11.09	15.52
	28.22	14.73	19.55	4.42	15.03	23.53
	33.48	16.25	66.9	8.17	15.78	32.06
	41.5	15.86	73.66	8.58	16.03	35.17
6	11.8	8.06	3.61	1.9	8	11.4
	17.17	10.54	6.97	2.64	10.19	15.52
	24.84	11.28	26.58	5.15	10.04	19.41
	33.08	12.11	35.2	5.93	10.52	26.3
	41.47	13.3	37.48	6.12	12.16	28.28

Table 6: The changes of LOF in *Interval* = 3

		Len				
the distance between the vehicles [m]	Relative speed [km/h]	*ave*	*var*	*stddev*	*med*	*max*
2	9.25	8.18	2.07	1.43	8.15	10.19
	23.14	11.04	33.15	5.75	8.06	26.3
	30.7	14.08	84.85	9.21	9.27	38.47
	33.76	16.46	134	11.57	11	42.01
	40.31	16.38	129.98	11.4	10.19	46.69
4	9.9	8.64	6.6	2.57	8.06	12.36
	18.71	13.36	19.62	4.43	13.53	21.58
	28.22	14.05	100.48	10.02	9.25	36.62
	33.48	19.04	83.23	9.12	17.11	35.44
	41.5	18.49	164.06	12.8	12.12	49.24
6	11.8	9.3	8.96	2.99	9.13	13.6
	17.17	11.69	31.98	5.65	10.04	21.84
	24.84	13.54	50.6	7.11	13.53	32.01
	33.08	15.72	96.25	8.91	10.68	40.6
	41.47	17.63	123.25	11.1	14.57	43.46

These results show that the changes of LOF in *Interval* = 8 do not depend on the increase of the relative speed. We suppose that the movement of objects between two images is too large in *Interval* = 8, and the feature points detected in a previous image may disappear in the next image, causing extraordinary LOFs to be generated. Consequently, LOF in *Interval* = 8 is not appropriate to generate optical flows.

Figure 8 shows the changes of variance (*var*) each *Interval*.

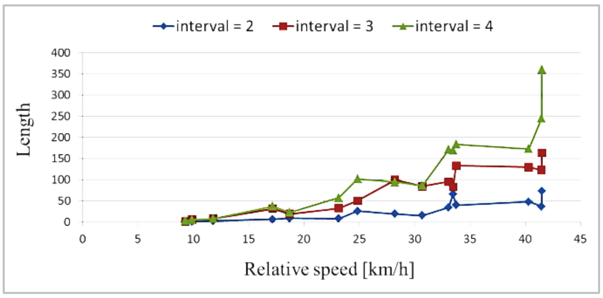


Figure 8: The changes of variance (*var*) each *Interval*

Var fluctuates widely in *Interval* = 3 and 4. In *Interval* = 2, *var* in the same relative speed are different in each the distance between oncoming vehicles. In addition, standard deviation (*stddev*) is similar to figure 8. We suppose that *var* and *stddev* are influenced greatly by the false detection of feature points. Therefore, variance and standard deviation are not appropriate to *Len*.

In *Interval* = 2, average (*ave*) and median (*med*) stand still regardless of the increase of the relative speed. In addition, maximum (*max*) increases depending on the increase of the relative speed, and the increased amount of it is small. We suppose that estimating SOV becomes susceptible to the noises of calculating optical flows if increased amount of LOF is small. In *Interval* = 3, *ave*, *med* and *max* increase depending on the increase of the relative speed, and the increased amount of *ave* and med are small. In addition, *max* increases with limited influence of the inter-vehicular distance. In *Interval* =4, *ave* and *max* increase depending on the increase of the relative speed along with figure 6, and the increased amount of *ave* is small. *Max* in the same relative speed is different in each inter-vehicular distance, and med fluctuates as the relative speed increases.

From these results we can deduce that maximum is appropriate to a statistic of LOF stored in the Optical Flow Length DB (*Len*). In addition, *Interval* = 3 is the candidate parameters for an estimation of SOV experiment. Figure 9 shows the interpolation function by using maximum in *Interval* = 3 and the relative speed.

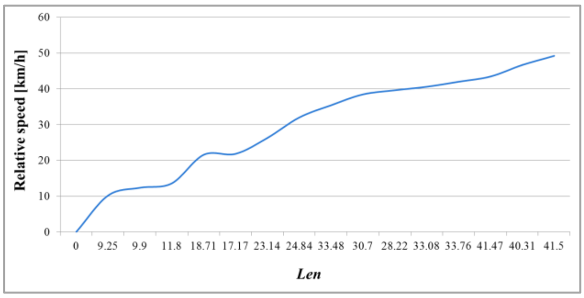


Figure 9: The interpolation function

According to Figure 9, relative speed does not necessarily increase depending on the increase of *Len*. Furthermore, the speed decreases when *Len* increases near 17.17. We suppose that *Len* is influenced by extraordinary LOF results caused by false detection of feature points, because they are generated due to the vehicle's only having driven past the oncoming vehicles once at each speed. Therefore, the errors influenced by false detection of feature points can be reduced by calculating the average of *Len* acquired by more than one drive.

5.2.2. Estimation accuracy

We evaluated the accuracy of SOV estimation by using datasets of *Len* and relative speed determined in the preliminary experiment. For this evaluation, in the same conditions as the preliminary experiment, a recording vehicle moves in the driving lane at 40 [km/h]. The recorded video is divided into multiple images. At the same time, oncoming vehicles are parked or moving slowly. We defined the interpolation function by using the datasets. LOF generated from objects in an oncoming lane in an image is given to the function as an argument. Then relative speed (R_{speed}) is calculated by the function according to the formula (1). SOV (O_{speed}) is calculated according to the formula (2). Figure 10 shows the results of estimation.

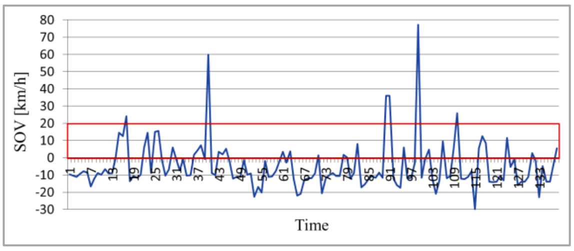
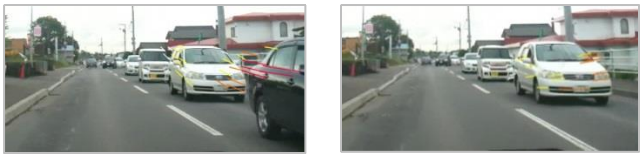


Figure 10: the results of estimation

In Figure 10, there are points at which SOV gets near to or surpasses 10 [km/h] with time. These speeds are calculated when the recording vehicle passes beside oncoming vehicles, as in Figure 11(a). At other points, SOV is less than 0 [km/h]. These speeds are calculated until the recording vehicle passes beside the next vehicle shown in figure 11(b).



(a) passing beside an oncoming vehicle
O_{speed} = 12.4
(*Len* = 54.74)

(b) passing beside the next oncoming vehicle
O_{speed} = -13.8
(*Len* = 26.47)

Figure 11: Changes in *Len* and estimated SOV in images

5.2.3. Discussion of experiment results

In this section, we discuss the results of the above experiments. In the experiment to determine parameters, we confirmed that the maximum of LOF is appropriate to generation of optical flows. In addition, we confirmed that the maximum of LOF increases depending on the increase of the relative speed in *Interval* = 2, 3, and 4. However, we confirmed that LOF decreases despite the increase of the relative speed at certain points in each *Interval*. We suppose that LOF is influenced by extraordinary LOF results caused by false detection of feature points because they are generated by only once the recording vehicle's moving in each speed. We suppose that the errors influenced by false detection of feature points can be reduced by calculating the average of LOF acquired by more than one drive at each speed. Moreover, we suppose that LOF is calculated accurately by diminishing the false detection of feature points, and the interpolation function that LOF increases depending on increasing of the relative speed can be defined. To diminish the noises of false detection of feature points, we consider the change of the size and position of the ROI and the change of parameters such as the number of detection of feature points. Then, the interpolation function can be defined accurately.

Through evaluating the accuracy of SOV estimation, we confirmed that SOV is estimated near 10 [km/h] when the recording vehicle passes beside oncoming vehicles. In addition, we confirmed that SOV less than 0 [km/h] is continually estimated until the recording vehicle begins passing beside the next oncoming vehicle. We suppose that the reason that LOF remains short is due to the following steps. First, an oncoming vehicle which the recording vehicle passes by slips from an image. Second, feature points generated on the oncoming vehicle are insufficient. Finally, feature points are generated anew on the next oncoming vehicle. We noticed certain points at which SOV is less than 0 [km/h] even though the recording vehicle is passing beside an oncoming vehicle. As we have discussed previously in the definition of parameters experiment, the system can estimate the SOV more accurately when processes to define a more accurate interpolation function are applied to the system.

6 CONCLUSION

This study aims to detect traffic congestion in an oncoming lane and to present images representing the causes of the congestion using an in-vehicle camera. We proposed a method that estimates the speed of oncoming vehicles using an in-vehicle camera to detect traffic congestion. In addition, we estimated the speed of oncoming vehicles.

In the future, to improve the accuracy of detection of traffic congestion in an oncoming lane, we will consider a method that compensates for the result of estimation of the speed of oncoming vehicles. Moreover, we will consider a method for detecting congestion on four-lane roads and divided roads. Our method is only applicable for single lane road. For example, in the case of four-lane roads, LOF generated by oncoming vehicles in each lane is different. In addition, optical flows in divided roads are generated from the median. We will define interpolation functions corresponding to multiple-lane roads and divided roads to improve our method.

REFERENCES

[1] Vehicle Information and Communication System Center, "VICS", HTML available at "http://www.vics.or.jp/index1.html".

[2] Hiroyuki Kitayama, "New Service and Platform by the Data Utilization from Cars," IPSJ Magazine, Vol.54, No.4, pp.337-343, 2013 *(in Japanese)*.

[3] Takayuki Morikawa, "Prospects of Telematics Based on Probe Vehicle Data (<Special Issue> Sophisticated Transportation Systems-Toward Transportation Services to Satisfy Individual Passengers)," Systems, Control and Information Engineers, Transactions of the Institute of Systems, Control and Information Engineers, Vol.54, No.9, pp.366-370, 2010 *(in Japanese)*.

[4] Junge Bai and Liang Zhao, "Research of Traffic State Identification Based on Probe Vehicle," Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium, pp.309-311, 2010.

[5] Bauza Ramon, Gozalvez Javier and Sanchez-Soriano Joaquin, “Road traffic congestion detection through cooperative Vehicle-to-Vehicle communications,” Local Computer Networks (LCN), 2010 IEEE 35th Conference, pp.606-612, 2010.

[6] Eiji Takimoto, Takashi Ohyama, Ryu Miura and Sadao Obana, “A Proposal and Consideration on a Management Method of Surrounding Vehicle in Vehicle-to-Vehicle Communication Systems for Safe Driving,” The Special Interest Group Technical Reports of IPSJ, ITS, Vol.2009, No.24, pp.47-51, 2009 *(in Japanese)*.

[7] Yuwei Xu, Ying Wu, Jingdong Xu and Lin Sun, “Multi-hop broadcast for data transmission of traffic congestion detection,” Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia (MUM '11), pp. 100-108, 2011.

[8] Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukuhara and Akihito Iwai, “Efficient Collection and Delivery of Video Data for Traffic Monitoring Utilizing Transition Rate of Congestion Situations,” The Special Interest Group Technical Reports of IPSJ, Vol.2012-MBL-61, No.29, pp.1-8, 2012 *(in Japanese)*.

[9] Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukuhara and Akihito Iwai, “An Image Processing-based Method for Efficient Collection and Sharing of Video Data about Conditions of Vehicular Traffic Congestion,” The Special Interest Group Technical Reports of IPSJ, Vol.2012-MBL-65, No.36, pp.1-8, 2012 *(in Japanese)*.

[10] Haruya Kutoku, Daisuke Deguchi, Tomokazu Takahashi, Yoshito Mekada, Ichiro Ide and Hiroshi Murase, “Detection of General Obstacles by Subtraction of Road-Surface with Past In-Vehicle Camera Images,” IEICE Technical Report Vol.109, No.470, pp.235-240, 2010 *(in Japanese)*.

[11] Kazuhide Hamao, Yutaka Suzuki, Masahiro Honma, Kenichi Hashimoto, Yasuhiro Ishikawa, Takashi takahi, Syuuji Ishiyama and Toshiaki Sakurai. “Methods for detection opposite lane traffic jam using a Smartphone,” Proceedings of the ITS IEICE, Vol.112, No.72, pp.19-24, 2012 *(in Japanese)*.

[12] OpenCV.jp, “Optical flow,” HTML available at “http://opencv.jp/sample/optical_flow.html”.

[13] Gary Bradski and Adrian Kaehler, Learning OpenCV Computer Vision with the OpenCV Library, California: O'Reilly Media, 2008.

Multi-region Aadaptive Geocast Enabling Two-way Communication

Hiroki Kobayashi*, Yoshitaka Nakamura**, and Osamu Takahashi**

*Graduate School of Systems Information Science, Future University Hakodate, Japan
**School of Systems Information Science, Future University Hakodate, Japan
{g2113014, y-nakamr, osamu}@fun.ac.jp

Abstract -Recently, there have been many researches on ad hoc networks that can be constructed without relying on existing communication infrastructures. Since mobile terminals perform routing functions in an autonomous distributed manner in MANETs (mobile ad hoc networks) using wireless communication, MANETs are expected to be applied to areas where it is difficult to construct communication infrastructure such as disaster areas or at the sea for example. Global positioning systems (GPSs) have been deployed in almost all mobile terminals.

If a disaster occurs, we must obtain the IP addresses of the terminals in the stricken area to communicate with the terminals. In this case, we have to use the flooding method to search for the IP addresses of the destination terminals. However, this method is inefficient because it can cause a lot of traffic in the network. Therefore, a communication method called "geocast" that corresponds with the destination terminals in such a situation is proposed.

Geocast is a casting technique using geographical information. Geocast can transmit data to all terminals existing in the area that we set by specifying the latitude and longitude. Recently, there have been many researches on routing protocols and algorithms for geocast. However, those researches focus mainly on one-way communication. This means that the source can only transmit data to the destination area by geocast. If we take into consideration two-way communication using geocast, we can transmit data from the source to the destination area and transmit the response data from the destination terminals to the source by a different communication technique. We propose a new geocast protocol by using two-way communication to solve the aforementioned problem.

Keywords**:** ad hoc networks, geocast, GGP, DSR

1 INTRODUCTION

The conventional geocast protocols [1] are basically one-way communication protocol as its purpose is to transmit a packet to an existing terminal in the specific area that we set. However, this method is not as efficient as two-way communication protocol because finding a second path is necessary for the return journey. We did not have any problems with one-way communication by advertisement delivery. However, two-way communication is necessary when we communicate with terminals in disaster areas.

When we communicate in an area where the infrastructure has collapsed, we realize two-way communication using geocast. In particular, we can use geocast in stricken areas to communicate with victims by, for example, issuing evacuation advisories. In addition, it is necessary to have many transmission areas to allow geocast to transmit a message to plural areas in actual use.

If we transmit a packet to plural areas using geocast, the network traffic becomes large when we use flooding, thus rendering the transmission ineffective. Therefore, a geometry-driven geocasting protocol (GGP) algorithm is proposed [2].

The GGP can transmit a packet to plural areas using geocast.

In this paper, we propose a multi-region adaptive geocast protocol considering for two-way communication based on the GGP and source routing protocol. And at the same time, we will show how to obtain the performance evaluation results by simulation.

2 RELATED WORKS

2.1 Geocast

Geocast is a communication scheme proposed by Julio. C. Navas in 1997. The expected acquisition of the GPS location tool facilitated the future of Navas as this tool was devised as a geocast protocol for position dependency. Geocast has the ability to transmit data to all terminals existing in an area that we set by specifying the latitude and longitude by utilizing geographical information obtained through casting techniques.

2.2 Reliability of Geocast

Recent researches on geocast have been studied to demonstrate how proposals for routing protocols and algorithms efficiency are being conducted and implemented. In addition, the application of geocast to such communication and disaster information has also been proposed. However, these researches mainly focus on one-way communication rather than taking in-to account two-way communication. These researches consider one-way communication. This means that because these researches consider one-way communication only, the source terminals cannot determine whether or not the packet was received at the destination terminal. In contrast, there is reseach that has reliable geocast communication by returning an ACK to the source terminal when the destination terminal area receives a packet [3]. However, in this research, communications will end once the ACK is returned, thus proving that the method is unable to perform continued communications thereafter. In reality, if we wish to communicate with terminals in stricken areas, reviving and using two-way communication will be necessary. Furthermore, the transmission destination is not limited to only one area as we are considering transmission via geocast to multiple areas.

2.3 ROUTING ALGORITHM OF GEOCAST

(1) Routing scheme of geocast

There are three types of routing scheme that are typical of geocast: flooding, no flooding, and directed flooding. In particular, directed flooding is a method of communication that uses position information. It is therefore often utilized in geocast researches. We have decided to use this method as a base protocol of our proposal.

(2) Overview of directed flooding method

In the directed flooding method, each terminal is in possession of the position information, and the method using a forwarding zone demonstrates how it controls the communication with position information. Generally, there are three types of algorithms: static zone scheme [4], adaptive zone scheme [5], and adaptive distance scheme (ADS) [5]. We have to select the scheme in the actual use of this method.

When source terminal transmits a packet to several areas using geocast, the network traffic becomes large especially if the terminal uses flooding. This basically makes the system ineffective. Therefore, a geometry-driven geocasting protocol (GGP) algorithm is proposed as the GGP can transmit a packet to several areas using geocast. This routing algorithm is Greedy forwarding [6]. The communication model is shown in Figure 1.

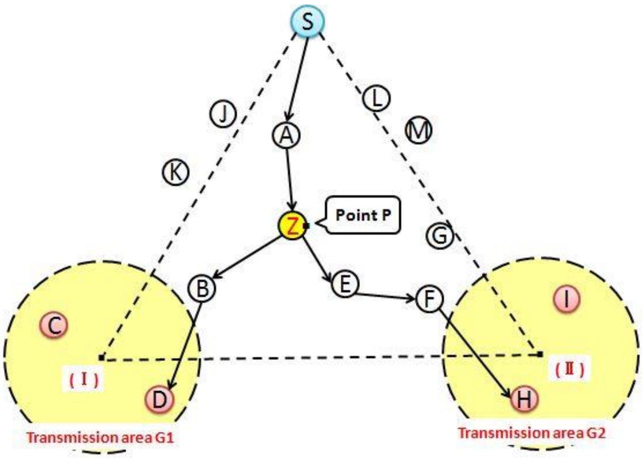


Fig. 1 GGP

In Figure 1, I (G1) and II (G2) are the point of the destination areas. Source terminal (S) is connected to two other transmission area points I (G1) and II (G2) thus creating a triangular center of gravity with middle point P. When the terminal (A) receives the packet, the terminal (A) sends packet to the point P. Source terminal (S) then transmits a packet towards P. When the terminal of point P receives a packet, the terminal transmits the other packet to two transmission areas. As a result, the communication path will become two from the terminal (Z). Through this way, the communication until the terminal (Z) is one path. The communication until the terminal in the destination areas is two paths. As a result, we can create an efficient network.

2.4 DSR Protocol

The DSR (dynamic source routing) protocol consists of two functionalities: route maintenance and route discovery [7] [8]. In addition, each terminal pathway that leads to any terminal is recognized and stored in the route cache. This route cache is updated each time a route to any terminal is discovered. In the beginning, route discovery automatically checks whether a route to the destination exists. If the route exists, we then use it. If the route does not exist, the terminal flooding route request packet (RREQ) captures its own IP address. The terminal that receives the RREQ checks whether the route to the destination exists in the route cache by itself. If the path exists in the route cache, the destination terminal notifies the sender of the route via routing reply (RREP). If the path does not exist in the route cache, the terminal transmits to the destination in the RREQ to achieve flooding and to capture its own IP address in the packet. If the RREQ reaches the destination, the destination terminal then transmits an RREP to the source terminal using a path that is described in the packet. DSR protocol is for performing construction on the pathway. The basic DSR protocol is shown in Figure 2.

When the source terminal captures its own position information in the packet header, the two-way communication geocast can be recognized by the IP address that is captured at the same time.

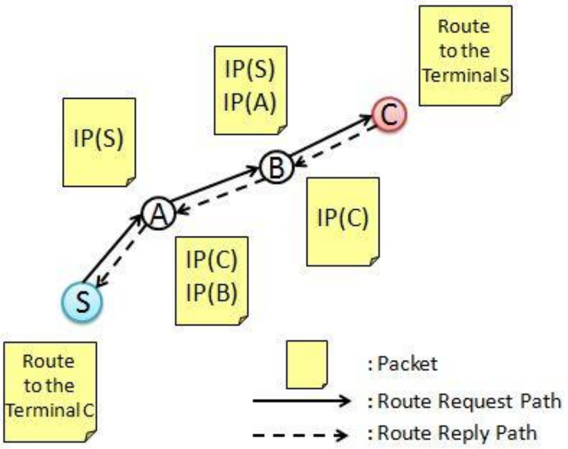


Fig. 2 DSR

3 PROPOSAL OF MULTI-REGION ADAPTIVE GEOCAST

(1) Basic concept

Conventional geocast communications use a different routing protocol in the forward path and the backward path. However, this method requires route searching again. Therefore, it is very inefficient because a delay occurs. To solve this problem, we propose an addition to the GGP algorithm to create the geocast DSR that fuses the DSR and source routing geocast.

(2) Selection of directed flooding scheme

We have evaluated and compared the number of packets that are sent over the network and the complexity of each terminal [9]. The terminal that receives the packet only determines the sending terminal itself that exists in the transmission area so the amount of calculation in the terminal is usually small. However, this approach always communicates using the flooded method. Thus, the number of packets in the network increases. The terminal that receives the packet determines the sending terminal itself existing in the transmission area. Then, we rebuild the forwarding zone, which involves calculating the one hop flooding. Therefore, the amount of calculation is higher than that of the static zone scheme. In addition, the number of packets in the network will be less due to some rebuilding of the forwarding zone. The terminal receiving the packet determines whether approaching the center of the transmission area is necessary to calculate the distance to the center. In addition, we examine the terminal just before forwarding to determine whether a destination area exists. This technique has high computational complexity because it requires determining its own override position information before forwarding. However, the number of packets in the network is reduced by transfer control of the distance between the center coordinates. There is not much difference in the number of calculation algorithms used by the three routing schemes as significant performance improvement of the mobile terminal was achieved. Therefore, in the proposed method, we used the ADS, which seems to operate with the lowest number of packets in the network.

(3)Two-way communication in real environment

Performing two-way communication using geocast when the communication infrastructure has collapsed is considered. It is possible to communicate with the victims using geocast when there is an evacuation and substantial damage to buildings. In addition, we performed geocast communication in a number of areas.

The purpose of the conventional geocast protocol is to transmit a packet to a terminal that is in the destination area. Accordingly, it is a one-way communication protocol. If we realize two-way communication using geocast, we transmit data from the source to the destination area and transmit the response data from the destination terminal to the source by a different communication technique. This method requires route searching for the return again. Accordingly, two-way communication with this approach is inefficient. We did not have any problems with one-way communication by advertisement delivery, but two-way communication is necessary when we communicate with terminals in stricken areas. We proposed a two-way communication geocast technique using source routing and GGP. However, the packet would be concentrated in the Fermat point terminal. In this method, we have not obtained a result that is higher than expected. In addition, when there is more than one destination area, packet loss due to redundant paths occurs.

In this paper, we propose a multi-region adaptive geocast protocol for two-way communication based on the GGP and source routing protocol.

4 IMPLEMENTATION OF TWO-WAY COMMUNICATION

4.1 One-to-one communication

If the destination area and the source terminal are in one-to-one communication, we create the route using the DSR and ADS.

There is a pattern for the three communication phases: first forward path, first return path, and round-trip path after the first path.

Phase 1: First forward path

The basic first forward path is shown in Figure 3.

First, the source terminal S decides the transmission area. Second, the source terminal S transmits a route request (RReq) packet to the transmission area. Third, terminals A and E, which have received the (RReq) packet, decide whether to forward it. This is the ADS algorithm. All terminals which received the packet perform this determination. In this case, terminal A is closer to the transmission area than source terminal S. Therefore, terminals A and E encapsulate the IP address and location information in the packet. Then, terminals A and E transmit the RReq packet to the nearest terminal. When terminal F receives an RReq packet, it determines whether to send an RReq packet it. However, terminal F is one more step away from the transmission area than terminal E. Therefore, terminal F discards the RReq packet. When terminal C (which is within the transmission destination area) receives the RReq packet, terminals C and D will receive the route information to source terminal S.

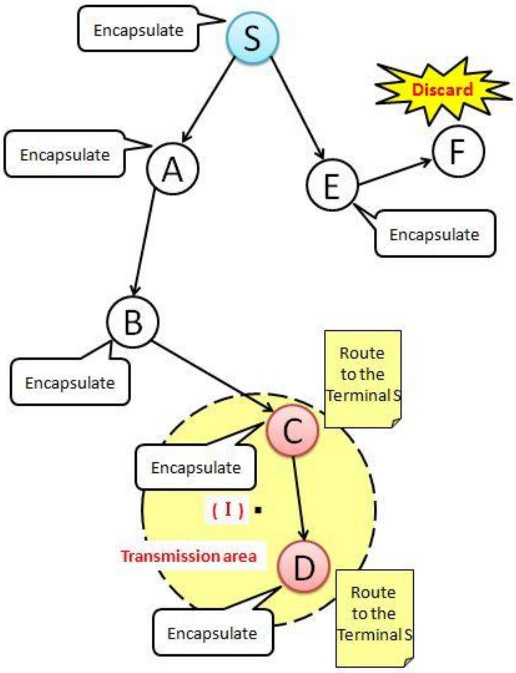


Fig. 3 First forward path

Phase 2: First return path

The basic first return path is shown in Figure 4.

Terminal D (which has received the RReq packet) knows route to source terminal S. Therefore, terminal D sends the route reply (RRep) packet to source terminal S. Terminal S (which has received the RRep packet) has the route to terminal D. Therefore, the destination terminal may not have to search for a route. As a result, the proposed method can suppress the delay.

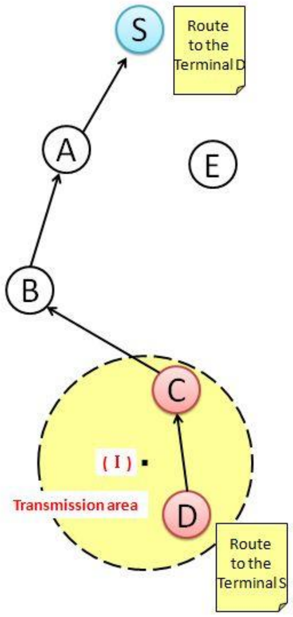


Fig. 4 First return path

Phase 3: Round-trip path after the first return path

The basic round-trip path after the first return path is shown in Figure 5.

When the first forward communication and the first backward communication are completed, the creation of the route by DSR is completed. Therefore, terminals D and S have a route to each other. Two-way communication with this route is possible.

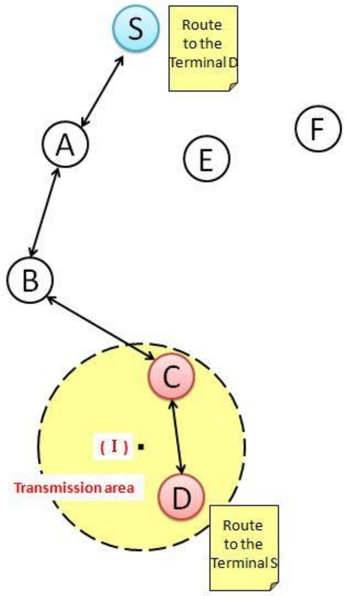


Fig. 5 Round-trip path after first return path

4.2 Basic one-to-n (n≧2) communication

The basic one-to-n communication is shown in Figure 6. First, source terminal S decides two destination areas. Second, terminal S decides the Fermat point P. At this time, we find out the Fermat point to determine the radius of a circle for any point P. And the terminal in this range is the Fermat point terminal. Third, we draw a tangent from terminal S towards the Fermat circle. We called this area the forwarding zone. Terminal S sends the RReq packet encapsulating the information towards the Fermat circle. When terminal A receives the RReq packet, it sends the RReq packet encapsulating its own information to terminal Z. Fermat point Terminal Z sends the RReq packet to the destination area of the two specified RReq packets that encapsulate its own information. At this time, the RReq packet that terminal G received is discarded by the ADS. Each terminal encapsulates its own information so that a destination terminal area knows the route to Terminal S. A destination terminal transmits the RRep packet to the source terminal using this pathway. There is only one route for each terminal in Figure 6, and the terminal flooded the forwarding zone. Thus, various routes are constructed in practice. We communicated by selecting the most efficient route. However, we considered the situation when a route cannot be used due to any failure. At that time, the terminal rebuilt the route by DSR and continued to communicate.

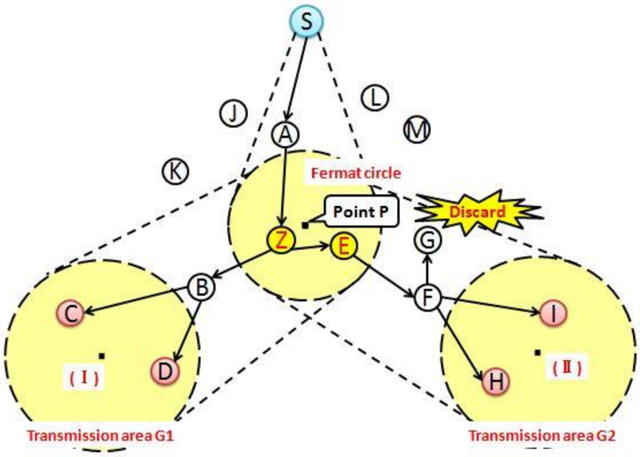


Fig. 6 One-to-n communication

4.3 Advanced one-to-n(n≧2) communication

The advanced one-to-n(n≧2) communication is shown in Figure 7.

First, the source terminal decides a destination area and divides it into three separate areas for every 120 degrees of the network area. Second, the source terminal does this to all the destination areas and chooses the most efficient way of dividing them. In Figure 7, the source terminal wants to set areas 1 and 2 and we communicate 1 to 2 in the network area. The source terminal communicates 1 to 1 in area 3.

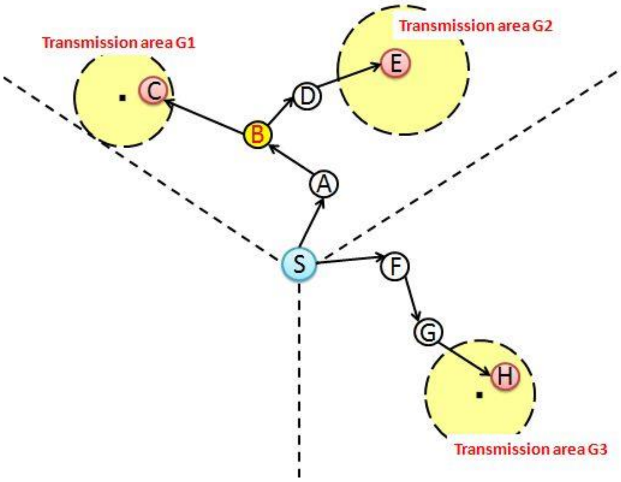


Fig. 7 Multi-directional transmission algorithm

The basic communication without the use of multi-directional transmission algorithm is shown in Figure 8.

The source terminal transmits one at a time toward the three transmission areas. Therefore, the path will become longer. As a result, congestion and packet loss occurs and the packet arrival rate will be lower.

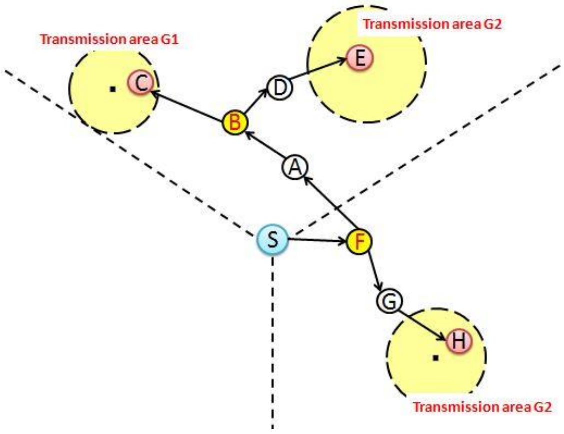


Fig. 8 Communication without the use multi-directional transmission algorithm

5 EVALUATION

5.1 Simulation

In this paper, we aim to send packets to more than one area by using geocast DSR. Also, we aim to further improve efficiency by adding the GGP function.

In this paper, we evaluate the proposed method by using the Qualnet network simulator.

We have compared the proposed method with the traditional geocast method.

We examined the number of RReq packets sent over the network and verified the effectiveness of the utilization efficiency of the network. The basic simulation model is shown in table 1.

Table 1 Simulation environment

Simulator	Qualnet 5.2
Area range	1000m×1000m
Coverage area	300m (max)
Total number of terminal	30, 60, 100
Number of the source terminal	1
Moving speed of the terminal	2.0m/s
Number of the destination terminal	5
Number of the destination area	3
Radius of geocast	100m
Radius of fermat circle	100m
The packet transmission interval	1packet/sec
Size of data packet	512kbyte

We have measured the packet arrival rate in the above environment. The packet arrival rate is the probability that the packet reaches the destination terminal from the source terminal. The situation set-up of the network is the same as that in Figure 7.

5.2 Results and discussion

The experimental results are shown in Figure 9 and 10.

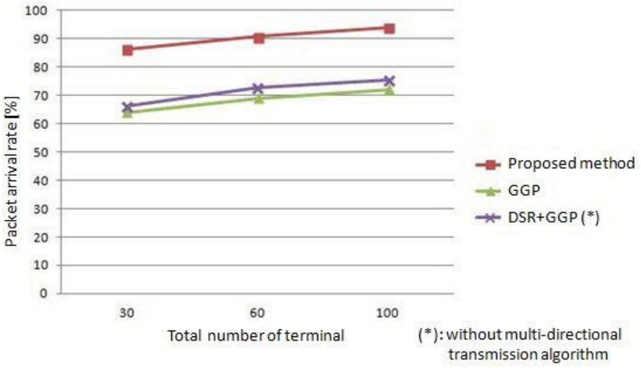


Fig. 9 Packet arrival rate

The packet arrival rate increases as the number of terminals increases in all methods. However, DSR+GGP and GGP send packets towards the multi-directional area. Therefore, the packet arrival rate is lower. The proposed method has a separate transmission to each network for each area. Therefore, the proposed method can construct an efficient route. In addition, the proposed method can perform route packet reassembly after multiple packet collisions caused by multiple algorithms at the Fermat point terminal. Therefore, the proposed method maintains a high packet delivery ratio.

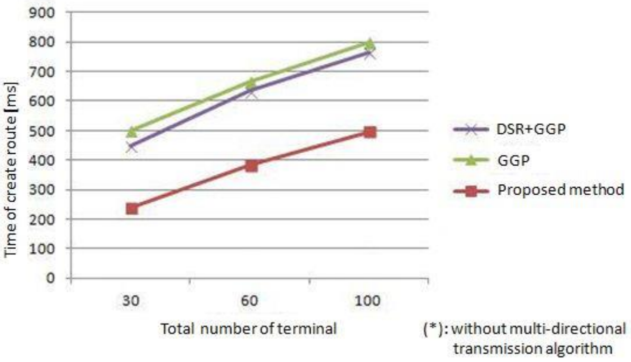


Fig. 10 Time of create the route

Then, we measured the amount of time to create the route. The proposed method uses a multi-directional transmission algorithm. This method has two paths, which increases the time to create the route. However, DSR+GGP and GGP does not use multi-directional transmission algorithm. This means that these methods have one path thus elongating the path of these methods causing a decrease in time of creating the route.

6 CONCLUSION

In this paper, we evaluated a geocast routing algorithm. In addition, we selected an algorithm with a good network utilization efficiency.

We examined the problems associated with the traditional geocast. These problems occur when two-way communication is difficult. And we propose a multi-region adaptive geocast protocol while considering the two-way communication. In addition, we had many Fermat point terminals. As a result, we determined the reconstruction path when the route failure occurred. We conducted experiments to compare the conventional method and the proposed method and discussed these experiments on the basis of the results. We found that the proposed method is superior to geocast DSR. Therefore, the proposed method can control avoiding wasteful packets when performing two-way communication using the geocast in the affected areas. Furthermore, we can realize flexible two-way communication with the terminal in the affected areas.

We now have some problems to tackle for the future. The first problem is a network area determination algorithm. The proposed method selects the network area. We must then be able to perform this process automatically. The second problem is a path reconstruction algorithm. When reconstructing the route, the proposed method uses another route stored in the route cache. In this case, we must consider the algorithm for selecting the optimum route. The third problem is when there are obstacles on the network. The proposed method does not simulate while simultaneously considering obstacles. When using the proposed method in the disaster area, there are obstacles in the disaster area. Therefore, we must propose a routing algorithm that considers the obstacles.

REFERENCES

[1] Julio C.Navas and Tomaz Imielinski , “Geocast-geographic addressing and routing”, Proc. MobiCom’97, pp.66-76,1997.

[2] Shinsuke TERADA , Takumi MIYOSHI , Kaoru SEZAKI, “Multiregion-adaptive Geocast Routing on Ad Hoc Networks , Technical Report of IEICE, Vol.107, No.524(NS2007-185), pp.299-302, 2008. (in Japanese)

[3] Kosuke YAMAZAKI, Kaoru SEZAKI, “A proposal of a reliable geocast protocol”, Technical Report of IEICE, Vol.101, No.715(NS2001-228), pp.111-118, 2002. (in Japanese)

[4] Young Bae Ko and Nitin H. Vaidya, ”Flooding-Based Geocasting Protocols for Mobile Ad Hoc Networks”, Mobile Networks and Applications, Vol.7, No.6, pp.471-480, 2002.

[5] Rajendra V. Boppana , “An Adaptive Distance Vector Routing Algorithm for Mobile, Ad Hoc Networks”, Proc. INFOCOM2001, Vol.3, pp.1753-1762, 2001

[6] Brad.Karp and H.T.Kung, “GPSR: Greedy Perimeter Stateless Routing for Wireless Networks,” Proc. MobiCom2000, pp. 243-254, 2000.

[7] David B.Johnson, David A.Maltz and Josh Broch, “DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks”, Ad Hoc Networking, ed.Charles E.Perkins, pp.139-172, Addison-Wesley, 2001.

[8] “DSR （ Dynamic Source Routing ） protocol” http://internet.watch.impress.co.jp/www/column/wp2p/wp2p05.htm

[9] Hiroki KOBAYASHI , Yoshitaka NAKAMURA , Osamu TAKAHASHI , “ Proposal of two-way communication geocast using the source routing” DICOMO2013, pp.109-115, 2013. (in Japanese)

Session 2:
Databases and Systems
(Chair : Takuya Yoshihiro)

Implementation and Evaluations of Recovery Method for Batch Update

Tsukasa Kudo[†], Masahiko Ishino*, Kenji Saotome**, and Nobuhiro Kataoka***

[†]Faculty of Comprehensive Informatics, Shizuoka Institute of Science and Technology, Japan
* Faculty of Information and Communications, Bunkyo University, Japan
** Hosei Business School of Innovation Management, Japan
*** Interprise Laboratory, Japan
kudo@cs.sist.ac.jp

Abstract - For the operation of the nonstop service systems, some methods have been put to practical use to perform the batch update concurrently with the online entries. However, the whole batch update can't be executed as a transaction by the conventional methods. So, in the case where a transaction failure occurs in the batch update, there is the problem that the rollback of the update can't be executed with maintaining the isolation from the online entries. For this problem, we have proposed the temporal update method, by which the batch update can be executed as a transaction. In this study, we show the consistency of the batch update result can be checked before the commit by this method, even in the case of the concurrent execution with the online entries. Further, we show the following: the recovery of the transaction failure by this method can be executed without affecting to the online entries; it is more efficiently than the conventional methods.

Keywords: database, batch processing, transaction, recovery, business system, nonstop service.

1 INTRODUCTION

In the actual business systems, databases are indispensable, and it is generally updated by two methods. First is the lump sum update by the great deal of data (here in after, “batch update”), that is the batch processing[4]. It is used widely in various fields: the settlement of account in the accounting systems, the great deal of account transfer for entrust company in the banking systems, and so on. Here, since its target data is a great many, the impact of the failure affects the extensive range of business. So, various kinds of mechanisms are introduced to maintain the safety of the system operations. For example, the temporary process is executed prior to the definite process. In the former, various kinds of confirmations are executed beforehand: validity of the input data, the consistency of the processing results and so on. After this confirmations, the definite process is executed to update the business data of the database.

On the other hand, users enter their data from the many online terminals concurrently (hereinafter, “online entry”): in the accounting systems, the sales information is entered from the POS (Point of Sales) terminals; in the banking systems, the deposit and withdrawal data is entered from the ATMs (Automatic Teller Machines). As for the online entry, small amount of data is entered at each time, and it is reflected into the database immediately. And, their concurrent execution is controlled by the transaction feature of the DBMS (Database Management System). Here, in the present time, non-stop service systems are used widely: such as the internet shops, ATMs, and so on. So, both of the above-mentioned two methods have to update the same table of the database concurrently.

As for the online entry, since it is executed as a transaction, its consistency properties is maintained. Moreover, the concurrency control is performed by the DBMS using the lock feature to execute the many transactions as a serializable schedule. So, transactions are executed without affecting each other based on the isolation. And, in the case of transaction failure, its rollback is executed to cancel the entry without affecting the other entries based on the atomicity and isolation. On the other hand, if the lock feature is used by the batch update for great deals of data, then the conflicting online entries are made to wait for a long while. That is, though the batch update can executed as a transaction, it causes the problem of the long latencies of the online entries.

For this problem, the mini-batch is used widely to shorten the latency time. It splits the batch update to the short time transactions, and executes them sequentially[4]. However, the atomicity and isolation of the ACID properties can't be maintained as for the whole mini-batch. That is, since the updated data is committed one after another, the rollback of the whole target data can't be executed even in the case of the failure. That is, the update must be canceled by the restore of the backup of the pre-update data or by the compensating transactions. However, in the case where it is executed concurrently with the online entries, they use the updated data immediately. So, it is difficult to cancel their result, and these methods aren't practical. As a result, there is the problem that the mini-batch must complete all the update by removing the cause of the failure.

This means that the above-mentioned safety of system operations can't be maintained. In particular, in the actual system operations, there are faults due to not only the program error but also the data quality, operation error and so on. Therefore, there is the problem that the complex or atypical batch update process must be often executed by stopping the online entries to separate the both updates. For this problem, we have proposed the temporal update method, which utilizes the data history of the time series, and shown that the batch update can be executed as a transaction without long latency of the online entries by this method[8]. However, we haven't evaluated this method in the case of failure yet. Our goal in this paper is to evaluate the recovery function of the temporal update for the transaction failure assuming the actual business

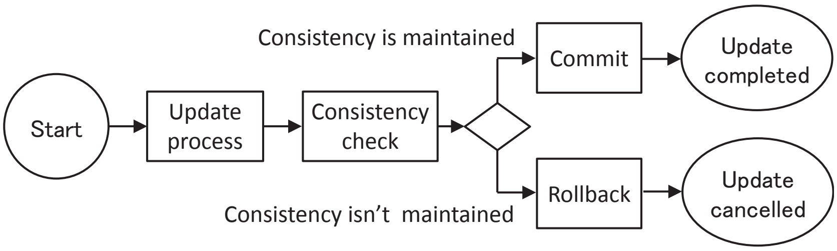


Figure 1: State transition of update transaction.

system operations. First, we show the requirements for the recovery of the batch update failure due to the actual business system operations. Here, we assume the batch updates are executed concurrently with the online entry in these system operations. Next, based on these requirements, we perform the experiment about this recovery feature by implementing the temporal update prototype.

The remainder of this paper is organized as follows. We show the batch update model and related works in Section 2, and show the requirement about recovery functions of the temporal update in Section 3. We show the implementation and evaluations of these functions in Section 4, discuss on the evaluation results in Section 5, and conclude in Section 6.

2 RELATED WORKS

2.1 Batch update operations and methods

To maintain the consistency of the database, various kinds of functions are implemented in the actual business systems. As for the update process of the online entry mentioned in Section 1, it is executed as transactions, and the consistency of the update result is checked before its commit. And, in the case where the consistency is not maintained, the rollback is executed to cancel the update as shown in Figure 1. For example, in the banking systems, if the updated result of the account transfer becomes minus, it can't be executed. So, the rollback of the transaction is executed, and the process is canceled. On the other hand, in the case where the consistency is maintained, the commit is executed.

As for the batch update, in the case of being executed in the different time period from the online entries such as the night batch, the operations like this can be performed. We show the dataflow of the batch update process in Figure 2. Since the batch update processes a great deal of data in a lump, commits are usually executed on the way based on the resource constraint. So, the backup of the target table is performed prior to the update. Then, in the case where the consistency of the update result isn't maintained, the target table is recovered by the backup data. That is, the update is canceled. On the other hand, in the case where the consistency of the result is maintained, the update completes and the target table is used by the business such as online entries.

Thus, even to execute the batch updates safely, the ACID properties of the transaction must be maintained. That is, as for atomicity, the state of database has to transit to a state of either: the commit state of the update in the case of successful

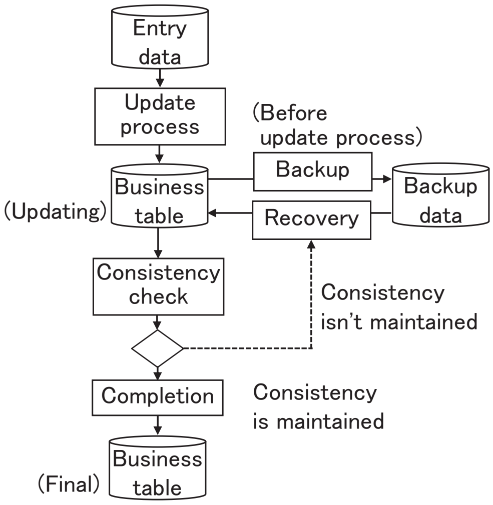


Figure 2: Data flow of batch update process.

completion; the state of the update canceled by the rollback in the case of failure. As for the consistency, the update result must be checked to satisfy the various constraints before its completion, by which the other transactions are permitted to access the table. As for the isolation, the target table of the update mustn't be accessed until the completion to avoid the influences on the other processing. In addition, the durability is maintained by the function of the DBMS as well as the online entry transaction.

This process is implemented by the method to lock the target table or data during the batch update. Though the ACID properties of the batch update can be maintained by this method, there is a problem that the online entries to access the target data are waited for a long while. However, as for the non-stop service systems, since the online entries are always performed, it is impossible to separate the processing time zone between the batch update and them. As a result, there is a problem that it is difficult to execute the batch updates with maintaining the ACID properties. So, as for the concurrency control between the batch update and online entries, some methods have been put into practical use [12], [13]. However, as for the conventional methods, there are some problems to apply it to the non-stop service systems.

First, the mini-batch splits the batch update to the short time transactions and executes them sequentially to shorten the time to lock each data. However, as for this method, there is the problem that the ACID properties can't be maintained

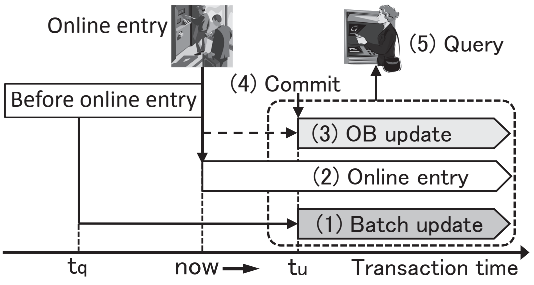


Figure 3: Outline of temporal update method.

as the whole mini-batch update as mentioned in Section 1. Next, as for the timestamp ordering, a unique timestamp is assigned to each transaction[2], [3]. And, if the transaction accesses the data updated by the larger timestamp transaction, it is aborted. However, since the batch update takes a long time, it has to be abort in most cases. Moreover, as for the multiversion concurrency control based on the snapshot isolation level, it also uses the above-mentioned method [1], [9]. So, there is the same problem.

2.2 Temporal update method

To solve this problem, we have proposed the temporal update method [8]. In this method, we use the extended transaction time. Here, the conventional transaction time expresses when a fact existed in the database [10]. For example, if a fact existed between the time T_a and T_d, its transaction time period T is expressed by $T = \{T_a, T_d\}$. That is, the fact is entered into the database at time T_a, and deleted at time T_d logically to remain the data history. As long as the data hasn't been deleted yet, the instance of T_d is expressed by now, which is the current time to query the database [11]. And, we extend this time to the future.

In Figure 3, We show the outline of this method. In this method, the batch update queries the data of the past time t_q, and inserts the update result in the future time t_u as shown by (1) in this figure. On the other hand, the online entry accesses the data of the current time now as shown by (2). So, it doesn't conflict with the batch update. Here, even for the online entry results during the batch update, it is also necessary to perform the batch update. So, it is executed by the online batch update (hereinafter, "OB update") as shown by (3), which updates the online entry results similarly to the batch update separately. Then, by the commit of the batch update as shown by (4), the batch and OB update results become to be queried by the other processing. However, since plural data exists after the commit, valid data have to be queried as shown by (5). The target data is selected as follows: firstly, the latest update data is selected for each key, and it can be identified by the above-mentioned time attribute T_a; secondly, if there is still plural data, the target data is determined in the order of the OB update, online entry and batch update. In other words, the online entry has higher priority to the batch update. And, the OB update, which reflects the batch update on the online entry result, has the highest priority.

We have shown that we can execute the batch update more efficiently by this method than by the mini-batch, in both of the centralized and distributed database environment [5], [6]. Moreover, we have shown that we can apply this method even if the completion time can't be predicted at the batch update starting [7]. However, these previous studies assume only the case where the temporal update completed successfully. And, the following studies have never performed: the study on confirmation of the consistency of the temporal update results before the commit; the study on rollback of the transaction in the case where some anomaly is detected as shown in Figure 1.

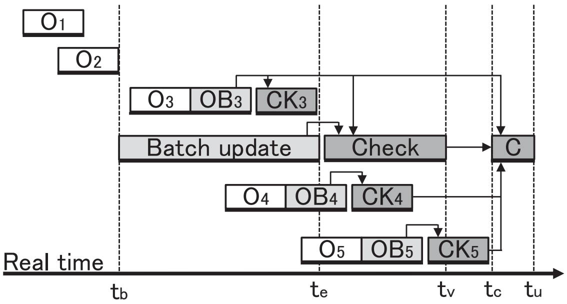


Figure 4: Execution timing of transactions.

3 REQUIREMENT FOR CONSISTENCY MAINTENANCE

In this section, we explore the requirements to maintain the consistency of the result of the temporal update, in the case where it is executed concurrently with the online entries. We show an execution timing example of transactions related to the temporal update in Figure 4. In this figure, O_i shows the online entry; OB_i shows the OB update; C shows the commit of the batch and OB update. OB_i is executed following to O_i during the batch update. Since the batch update processes a great deal of data, it is often composed by plural transactions. And, at the batch update completion time t_e, from the viewpoint of the transaction feature of DBMS, the whole batch update and the completed OB update results mustn't be queried by the other transactions until the completion of the commit C. In other words, as for the example in this figure, the transaction of the batch update and OB_3 are prepared to commit as the temporal update. We call this status "pre-commit". And, with the passage of time, the other OB updates complete one after another such as OB_4, OB_5, and they are committed in a lump with the batch update at the time t_c.

Here, the consistency of the batch and OB update results have to be checked before the commit as shown in Figure 2. And, if their consistencies aren't maintained, the rollback of these updates must be done. However, since the OB updates are performed until the commit C, their results have to be also checked individually by CK_i as shown in this figure. And, if the consistency of OB_i isn't maintained, both of its rollback and the corresponding online entry O_i's rollback are done.

Assuming the above, the requirement to maintain the ACID properties in the temporal update is as following. First, as for the consistency, it has to be checked prior to the commit as shown in Figure 4. Here, the consistency has to be checked

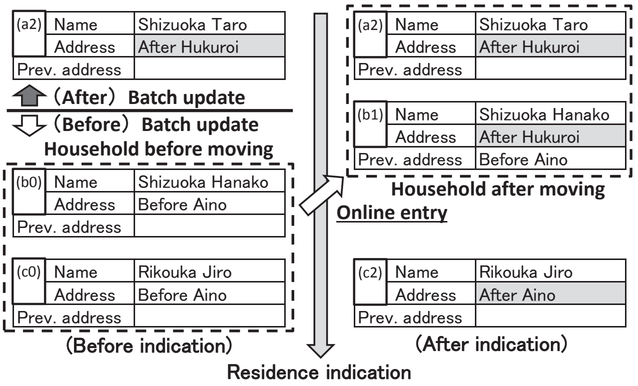


Figure 5: Update example of mini-batch.

by both of the DBMS functions and the business logics. The latter indicates that these update results have to be queried from only the batch update application program prior to the commit. In other words, prior to the commit of the temporal update, its results and the committed online entries have to be queried to verify the consistency of the database.

On the other hand, as for the isolation, the batch update and OB update results mustn't be queried by the online entry application program until the commit. In addition, even if rollback due to the failure is not completed until the commit time t_u that was scheduled, it mustn't be queried similarly. And, as for atomicity, both rollbacks of the batch update and the OB update have to be executed in the case of failure. That is, the database transitions to the state where only the online entries were performed. Moreover, in the actual system operations, if the batch update was aborted due to the failure, it is necessary to be rerun by removing the cause of the failure immediately. So, the rollback also has to be executed efficiently.

In addition, since the OB update accompanies the online entry, it isn't executed if the latter is failed. Conversely, when a failure is detected in the OB update, its rollback needs to be executed by one of the following two methods. First, in the case where a transaction failure or consistency anomaly is detected during the OB update, both rollbacks of it and the corresponding online entry must be executed to prevent the anomaly between them. That is, the online entry and OB update have to be executed as a single transaction. Second, in the case where some consistency anomaly is detected by the check of the batch update, the commit of only the online entry can be executed. So, the batch update and all the corresponding OB updates must be canceled by the rollback.

For this example, we show the data manipulation of the residence indication. In this business, all the address is changed to be easy to understand, and it is performed in the whole target district at the same. Here, the data of one household must have the same address, and a resident after moving has the previous address, which must be consistent with the address as timing. So, since the data are related with each other, it is necessary to process the whole batch updates as a transaction. Therefore, it can't be executed by the mini-batch concurrently with the online entries. We show this example in Figure 5, and "Prev. Address" shows the previous address in it. In this case, household address is changed by the mini-batch update. Concurrently, by the online entry, a member of a household (b0) that hasn't changed yet moves to another household that has already changed, and its result is shown by (b1) in this figure. Here, the resident card has the previous address, and it doesn't reflect the residence indication. However, the (current) address reflects it, though it is same timing with the previous address. Therefore, an anomaly as for the address occurs.

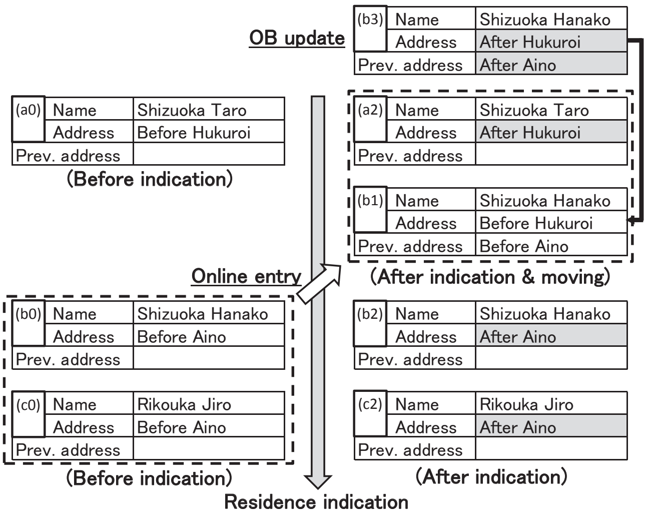


Figure 6: Update example of temporal update.

On the other hand, we show the case of executing the same processing by the temporal update method in figure 6. As for this method, its batch update results aren't queried until the commit. So, as shown by (b1) in this figure, both of the address and the previous address of the online entry result are based on the data before the residence indication. Similarly, the following data is generated: the batch update result (b2), which doesn't reflect the online entry; the OB update result (b3), which reflects the residence indication on the online entry result. As shown in Section 2.2, since the OB update result (b3) has the highest priority, the moving result reflecting the residence indication is queried.

And, in the case where the transaction failure occurs in this processing, the following recovery is executed. First, if the OB update fails, then the address and previous address shown by (b1) remains. So, it causes anomaly among (b1) and (a2). Therefore, both of the OB update and the online entry must be canceled by the rollback. As a result, only the residence indication result (b2) remains, which maintains the consistency. Second, in the case where the online entry fails, the result is same as this. Third, if the batch update fails by some reason, the batch update and OB update are canceled. They are (a2), (b2), (c2) and (b2). So, the online entry result (b1) remains instead of the original data (b0). That is, only the moving result remains.

As shown above, the requirement to maintain the consistency of the temporal update result is the following. First, OB update has to be executed in the single transaction with the corresponding online entry. Second, prior to the batch and OB update commit, the consistency of their result have to be checked. Third, if the anomaly is detected, the rollbacks

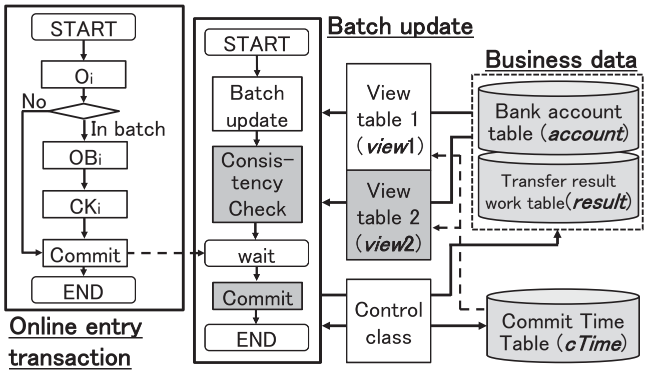


Figure 7: Program composition of temporal update.

of the batch and OB update have to be executed without affecting to the online entries. Fourth, this rollback has to be executed efficiently.

4 IMPLEMENTATION AND EVALUATIONS

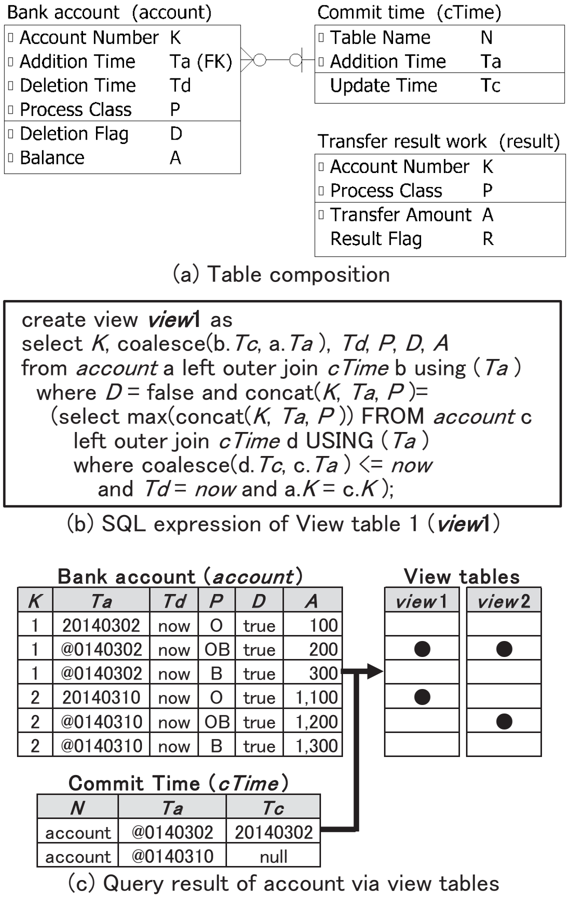
We perform the experiment by the prototype of the bank account system, because it is a simple business system. And, verify that the requirements mentioned in Section 3 can be realized by the temporal update method.

4.1 Implementation of temporal update

In Figure 7, we show the program composition of the temporal update to realize the data manipulation shown in Figure 4. As for the online entry transaction, in the case where the online entry O_i conflicts the batch update, the OB update OB_i and its consistency check CK_i are executed. In addition, in the case where some anomaly is detected, the rollback of this transaction is executed. On the other hand, as for the batch update, its status becomes pre-commit shown in Section 3, when its execution and the commit of DBMS has completed. Next, its consistency check is executed. Then, it waits the completion of the running online entry transactions that update the target data of the batch update, and executes the commit.

As for the table of database, we use the following three tables. Bank account table (hereinafter, "$account$") stores the deposit balance of the account, and Commit time table (hereinafter, "$cTime$") stores the data to manage the temporal update. In addition, Transfer result work table (hereinafter, "$result$") stores the transfer amount and the result of account transfer, which is performed by the batch and OB update.

Here, the data of $account$ is queried via following two view tables as shown in 7, and $cTime$ is used for these view tables. First, View table 1 (hereinafter, "$view1$") is used by the online entry transactions, and the data of the batch update and OB update data isn't be queried until the commit of the batch update. Of course, the transaction can query the OB update data made by itself. Second, View table 2 (hereinafter "$view2$") is used by the batch update, and has the same function as $view1$. Also, the batch update program can query the batch and OB update data as for it before the commit.



Bank account (account)

□	Account Number	K
□	Addition Time	Ta (FK)
□	Deletion Time	Td
□	Process Class	P
□	Deletion Flag	D
□	Balance	A

Commit time (cTime)

□	Table Name	N
□	Addition Time	Ta
	Update Time	Tc

Transfer result work (result)

□	Account Number	K
□	Process Class	P
□	Transfer Amount	A
	Result Flag	R

(a) Table composition

```
create view view1 as
select K, coalesce(b.Tc, a.Ta ), Td, P, D, A
from account a left outer join cTime b using (Ta )
  where D = false and concat(K, Ta, P )=
    (select max(concat(K, Ta, P )) FROM account c
       left outer join cTime d USING (Ta )
       where coalesce(d.Tc, c.Ta ) <= now
         and Td = now and a.K = c.K );
```

(b) SQL expression of View table 1 (*view1*)

Bank account (*account*)

K	*Ta*	*Td*	*P*	*D*	*A*
1	20140302	now	O	true	100
1	@0140302	now	OB	true	200
1	@0140302	now	B	true	300
2	20140310	now	O	true	1,100
2	@0140310	now	OB	true	1,200
2	@0140310	now	B	true	1,300

View tables

view1	*view2*
●	●
●	
	●

Commit Time (*cTime*)

N	*Ta*	*Tc*
account	@0140302	20140302
account	@0140310	null

(c) Query result of account via view tables

Figure 8: Data and view tables of temporal update.

In (a) of Figure 8, we show the table composition and its query result via view tables. Firstly, as for $account$, its relation scheme R_a and its attributes are as follows. In addition, T_a and T_d are the transaction time attributes mentioned in Section 2.2, and they are shown in the form of date for the simplicity.

$$R_a(K, T_a, T_d, P, D, A) \tag{1}$$

- K**:** primary key attributes. It is the primary key of the projection (K, A), which is the attributes for business.

- T_a**:** addition time of the data. Here, as for the batch and OB update, it is the start time of the temporal update and, "@" is set for the first place as shown at (c) of Figure 8. It is for the case where the completion time of the batch update can't be predicted.

- T_d**:** deletion time of the data.

- P**:** process class. This shows the process that updated this data: the OB update, the online entry, and the batch update. The corresponding value set is expressed by $\{\mathrm{P_{ob}}, \mathrm{P_o}, \mathrm{P_b}\}$. Here, we make $\mathrm{P_{ob}} > \mathrm{P_o} > \mathrm{P_b}$.

- D**:** deletion flag. This shows whether the queried data is the target of the query. So, it has the logical value $\{true, false\}$. And, if $D = false$ then the data doesn't be queried. It is used by the OB update to hide the corresponding batch update result, which data was deleted by the online entry.

- A**:** the other attribute. As for the Figure 8, it shows the balance of the bank account.

Next, as for $cTime$, its relation schema R_c and attributes are as follows.

$$R_c(N, T_a, T_c) \tag{2}$$

- N**:** table name updated by the target batch update.
- T_a**:** addition time of the target batch update and OB update data.
- T_c**:** commit time of the target batch update. Until the completion of the commit C, it is set to "null".

Here, $cTime$ controls the query results of view tables. As for $view1$, by the commit of the batch update, the time of this commit is set to T_c. So, if $T_c \neq null$, then the valid data is queried as follows. We show its SQL expression in (b) of Figure 8. First, as for the data $P = \mathrm{P_b}$ (batch update result) or $P = \mathrm{P_{ob}}$ (OB update result) in $account$ and $cTime$, if $account.T_a = cTime.T_a$ then $account.T_a$ is replaces by corresponding the value of $cTime.T_c$. Next, the data is queried for each K, which has the latest T_a then the largest P. Incidentally, "now" is the current time mentioned in Section 2.2.

Similarly, we define $view2$. As for it, the data is queried for each K, which has the latest T_a then the largest P. Here, since "@" is larger value than the number, the batch and OB update result has the latest T_a. For example, if the online entry is executed concurrently with the batch update, OB update result is queried.

We show an example of the query result via view table in (c) of Figure 8. As for the data $K = 1$, since its batch update is committed, the query results are same in the both of two views. As for uncommitted data $K = 2$, the online entry result being executed concurrently with the batch update is queried by $view1$; its OB update result is queried by $view2$.

In addition, $result$ is a table about the interface with the company which entrusted the account transfer: the account number, transfer amount of each account is indicated by the company; the result flag is set by the system to each account data corresponding to the successful or failure, about the account transfer. Since this table is used as a work file, the data history isn't necessary. However, the result flags are set by both of the batch and OB update processes. Here, the execution order of them is undecided. So, we added the process class P to this table to store both of the results, and select the data in the same way as $account$.

4.2 Experimentations and evaluations

We performed this experiment by the stand alone PC: CPU is Xeon CPU E5-1620 3.60 GHz with 8 GB memory, and its OS is Windows 7 (64 bit); DBMS is MySQL Ver. 5.6.17; the transaction control is performed by its database engine InnoDB; the concurrent execution is controlled by Thread class of Java;

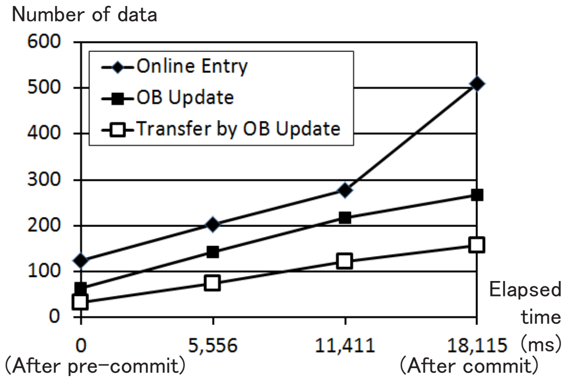


Figure 9: Transition of number of data via View table 2.

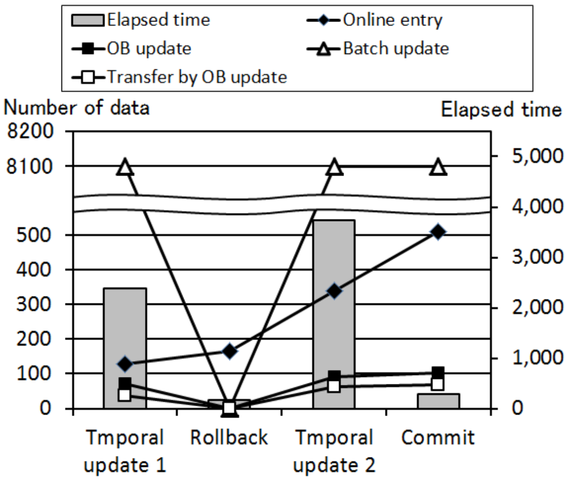


Figure 10: Transition of number of data about rollback.

4.2.1 Consistency check before commit

To evaluate the requirement mentioned in Section 3, we performed experiments using the tables shown in Figure 8. Prior to each experiment, we set 10 thousand data, which bank balance is one million, to the $account$; set 9 thousand data of transfer amount A to $result$ to perform the bank transfer, here result flag R (transfer result) is set to "null". In the latter, transfer amounts A is calculated in accordance with the account number K as

$$A = [111 - (K\ mod\ 110) + 1] \times 10^4$$

to examine the account transfer about both of the success and failure cases. Concurrently, the online entries are executed from 5 terminals, by which 50 % of the account balance of each bank account is transferred to its corresponding bank account. Each is executed in the interval of about 0.3 sec. And, if the online entry conflicts with the batch update, then its OB update is executed.

To examine whether the consistency of the batch and OB update can be checked before the commit, we experimented to query the data via $view2$. In Figure 9, we show the experimental results from the pre-commit until the commit completion in the temporal update. In this figure, "Online En-

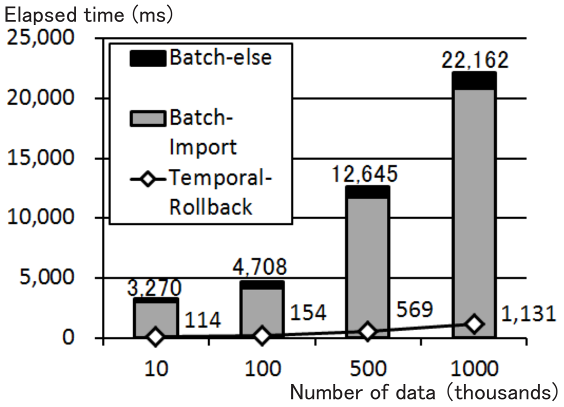


Figure 11: Comparison of elapsed time for recovery.

try" shows the change of the number of the data in *account* updated by the online entries with the elapsed time; "Transfer by OB Update" by the OB update. Similarly, "OB Update" shows about the OB update in *result*. Here, if the bank transfer is failed due to the lack of the account balance, then *account* is not updated. So, the numbers of the latter two is different. In addition, though not shown in this figure, the unchanging number of the data updated by the batch update could be queried, because the batch update process had already completed.

As shown in this Figure, we could query the data of each query time, including the pre-committed batch and OB update data. Thus, using $view2$, we could check the consistency of the database any time before their commit.

Similarly, we examined the case of the rollback, and in this case, the experimental process to manipulate the data is the follows. In addition, the above-mentioned online entries are executed over this experiment, and the data of target tables is queried after the each stage. Firstly, the temporal update and its pre-commit are executed; secondly, the rollback of the batch and OB update is executed; thirdly, the temporal update is rerun for the data of this time, and its pre-commit is executed; finally, the commit of the temporal update is executed.

We show the experimental result in Figure 10 similar to Figure 9. Here the elapsed time of the each stage is shown by the bar graph and right axis, instead of the total elapsed time from the beginning. Figure 10 shows, in addition to the result shown in Figure 9, the number of the batch and OB update result becomes to 0 after the rollback. That is, the status, in which the batch and OB update were canceled at this stage, could be queried. On the other hand, as for the requirement about the affecting to the online entries, since this rollback didn't affect to the online entries, the number of the online entry data was constantly increased.

4.2.2 Efficiency of rollback

To evaluate the efficiency of the rollback in the temporal update, we compared its elapsed time with the time for the recovery in the conventional batch and mini-batch update.

Here, as for the temporal update, its rollback can be executed by deleting the data from both of the business table and Commit time table $cTime$: as for the former, the target data set X of $account$ is expressed as following.

$$X = \{x \in R_a | (x[P] = \mathrm{P_b} \cup x[P] = \mathrm{P_{ob}}) \cap T_a = \text{@time}\}$$

Here, $x[P]$ shows the value of attribute P in x; similarly, "@time" shows the value of T_a of this temporal update, which is "@0140310" in Figure 8. Similar to this, as for $cTime$, the target data set Y is expressed as follows.

$$Y = \{y \in R_c | y[N] = \text{account} \cap T_a = \text{@time}\}$$

On the other hand, the recovery for the batch and mini-batch update has to be executed by the following process as shown in Figure 2: firstly, the backup of the target data is always executed before the update; next, in the case of failure, the data of the target table is cleared, and the target table is restored by the backup.

In Figure 11, we show the experimental result for the above-mentioned recovery in the four cases of the number of data in $account$. Here, "Batch-else" shows the total elapsed time of the backup and clear, and in this experiment we executed the clear by the "truncate" statement of SQL; "Batch-import" shows the elapsed time for the restore by the backup data, which is executed by the "import" statement of MySQL.

As shown in this figure, as for the both methods, as the data increases, the elapsed time is longer. However, the rollback of the temporal update could be executed so efficiently as compared with the method by the backup and restore. For example, in the case of the maximum number (1000 thousands), the elapsed time of the former was about 1/20 of the latter. Furthermore, most of the elapsed time of the recovery by the backup was spent for the restore.

5 DISCUSSIONS

Through the implementation and experiments, we confirmed that the requirements mentioned in Section 3 can be meet by the temporal update. First, as shown in Figure 7, both of the online entry and OB update could be composed as a single transaction. Second, as shown in Figure 9, the batch and OB update results could be queried before their commit. That is, the consistency about them could be checked before the commit. Third, as shown in Figure 10, the rollback of them could be executed without effect on the online entry result. That is, if anomaly is detected by the above-mentioned check, the update is canceled by the rollback. Finally, as shown in Figure 11, this rollback is very efficient comparing with the conventional recovery method: about 20 times in the experimental case. Moreover, as shown in Figure 10, we performed the verification of the case of rerun. In the actual system operations, if the anomaly is detected, the cause has to be removed and the job has to be rerun to complete the business. We think the result of this verification shows that this method is useful for the actual system operations.

Here, for the reason of the efficiency of the rollback, it can be pointed out that the rollback of the temporal update can be performed by a simple delete command as shown in Subsection 4.2.2. That is, in the temporal update, the data histories about the transaction time are managed. So, as shown in Figure 8, since the batch and OB update results are stored in the table as unrelated records to the online entry results, they are classified by only the attribute of Process class P. As a result, the target data of the rollback can be deleted efficiently.

In the actual system operations, various kinds of failures are detected in the batch update and often it has to be rerun. So, we consider this efficient rollback is useful to shorten the turnaround of the batch update. Moreover, as shown in Figure 10, we performed the verification of the case of rerun. In the actual system operations, if the anomaly is detected, the cause has to be removed and the job has to be rerun to complete the business.

Also, for example, even in the case where the batch update can be executed concurrently by using the mini-batch, the online entries are often stopped for the safety. As a reason for this, it can be pointed out that above-mentioned various failures, such as manipulation errors, dead-locks and so on, become threats of system operations, which may cause the error of the online entry result and to disturb the online entry. From the viewpoint of the safety of the batch update operations, the updated results should be separated from the online entry results; the consistency of the updated results can be checked before their commit. And, these requirements can be realized by the temporal update. Therefore, We think that this method is useful for the actual system operations.

As for the implementation of the temporal update, some attributes have to be added to the target and related tables: such as the transaction time, process class, and deletion flag. Also, some functions have to be composed: the view tables to query these tables; the OB update for the online transaction. On the other hand, the temporal update can be composed without considering the online entry. That is, it isn't necessary to commit in each short time period, nor to implement the recovery methods such as the compensating transactions. So, it can be composed easier than the mini-batch.

Lastly, we would like to discuss about the advantages and disadvantages of this method comparing with the conventional method. As for the advantages, firstly, this method can execute the whole batch update as a transaction concurrently with the online entries. That is, in the conventional method, there are the following problems about their concurrent executions: the batch update with table lock can be executed as a transaction, but the online entries must wait for its completion; the mini-batch can be executed concurrently with the online entries, but it isn't a transaction as the whole processing. So, the latter can't do its rollback and can't maintain the isolation. And, the data has to be recovered by the backup data in the case of failure as for the both. Therefore, this method is useful for the following batch update: the batch update on the data related with each other, as shown in Figure 5; the batch update which recovery for the failure has to be performed in a short time. On the other hand, as for the disadvantage, above-mentioned functions must be implemented for this method as shown in Figure 7. So, the application of this method should be decided based on this trade off.

6 CONCLUSIONS

At the present time, since most of the business systems provide the nonstop services, the batch update has to be executed concurrently with the online entries. However, in such the environment, since it can't be executed by the conventional methods as a transaction, some problems remain. So, we have proposed the temporal update method, and shown it can be executed as a transaction. However, to apply this method to the actual business systems, it has to equip the functions for the failures.

In this paper, we analyzed the batch update operations in the actual business systems, and showed the requirement to execute the temporal update safely during the online entries. Moreover, through the experiment by the prototype, we confirmed that the temporal update satisfy these requirements. Especially, we find that the rollback can be executed so efficiently comparing with the conventional methods. Therefore, we can extract conclusions that this method is useful in the actual system operations.

Future studies will focus on the investigation of the application fields of this method, and its application evaluations.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 24500132.

REFERENCES

[1] Berenson, H., et al., A Critique of ANSI SQL Isolation Levels, Proc. ACM SIGMOD 95, p. 1-10 (1995).

[2] Bernstein, P. A., Hadzilacos, V., Goodman, N., Concurrency Control and Recovery in Database Systems, Addison-Wesley (1987).

[3] Connlly, T. M., Begg, C. E., Database Systems: A Practical Approach to Design, Implementation and Management, Addison-Wesley (2009).

[4] Gray, J., Reuter, A., Transaction Processing: Concept and Techniques, San Francisco: Morgan Kaufmann (1992).

[5] Kudo, T., Takeda, Y., Ishino, M., Saotome, K. and Kataoka, N., Evaluation of Lump-sum Update Methods for Nonstop Service System, Int. J. of Informatics Society, Vol. 5, No. 1, pp. 21–28 (2013).

[6] Kudo, T., Takeda, Y., Ishino, M., Saotome, K., Kataoka, N., A Mass Data Update Method in Distributed Systems, 17th Int. Conf. in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, Procedia Computer Science, Vol. 22, pp. 502–511(2013).

[7] Kudo, T., Takeda, Y., Ishino, M., Saotome, K. and Kataoka, N., Application of a Lump-sum Update Method to Distributed Database, Proc. of Int. Workshop on Informatics (IWIN2013), pp. 49–56 (2013).

[8] Kudo, T., Takeda, Y., Ishino, M., Saotome, K., Kataoka, N., A batch Update Method of Database for Mass Data during Online Entry, Procs. 16th Int. Conf. on

Knowledge-Based and Intelligent Information & Engineering Systems – KES 2012, pp. 1807–1816 (2012).

[9] Silberschatz, A., Korth, H. F., Sudarshan, S., Database System Concepts, McGraw-Hill Education (2010).

[10] Snodgrass, R., Ahn, I., Temporal Databases, IEEE COMPUTER, Vol. 19, No. 9, pp. 35–42 (1986).

[11] Stantic, B., Thornton, J., Sattar, A., A Novel Approach to Model NOW in Temporal Databases, Procs. 10th Int. Symposium on Temporal Representation and Reasoning and Fourth Int. Conf. on Temporal Logic, pp. 174–180 (2003).

[12] Wang, T., Vonk, J., Kratz, B., Grefen, P., A survey on the history of transaction management: from flat to grid transactions, Distributed and Parallel Databases, Vol. 23, Issue 3, pp. 235–270 (2008).

[13] Yadav, D.S., Agrawal, R., Chauhan, D.S., Saraswat, R.C., Majumdar, A.K., Modelling long duration transactions with time constraints in active database, Procs. the Int. Conf. on Information Technology: Coding and Computing (ITOC’ 04), Vol. 1, pp. 497–501 (2004).

A Simulator for the Execution Efficiency Measurement of Distributed Multi-database Virtualization

Daichi Kano* Hiroyuki Sato* Jun Sawamoto* and Yuji Wada**

* Graduate School of Software and information Science, Iwate Prefectural University, Japan
** Department of Information Environment, Tokyo Denki University, Japan
sawamoto@iwate-pu.ac.jp

Abstract –In database virtualization technology, the database of a different kind can be used as if it were a kind of database. However decline of execution efficiency is left as one of the research subjects. In improving the execution efficiency, it is necessary to measure the execution performance of the virtualization processes, especially in a distributed environment where multiple databases are connected via a network. In this study, we have designed and implemented the simulator for the execution efficiency measurement. This simulator measures the execution efficiency by calculating the processing time of virtualization processes, database processes and communication processes, and totaling them.

Keywords: Distributed database, Multi-database virtualization, Simulator, Performance evaluation and improvement.

1 INTRODUCTION

Today, it is important to discover and analyze the knowledge and trends which are hidden in large collections of data on ubiquitous network environment using data mining technology, and to use them for decision making of business, etc. However, since those data exists in various types of distributed databases, an appropriate database has to be chosen from a variety of databases and accessed properly. The work of the preparation process of data mining of acquiring appropriate data is needed, and it becomes a burden for the data analysis engineer who performs data mining in the distributed database environment.

To reduce this burden, the multi-database virtualization technology which enables a user to access various types of databases as if accessing a single type of databases has been studied [1-3]. The usefulness has been shown when database virtualization technology is used to perform data mining.

However, some research issues are pointed out. Degradation of the execution efficiency by virtualization processing among the research issues remain by the previous work as one of the main subjects to be solved. Since virtualization processing is performed in addition to normal database processing, it causes execution degradation. Virtualization processing transforms commands and the processing result based on the schema, and when especially the processing result becomes extensively large, virtualization processing becomes a burden. An improvement can be expected by using load sharing technology and parallel processing technology for this issue. While each load decreases by distributing data processing and parallel processing, we anticipate the generation of network delay by low line speed, congestion, etc. Therefore, factors about the network, such as communication time and transmission speed, become important as well as processing of databases.

In improving the execution efficiency, it is necessary to measure the execution performance of the virtualization processing, especially in a distributed environment where multiple databases are connected via a network. But it takes a lot of databases and large-scale network structure, and preparation of actual measurement environment is costly and very difficult. Therefore, the measurement environment using a simulator is considered.

In this study, we have designed and implemented the simulator for the execution efficiency measurement. This simulator measures the execution efficiency by considering the processing time of virtualization processes, database processes and communication comprehensively. And we aim to contribute to quantitative verification and evaluation of the execution efficiency improvement technique of virtualization processing.

The rest of this paper is organized as follows: In section 2, we describe related works. In section 3, we present our proposed solution for database virtualization. In section 4, details of the design of the proposed simulator are described. In section 5, we report the process and some results of acquiring reference parameters for the time of virtualization processing. Finally, the paper is concluded in section 6.

2 RELATED WORKS

The performance estimation of database system is an active research area. They mainly approach this subject by building performance models of database servers and running the models for the simulation [1-3].

Garcia [1] presents a simple model based on the queuing network paradigm using fixed distribution for the service times of the queues. The parameters used in the model are adjusted using measurements taken from real servers. This work demonstrates that extreme simple model is capable of predicting the performance of metrics of real database servers with high accuracy and capturing the essential performance aspects of database servers.

Wentao, et al [2, 3] propose a method for predicting query execution time for concurrent and dynamic database workloads. Their approach is based on analytic model rather than machine-learning model. They use optimizer's cost model to estimate the I/O and CPU operations for each individual query, and then use a queuing model to combined these estimates for concurrent queries to predict their execution times. A buffer pool model is also used account for the cache effect of the buffer pool.

These related works are all targeted for real database servers. On the other hand, our target is virtualized distributed multi-database system. And we have designed and implemented the simulator for the execution efficiency measurement by considering the processing time of virtualization processes, database processes and communication processes. Our main goal is to discover the bottlenecks of the database virtualization processing.

Some earlier reports in [4], [5] and [6] have described the study of database virtualization technology.

Mori et al. [4] proposed development of a system to disseminate information actively to all users in a mobile computing environment. They implemented an experimental system using the meta-level active multi-database system as the platform in a mobile computing environment. By mapping the data of the local database group to a meta-database through the basic search and build operations, the system intends to combine data and include different types of local database group.

The data integration technique, Teiid [5], enables virtualization of various types of databases; through such virtual databases, one can access such data sources as relational databases, web databases, and application software such as ERP and CRM, etc. in real time. They can all be integrated for use. In fact, Teiid has a unique query engine. Furthermore, the real-time data integration is accomplished by connecting business application software through the JDBC/SOAP access layer with data sources which are accessed through the connector framework.

In [6], they similarly describes a module known as a wrapper that allows accessing and integrating data from various sources such as RDBs, the Web, and Excel files.

In our previous study [7], we considered the metadata, UML, ER model, and the XML schema as candidates for use to accomplish database virtualization. Thereby, ubiquitous databases can be used as if they were a single database. We then compared the advantages and disadvantages of each to analyse them as follows.

In our previous studies [7-10], we examined XML schema advantages and proposed a virtualization method by which such ubiquitous databases as relational databases, object-oriented databases, and XML databases are usable, as if they all behaved as a single database.

3 DATABASE VIRTUALIZATION [7]

Databases of many kinds exist in terms of their associated data model differences and vendor differences. Regarding differences among data models, each has different data representation, and unique associated manipulation. Some typical examples include the table type of relational databases (RDB), XML-representation type of XML databases (XMLDB), and object-oriented databases (OODB). Even the same model database might have different features among vendors. Regarding RDB for example, there might be some differences in SQL and/or data type representation. The typical example is that we have MySQL, PostgreSQL, and SQLServer from different vendors.

These differences according to the model and vendor bring some undesired results. For example, we might end up spending more time and labor during application system development because of the different data models that must be confronted. For example, we might need to acquire the right API to handle data of every different type of database. Virtualization of such different types of modelled databases to unify the procedures for all of them would probably impart less of workload and cost, and facilitate their management in a more flexible manner. Consequently, virtualization of databases, if it could be done, would facilitate application system design and database management as well.

To have a virtualization feature, we will consider the inclusion of features to manage distributed databases of similar types, the distributed databases of different types, and provide location transparency for users, such that they notice no differences of database structure or location and become able to use databases of all kinds in a flexible fashion. Figure 1 portrays an example view of the database virtualization technique.

For virtualization of ubiquitous databases in our study, we will describe the schema information of the real databases, of which more than one always happens to exist, by creating and using one common XML schema. We also provide functionality of data search and update with the XML-based common data manipulation API.

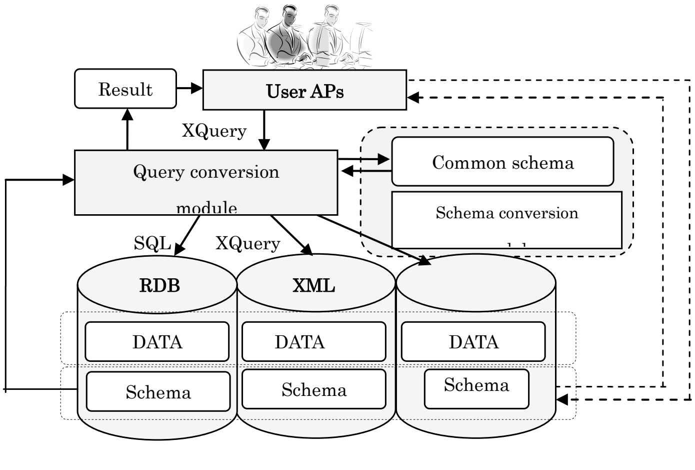


Figure 1: An example view of database virtualization.

3.1. XML conversion program

We will use an XML schema that provides a flexible representation capability and a high transparency capability.

To do so, we will produce such a virtualization concept in which the user would feel as if he or she were locally manipulating the remote site RDB from a local RDB process environment. That can be accomplished by converting the schema information and data information of the local RDB into the XML schema, and then storing that information into the RDB that the user would like to operate.

We developed an XML conversion program, XML Export/Import, as depicted in Figure 2. We then used such different vendor RDBs as MySQL, PostgreSQL, and SQLServer2005 because they are available in the RDB virtualization system creation environment. We have to rebuild the XML tree with our XML conversion program when the distributed database is redefined.

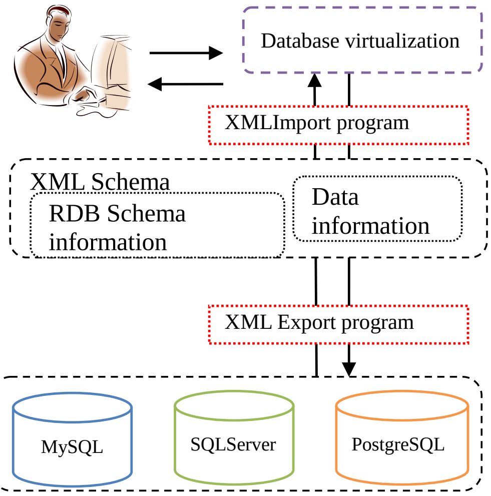


Figure 2: Virtualization technique for RDB databases.

3.2. RDB schema conversion into XML

The following describes how the RDB schema is converted into XML. Figure 3 presents results of reading the schema information from the RDB and converting it into XML. The RDB schema information that is converted into an XML format includes "table names", "field names (associated data types and default values)", and "constraints (primary key constraint, unique constraint, check constraint, NOT NULL constraint, and foreign key constraint)" capability.

Regarding the XML tree structure, we described the table information in the table structure node with its elements of Field="column name", Type="data type", Null="TRUE or FALSE" (NOT NULL constraint). We described the schema information in the schema node with its elements of TYPE="constraint name", Table="table name", Column="column name", ReTable="referenced table name", ReColumn="referenced column name", and Check="rule".

3.3. RDB data conversion into XML

The manner in which the RDB data are converted into XML is described next. Figure 4 portrays results of reading the data information from the RDB and conversion into XML. Because of the XML tree structure, we had dbname="database name", tblname="table name" , and the succeeding column name="actual data".

```
<?xml version="1.0" encoding"UTF-8" standalone="yes"

 <root>

<rdb Name="mysql">

   <database Name="questionnaire" >
       <table_structure Name="member">
       <field Field="samplenum"
                 Type="integer" Null="FALSE"  Default=" />
    <field Field="answerday" Type="text"
     Null="FALSE"  Default=" />

    ….

  </table_structure>

 <schema>
  <constraint Type="PRIMARY KEY"
       Table="member"  Column="samplenum"  />

….

     <constraint Type="FOREIN KEY" Table="questionnaire"
     Column="samplenum" Retable="member"
     ReColumn="samplenum" />

….

 </schema>
</database>
</rdb>
</root>
```

Figure 3. Example of RDB schema information conversion into XML.

Code	Name	Latitude	Longitude
47401	Wakkanai	45.25	141.41
47404	Haboro	44.22	141.42
…	…	…	…

RDB

```
<?xml version="1.0" encoding="utf-8" ?>
<dataset dbname="chihou">
    <data tblname=" AreaInfo ">
       <Code>47401</Code>
       <Name>Wakkanai</Name>
       <Latitude>45.25</Latitude>
       <Longitude>141.41</Longitude>
   </data>
     <data tblname=" AreaInfo ">
       <Code>47404</Code>
       <Name>Haboro</Name>
       <Latitude>44.22</Latitude>
       <Longitude>141.42</Longitude>
   </data>
   <…>
```

Figure 4: Example of actual RDB data conversion into XML.

3.4. Virtualization of databases

We discuss the virtualization of modelled DBs of different types. For virtualization of different types of modelled DB, we describe the schema information of each model using a single common schema. The common schema we will use is an XML Schema. Around it, we will perform virtualization. Figure 1 shows a virtualization method for different database types. To accomplish schema conversion from a different modelled database, we first get the schema information from an RDB to work on. Then we convert it into the correct XML schema for that RDB. We currently have to re-build the XML tree with our schema conversion module when the distributed database is redefined.

Table 1 presents schema conversion correspondences between the two. Because any XML DB is already described in the XML format, we extract the schema information without conversion. On the other hand, when the data are manipulated, our query conversion module automatically transfers the access results to the application program.

3.5. Techniques of execution efficiency improvement

Methods of the execution efficiency improvement of virtualization processing (improvement in the speed) are as follows.

- The place of virtualization processing

In order to accelerate, the virtual database environment which uses load sharing technology and parallel processing technology is shown in Figure 5, and we use both user side virtual DBMS and data side virtual DBMS.

Since the database is distributing through a network and communication time influences the whole processing time greatly, it becomes important to reduce the amount of data transfer for the improvement of the processing speed. Under the virtualization processing the data volume changes. Even if the same data is processed, data volume differs by the schema expression, RDB schema or XMLDB schema. Therefore, the place where the virtualization processing is performed could be changed, so that the amount of data transferred is reduced, and communication time is reduced.

- Database selection

When the same table and data are stored in different databases, it could be considered to make the load of each database uniform by acquiring data from a database with little load. In database virtualization technology, since virtual processing is added in addition to processing of the usual database, balancing of the database load becomes important.

4 DESIGN OF THE SIMULATOR

In improving the execution efficiency, it is necessary to measure the execution performance of the virtualization processing, especially in a distributed environment where multiple databases are connected via a network. But it takes a lot of databases and large-scale network structure, and preparation of actual measurement environment is costly and very difficult. Therefore, the measurement environment using a simulator is considered.

Two of the followings are the basic requirements needed by the simulator.

- Measurement for discovering the causes (bottlenecks) of delay of database virtualization processing can be performed.
- Measurement when the number of databases connected or the volume of each database becomes large on the virtual database environment using load sharing technology and parallel processing technology can be performed.

Table 1: SQL and associated XML

	SQL	XML
Table definition	CREATE TABLE table name…	<xsd: element name=“table name”…
Column definition	CREATE TABLE… column name...	<xsd: element name=“column name”…
Data type definition	CREATE TABLE… data type..	<xsd: element… type=“data type”…
Default values	CREATE TABLE… column name DEFAULT value	<xsd: element… default=“value”…
Primary key constraint	PRIMARY KEY	<xsd: key…
Unique constraint	UNIQUE	<xsd:unique …
Foreign key constraint	FOREIGN KEY	<xsd: keyref … refer =…
NOT NULL	NOT NULL	<xsd:… nillable=“false”...
Method	CREATE METHOD	
Inheritance	CREATE TABLE… UNDER upper level table name	<xsd: complexType …

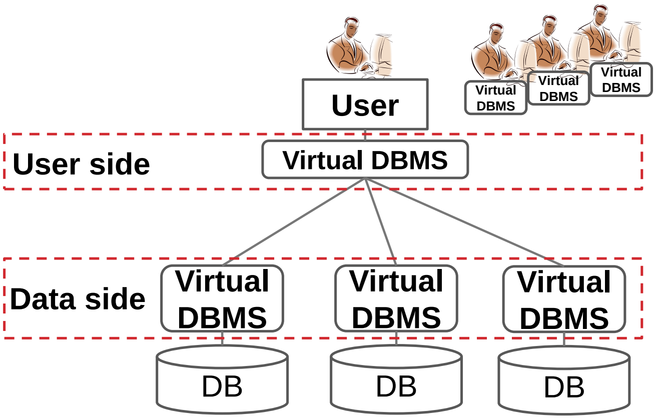


Figure 5: Virtual database environment which uses load sharing technology and parallel processing technology.

4.1 Outline of the simulator

The main purpose of the simulator is the bottleneck discovery of database virtualization processing. For the purpose of this bottleneck discovery, actual processing, such as virtualization processing, database processing and communication processing, are not needed and actual processing is not performed in the simulator. The execution efficiency is computed simulating and integrating each processing time. Random elements such as network congestion, user's command input timing are simulated and computed repeatedly to obtain average and variance.

Prerequisites for database access for the simulation are specified as follows. The data mining and distributed database environments are considered in the simulator and it assumes a limited range of database operations here. For example, database updating and join operations are excluded in the simulator.

By realizing each component such as database processing of the simulator as a process and performing inter-process communication with TCP protocol, the simulator can be implemented on a single PC or on two or more PCs. Followings are prepared as an item which can be changed by setup.

- Number of users
- Number, scale, and kind of databases
- Network line speed

4.2 Measurement items

The following measurement is performed for the overhead identification of virtualization processing. About the reference parameters for the simulation, some preliminary simple virtual processings are performed beforehand and they are determined from the result at the time of implementation.

- Time of virtualization processing

This mainly considers time of conversion such as query conversion from XQuery to SQL and result conversion from RDB result into XML format. The measuring method computes and converts the processing time according to the length of a query, the data volume of the result, etc. based on the reference parameters.

- The change in the data volume after virtualization processing

The data volume fluctuated by virtualization processing of query result is measured.

- Processing time of a database

The processing time of a database is computed from a query. For example, in 'Selection', processing time changes by the existence of indexes. Processing time is changed also by the timing of the database usage and the number of users. If there are some database processing performed during system usage of a user, the wait time of the database processing will be added to the processing time for the user.

- The amount of data transfer

The data transfer rate is adjusted by changing the network utilization factor according to the number of users, users' usage timing, etc. of databases. The system determines the amount of data volume by what kind of query is issued to which database by each user, then decides the amount of data transfer by which network is used for the data transfer.

- Communication time

Communication time is computed using the following formulas.

$$Communication\,time = \frac{Amount\,of\ data\,transfer \times (1 + Rate\ of\ control\,data)}{Line\,speed \times Network\,utilization\ factor}$$

Since the network of a database is classified to class 3 in Network Quality of Service (QoS) of Y.1541 of ITU, delay by congestion is generated in the probability of 10^{-3} based on the class 3 of QoS. Time to be delayed in this case, being unspecified in the class 3 of QoS and not restricted, we make it the interval of the retransmission-of-message packet. The process on the data reception side performs the measurement of communication time.

4.3 Size of packet

Packet size is needed for the determination of the rate of control data or the number of times of communication. The maximum size (MSS: Maximum Segment Size) of the packet changes with MTU (Maximum Transmission Unit) of the data link assuming that the database uses TCP.

The main current data links are Ethernet and PPPoE, and assuming the protocol uses TCP, MTU of Ethernet is used.

The maximum data volume per packet is set to 1460 bytes, and the number of times of communication is (Amount of data transfer /1460) and the rate of control data is (1-1460 / 1518).

4.4 System configuration

Each component is realized by a process so that the each component, such as virtual DBMS, can be executed concurrently. Each component performs inter-process communication with TCP protocol, and the simulator is run on a single PC or two or more PCs. Development language is C and execution environment is Linux.

In order to decide to implement virtualization on user side or data side depending on the measurement result, virtualization process could be performed on both sides. Although designed supposing virtualization of RDB and XMLDB at this time, when adding virtualization of other DB kinds, it is made to be easy to extend. By saving the last setting environment in a file, and calling it easily, the time and effort for the setup for every simulator use is reduced.

The system configuration of the simulator based on Figure 5 is shown in Figure 6. And the component processes of the simulator are classified into following three.

- Interface process for the simulator user

Processing of a simulator user's interface and management of the whole simulator are performed. The setup of the simulator and directions of a simulation start are performed.

- User's process

Processing corresponding to each user using a database is performed. Execution of XQuery, reference of an XML schema, etc. are performed and processing time is sent to the interface process for the simulator user. In a communication module, calculation and conversion of communication time are performed from the data volume of the received result. In a virtualization process module, calculation and conversion of time of virtualization processing from the data volume of a result are performed.

- Handling process of each database

Processing of data side virtual DBMS and database accesses are performed. The processing time for processing of a database and virtualization processing according to a setup of the number of data etc. is computed and converted. In a communication module, calculation and conversion are performed for the communication time of query reception based on the received query. By DB module, calculation and conversion of the processing time concerning query execution are performed and data size or the number of data of the result data are determined. In the virtualization process module, calculation and conversion of time for virtualization of data from the number of result data, etc. are performed.

In a communication module, since a transmitting side process does not need to consider the existence of delay, such as a collision, about measurement of a communication time, the communication module of the receiving side process measures communication time. Specifically, measurement of communication time in case a command is sent to data side virtual DBMS from user side virtual DBMS is performed by the database side communication module and in case a result is sent to user side virtual DBMS from data side virtual DBMS, measurement is performed by the user side module.

5 REFERENCE PARAMETERS FOR THE TIME OF VIRTUALIZATION PROCESSING

Simple and preliminary virtualization processing was performed and the reference parameters of the processing time of virtualization processing and the fluctuation of the data volume after virtualization processing were determined. Although implementation was carried out in Java by the previous work [7], since Java operates on a virtual machine and delay by insufficient memory occurs, we re-implemented the system in C.

At this stage, since database virtualization of only RDB and XMLDB is assumed, only the reference parameters of these virtual processings are obtained. Moreover, execution using an actual database is not performed about processing of a database, but the function which returns dummy data is prepared. The execution environment of preliminary virtualization processing is as shown in Table 2.

The reference parameters obtained in this section are the references only for the environment shown in Table 2. The reference parameters should be reconsidered and modified under other environments.

About the composition of a database, RDB 'Chihou' assumes the database with the table and column shown in Table 3, and assumes the XMLDB database 'Tenkou' which is shown in Figure 7.

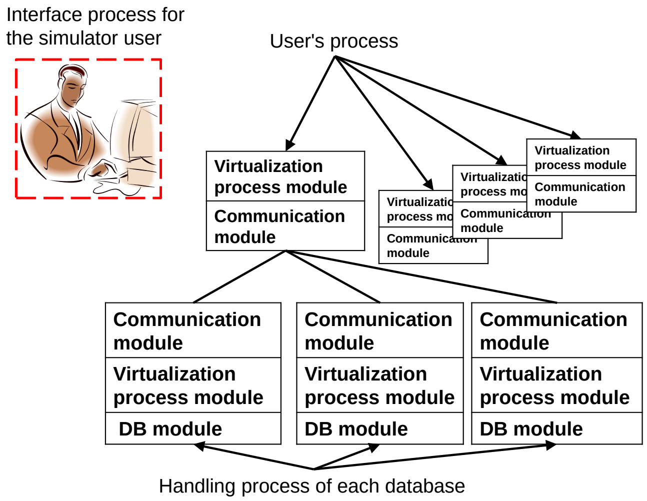


Figure 6: System configuration.

Table 2: Preliminary virtualization process execution environment

OS	Red Hat Enterprise Linux 5 Server
CPU	Quad Core Xeon 2.50GHz × 2
Memory	16GB

The XQuery used for the execution is as follows.

```
for $A in fn:doc(‘Tenkou’)//Item let $B :=
fn:doc(‘Chihou’)//areainfo[@Code=$A/Station/Code]
let $C := fn:doc(‘Chihou’)//observ[@Code=$A/Station/
Code] return <result>{$B/@Code, $B/Area, $B/Kana,
$C/Observ, $A//Precipitation, $A//Precipitations}
</result>
```

This XQuery is a query which acquires data from RDB named 'Chihou' and XMLDB named 'Tenkou'. It is the query of returning the result which acquired from 'Chihou' of the 'let' phrase based on the result of 'Tenkou' acquired with the 'for' phrase, in the form described after 'return' phrase. From the simple execution result of virtualization processing, the following processing execution has measured time.

- Extraction of the result
  When the query of the 'let' phrase in XQuery is transformed and created, in order to describe a conditional sentence using the result performed with the 'for' phrase, the result is stored in the memory temporarily and the extraction time of the result from the memory is required .
- Virtualization processing of a query execution result

Table 3: Structure of RDB 'Chihou'.

Table name	Column name
Areainfo	Code, Area, Kana
Observ	Code, Observ

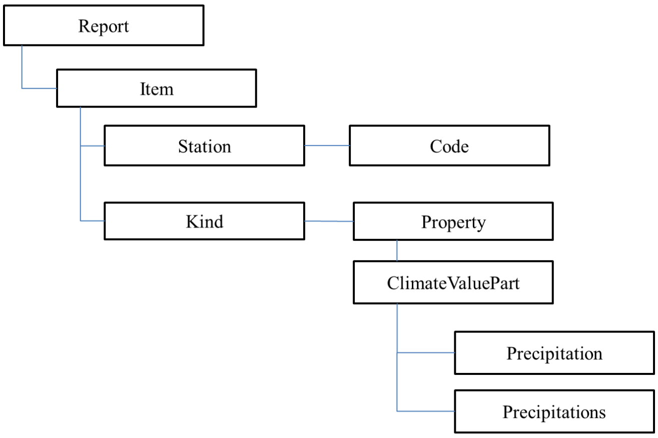


Figure 7: Structure of XMLDB 'Tenkou'.

To determine the reference parameters, queries for above mentioned processing which return 1,000,000 or 2,000,000 result data, are created and executed multiple times. From the execution results, reference parameters are determined as shown in Table 4, 5, 6.

For the read-out time of the result, it is proportional to the number of result data. It takes of 0.130 microseconds per one result data on average.

For the virtualization processing time of the execution result of the query of RDB, it is proportional to the number of result data. Moreover it can be expressed by a linear equation of 0.149 microseconds of inclination and 0.135 microseconds of intercept of the number of columns.

For the virtualization processing time of the execution result of the query of XMLDB, it is proportional to the number of items to acquire and also the number of result data. Therefore, it has taken about 0.420 microseconds per one data item and one result data.

The determined reference parameter of each processing time is shown in Table 7.

Table 4: Read-out time of the result (microseconds).

Number of result data	Average processing time
1,000,000	130,180
2,000,000	261,960

Table 5: The virtualization processing time of the execution result of the query of RDB (microseconds).

	Number of columns	
Number of result data	1	5
1,000,000	285,756	880,766
2,000,000	562,815	1,753,933

Table 6: The virtualization processing time of the execution result of the query of XMLDB (microseconds).

Number of items	Average processing time
1	418,336
3	1,195,618

Table 7: Reference parameter of processing time (microseconds).

Processing	Processing time
Read-out time of the result	0.130 x Number of result data
Virtualization processing time of the execution result of the query of RDB	(0.149 x Number of columns +0.135) x Number of result data
Virtualization processing time of the execution result of the query of XMLDB	0.420 x Number of result data x Number of items

6 CONCLUSION

In this research, the design and implementation of the simulator which measure the execution efficiency of the database virtualization processing in the distributed environment where multiple heterogeneous databases were connected with the network have been performed.

However, verification and evaluation of this simulator itself is left yet. Therefore, it is necessary to advance to the next stage of performing verification and evaluation of the simulator, and perform quantitative measurement of database virtualization processing. From the result, we discover the bottleneck of database virtualization processing, and plan to accelerate the bottleneck parts in the future.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 24500122.

REFERENCES

[1] Garcia, Daniel F. Performance Modeling and Simulation of Database Servers. OJEEE, 2010, 2: 183-188.

[2] WU, Wentao, et al. Predicting query execution time: Are optimizer cost models really unusable?. In: Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013. p. 1081-1092.

[3] WU, Wentao, et al. Towards predicting query execution time for concurrent and dynamic database workloads. Proceedings of the VLDB Endowment, 2013, 6.10: 925-936.

[4] K. Mori, S. Kurabayashi, N. Ishibashi, and Y. Kiyoki, “An Active Information Delivery Method with Dynamic Computation of Users' Information in Mobile Computing Environments”, DEWS2004 1-A-04, 2004.(in Japanese)

[5] teiid, http://www.jboss.org/teiid, Red Hat

[6] DB2: Information Integrator V8.1, http://www.jpgrid.org/documents/pdf/WORK4/sugawara_ws4.pdf

[7] Yuji Wada, Yuta Watanabe, Keisuke Syoubu, Hiroshi Miida, Jun Sawamoto, Virtual Database Technology for Distributed Database in Ubiquitous Computing

Environment, American Journal of Database Theory and Application 2012, 1(2): 13-25 , DOI: 10.5923/j.database.20120102.02

[8] Y. Wada, Y. Watanabe, K. Syoubu, J. Sawamoto, and T. Katoh,“Virtualization Technology for Ubiquitous Databases”, Proc. 4th Workshop on Engineering Complex Distributed Systems (ECDS 2010), pp. 555-560, 2010.

[9] Y. Wada, Y. Watanabe, K. Syoubu, J. Sawamoto, and T. Katoh, “Virtual Database Technology for Distributed Database”, Proc. IEEE 24th International Conference on Advanced Information Networking and Applications Work-shops(FINA2010), pp.214-219, 2010.

[10] Y. Wada, Yuta Watanabe, Keisuke Syoubu, Hiroshi Miida, Jun Sawamoto and Takashi Katoh,” Technology for Multi-database Virtualization in a Ubiquitous Computing Environment”, International Workshop on Informatics (IWIN2010),pp. 89-96, 2010.

A Construction Method of Efficient SkipGraph Using The Performance of Peers in Heterogeneous Environment

Yohei Yasutomo*, Yoshitaka Nakamura** , and Osamu Takahashi**

*Graduate School of Systems Information Science, Future University Hakodate, Japan
**School of Systems Information Science, Future University Hakodate, Japan
{g2113033, y-nakamr, osamu}@fun.ac.jp

Abstract - SkipGraph is an overlay network that was applied to the data structure of SkipList in a peer to peer (P2P) network. Conventional SkipGraph does not take into account communication or the environmental performance of peers and uniformly treats all peers. However, communication and environmental performance differ for individual peers in real environments, and in some cases search efficiency deteriorates depending on the configuration of the topology of SkipGraph. We propose a method of constructing SkipGraph where the transfer delays between peers in small enough. In this method, peers are classified into three types by taking their processing speeds and communication speeds into consideration. We also evaluated the performance of the method of construction.

Keywords: P2P, Structure overlay, SkipGraph.

1 INTRODUCTION

Peer to Peer (P2P) network is network architecture. P2P networks are constructed from individual nodes (such as terminals) without servers, in contrast to the server-client model, and are superior in fault tolerance, scalability, and load distributions. All terminals of P2P network constitute an overlay network in cooperation with each other. These terminals search for contents data and transfer them on this overlay network. Many methods of constructing overlay networks have been proposed. The distributed hash table (DHT) [1] and SkipGraph[2] are typical methods. DHT is a method of effective searching data by managing the keys of data mapped to the same space with peers by hash function between multiple distributed peers. However, it is difficult to carry out range queries with DHT because the order of keys collapses due to the hash function. SkipGraph is an overlay network that applies SkipList [3] to P2P. SkipList is a forward linked list type data structure constructed by a probabilistic algorithm. It is easy to carry out range queries with SkipGraph because it does not treat hash values.

The opportunity to obtain services using P2P technology is increasing for mobile users, because mobile terminals such as smart phones, tablet terminals, and wireless communication technologies such as 3G or Wi-Fi have been developed. However, there are differences in terminal processing performance and communication environments between mobile terminals and fixed terminals. Mobile terminals generally perform worse than fixed terminals, and wireless communication causes long transmission delays. Therefore, mobile terminals have adverse effects in searches of the whole P2P network, when many mobile terminals become relay nodes.

Some researches have proposed solutions to this problem caused by differences in communication environments and peer processing performance. For example, Ref.[4] proposed asymmetry type P2P technology to perform network construction and data transmission processing in consideration of the characteristic of the terminal, and Ref.[5] proposed hierarchical P2P technology using the super node. These researches mainly use DHT as P2P construction method, but much less use SkipGraph.

There are differences in searching costs between peers in SkipGraph due to the participating positions of peers. P2P network with low search efficiency is constructed when terminals with low processing performance and poor communication environment participates in the position taking many search processing, and when terminals with high processing performance and good communication environment participates in the position taking few search processing. In this way, there are many problems to apply SkipGraph in real communication environment. We propose a method to construct efficient SkipGraph by giving priority to the terminals with small delay to take many search processing and limiting the terminals with large delay to take few search processing. We assumed a P2P network where various communication environment and terminal processing performance are mixed and classified terminals into three types by using these characteristics.

2 RELATED WORKS

2.1 SkipGraph

SkipGraph is an overlay network that applies SkipList to P2P. The structure of SkipGraph is outlined in Fig.1. SkipGraph has a number of hierarchies called “level”. “Level” is expressed as a number in the squares in Fig.1. Each peer is expressed with a square in Fig.1, and the number in the square expresses the key of data which the peer holds. This key plays a role as the node ID, and peers from a line in order of a key. The peer of SkipGraph has a bidirectional link in each level. It is determined which peers are linked by “Membership vector” which are the random binary digits. “Membership vector” are three digit numbers under the peers in Fig.1. The peers whose *n* digits prefix of Membership vector is match each other link in Level *n*. The set of all peers linking each other is called “List”. The highest level linked to peers was called “Highest level”. And all peers in Level 0 are linked to peers in ascending order.

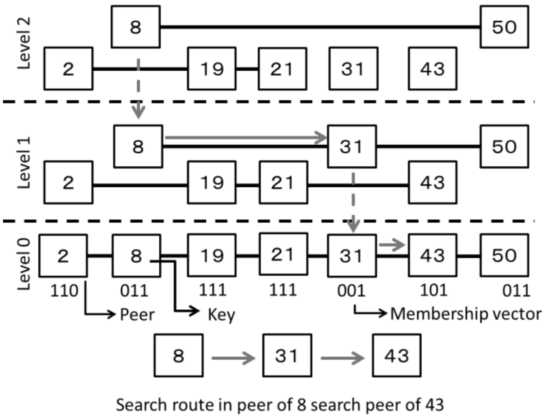


Figure 1: SkipGraph

2.1.1. Searching Process of SkipGraph

A peer (Peer S) in SkipGraph starts to search for a target peer (Peer D) from the highest level. The peer which received a search message compares its own key to the search key. If its key equates to the search key, it sends a completed search message to the Peer S. If the search key is larger than its own key, the peer starts to search other peer with lower key than the search key in the same level. If the peer is not able to find the target peer, it searches for it on the next below level. We explain the search method of SkipGraph with Fig.1. In this example, it is assumes that the peer with the key of 8 (Peer 8) searches for the peer with the key of 43 (Peer 43). First, the Peer 8 searches for the target Peer 43 on Level 2 that is the highest level. The peer searches for a peer having key between 8 and 43 because the target peer has the key of 43. Because Peer 8 cannot find the peer satisfying the search conditions on Level 2, Peer 8 lowers one hierarchy of SkipGraph and searches again on Level 1. The peer on Level 1 is able to find the peer with the key of 31 which is between 8 and 43, and forwards the search message to Peer 31. Peer 31 which received the search message starts to search the target Peer 43. Peer 31 lowers one hierarchy of SkipGraph and searches again on Level 0, because Peer 31 and 43 are on Level 1. And target Peer 43 is found on Level 0. Peer 31 sends the search message to target Peer 43 and completes the search process. The average number of hops which is necessary until search completion is log *N* (the number of all peers is *N*). SkipGraph streamlines searches because its topology is able to forward messages to distant peers on high levels.

2.1.2 Join and Leave Process of SkipGraph

A joining peer sends a message to an existing peer (agency peer) to inform its joining. The agency peer finds neighboring peers of the joining peer by using its key on Level 0, and inform the joining peer about the neighboring peers of the joining peer. After that, the joining peer sends the own membership vector to the neighboring peers on Level 0. Next the joining peer searches the neighboring peers on Level 1. Repeating this process on each hierarchy higher than Level 1, the joining peer knows the neighboring peers on each hierarchy. The average number of the messages becomes log *N* at the time of peer participation (*N* is the number of all peers). A leaving peer in SkipGraph sends messages to neighboring peers to inform its leaving from the highest level to the lowest level of SkipGraph. The neighboring peers reconstruct a topology with the messages. The average number of the messages becomes log *N* at the time of peer leaving.

2.1.3 Extension of SkipGraph

The research trend in SkipGraph is to extend multidimensional range searches [6] or to share multiple keys with peers [7].

Reference [8] proposed a method of construction with proximity in SkipGraph. Conventional SkipGraph may have links with too much delay to communicate because it does not assume physical localization or communication time between peers. To solve these problems, this method constructs SkipGraph with small delay by measuring the communication rate between peers.

2.2 Problems with SkipGraph

In the real communication environment, peers of P2P network are categorized into three types, such as long transfer delay, medium transfer delay, and short transfer delay according to the communication speeds and the processing speeds. Forwarding frequencies of peers in SkipGraph are changed by the position of the peer in the topology. We investigated the number of forwarding of peers at the each highest level when peers search another peers in the SkipGraph constructed by 4000 peers. Figure 2 plots relative values of the number of forwarding in every highest level with Level 12 being 100%. It is a standard at Level 12 because the average of highest level is Level 12 when 4000 peers construct SkipGraph. In Fig.2, we can see that there is 15% of difference with the rate for the number of forwarding of messages between peers under Level 10 and those over Level 17. In brief, the forwarding time increases as the highest level becomes higher. Forwarding messages on high levels in SkipGraph are more than low levels because searches in SkipGraph start from the highest level. Therefore, the forwarding efficiency in higher Level becomes important in search processing. When the peer which needs long time for message forwarding locates on higher highest level, the SkipGraph is inefficiency. However, when the peer with short transfer time is on higher highest level, the SkipGraph is efficient topology.

However, many peers in Fig.2 that have long transfer delay may be located on Level 16 or 17 because the membership vector deciding level is given at random. The position of the peers with long transfer delay influences the efficiency of the topology. Figure 3 is a specific example of construction of the inefficient topology. When Peer A and Peer B search for Peer C in this topology, the forwarding of messages goes through two low performance peers and increases delay. Therefore, this topology increases delay in searching.

Another problem is that Peer D only forwards messages to proxemics peers despite that there are other peers with lower delay.

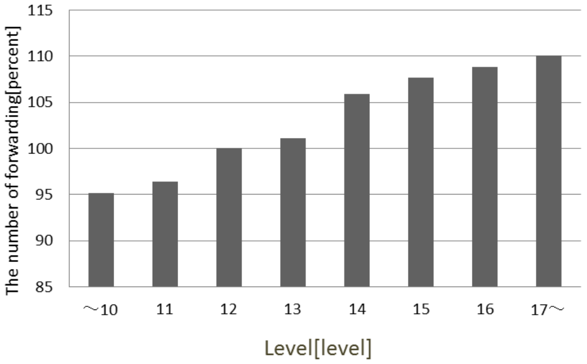


Figure 2: Number of forwarding for each Level

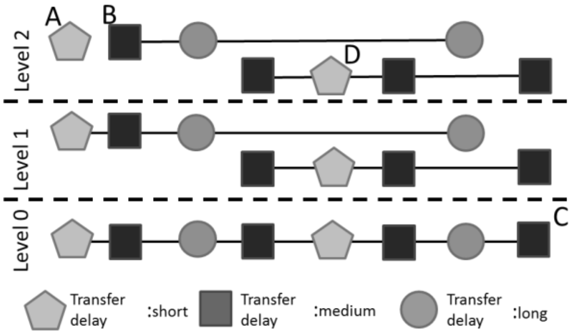


Figure 3: Example of inefficient SkipGraph

3 PROPOSED METHOD

3.1 Basic Concept

Our proposed method classifies peers joining SkipGraph in following three types from the viewpoint of transfer delay.

High Performance Peers A high performance peer is a terminal such as a server and a terminal located on a backbone network. It can forward large volumes of data with wider bandwidth than that of other peers.

Medium Performance Peers A medium performance peer is connected to a network with a general fixed line. Therefore, it is possible to communicate stably.

Low Performance Peers A low performance peer uses wireless communications such as Wi-Fi or 3G. It has slower communication speeds than medium performance peers. A peer using 3G has much more transmission delay than other peers. This type of peer is usually a mobile phone, smart phone or tablet terminal. Its IP address is frequently changing because of switching access points by the movement of the terminal. Therefore, its communication environment is unsteady.

The proposed method constructs topology such as Fig. 4. There are three types of peers in the topology. High performance and medium performance peers account for high forwarding rates at higher levels. To achieve this purpose, the proposed method applies joining and reconstruction method to each three types of peers in SkipGraph. In the joining methods of high and low performance peers, the joining peer find the highest levels that are linked to neighboring peers by the number of all peers. The average of these levels are called ``average levels'' in this paper. And the proposed method sets the membership vector so that high performance peers are located on higher than the average level and low performance peers are located on lower than the average level. Most search messages were assumed by high performance peers in this way. The proposed method decreased the number of search messages sent by low performance peers. Therefore, the proposed method can construct efficient SkipGraph.

The proposed method assumes that peers manage the number of peers participating target SkipGraph to find the average level. We explain the flow of joining and reconstruction methods with high and low performance peers in next subsection. Joining and reconstruction methods of medium peers are omitted from the explanation because a general method used in SkipGraph is applicable.

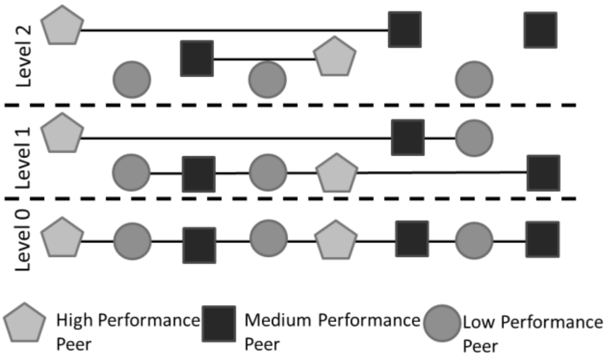


Figure 4: SkipGraph using the proposed method

3.2 Flow of the Proposed Method

Joining peers just get the key of the neighboring peer in level 0, the information about the average level of peers and key of the managing peers from agency peers when joining peers join SkipGraph. All types of joining peer joins SkipGraph with the proposed joining method. Agency peers sent joining messages about joining peers to the managing peer. The managing peer just calculates the average level with the number of peers. And the managing peer informs all peers on the recent average level when the average level is changed. High and low performance peers receive messages reconstruct the topology using the proposed method with the recent average level.

The average level can be calculated by log$N$ when the number of all peers is N. The managing peer informs all peers of the average level which is log$N$ truncated by a decimal point when the recent average level is below the last notified average level. The reason for this is that we could avoid frequent changes in the average level. For example, if the average level is a rounded value, the managing peer has to inform all peers of the average level, whenever the average level frequently keeps changing between three or four when log$N$ ranged in the neighborhood of 3.5.

3.3 Joining and Reconstruction Process of High Performance Peers

First, high performance peers use the general method of participation. When the highest levels of high performance peers is under the average level, they reconstruct the topology, and the highest level of high performance peers are over the average level. In this way, high performance peers can often forward messages at higher levels than other types of peers. We can also shorten the processing time for searches.

The highest level of high performance peers is Level i. High performance peers in Level i send messages to all peers belonging to the same list on Level i and investigate membership vectors of these peers. High performance peers compare these membership vectors with the membership vector of high performance peers that is inverted i +1 digits of prefix. And the high performance peer calculates the new highest level. When the calculated new highest level is higher than the average level, high performance peers reconstruct the topology using the inverted new membership vector. If the highest level does not attain the average level, they send messages to the all peers of the list on the next level below. They investigate the membership vectors and calculate the highest level again. In reconstructing the topology, high performance peers leave from the same level as the inverted digits. After that, high performance peer renew the membership vector, and join to the level using the new membership vector by the general method.

We explain the proposed process to achieve high performance peer (Peer 19) by using Fig.5. The Peer 19 reconstructs the topology in order to set over the highest Level 2 because the highest level of the Peer 19 is 1 while the average level is 2. The Peer 19 sends messages to the peers belonging to the list in level 1 that is highest level. The Peer 19 compares these membership vectors with the Peer 19's new membership vector that is ``110'' and calculates the highest level. Therefore, the Peer 19 understands the new highest level (Level 2) is over the average level. The Peer 19 sends the neighboring peers the renewed membership vectors ``110'' and commonly joins SkipGraph. Finally, the highest level of Peer 19 becomes 2, which is over the average level.

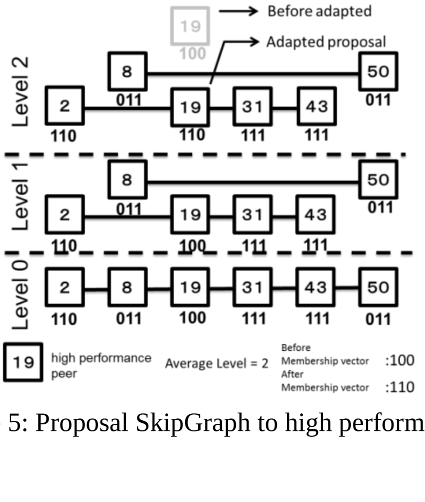


Figure 5: Proposal SkipGraph to high performance peer

3.4 Joining and Reconstruction Process of Low Performance Peers

Forwarding by low performance peers on higher levels is limited by setting the highest level of low performance peers under the average level. We can control the increase of time that occurs when searching with low performance peers on high levels.

When low performance peers join SkipGraph, they set the upper limit of the highest level (limit level) by using the average level. Limit level is calculated as follows.

$$\text{Limit Level} = \text{Average Level} - k \qquad (1)$$

k is one fixed value. We mask the higher digits of membership vector over the digit of the limit level. Low performance peers join SkipGraph with the masked membership vector. If the average level is increased after joining of low performance peers to SkipGraph, they only clear the masks of rising levels. Low performance peers participate in SkipGraph from the current highest level with the renewed membership vector again. If the average level is decreased, the low performance peers mask their membership vector and leaves to prevent the limit level from decreasing.

We explain the method of participation by low performance peers using Fig.6, where the low performance peers of Peer 19 mask the prefixed one digits of the membership vector before joining the topology. In this case, k is 2. The limit level becomes Level 1 by calculated by expression (1). Therefore, the low performance peers mask 2 digits of the membership vector. The low performance peers join with the new membership vector, which is one digit because of masking of two digits. Therefore, the low performance peers can fix the highest level at Level 1, which is lower than the average level while the highest level of low performance peers is Level 2, which is achieved with the general method.

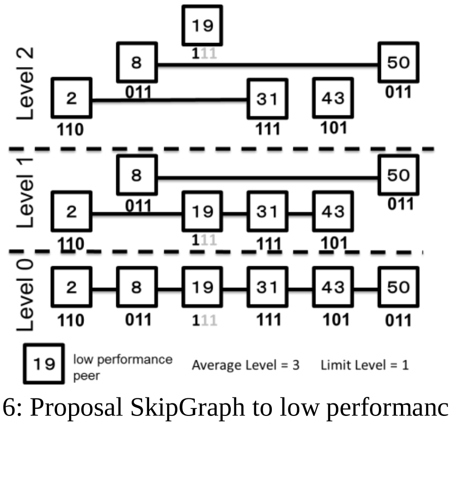


Figure 6: Proposal SkipGraph to low performance peer

4 EVALUATION EXPERIMENT AND CONSIDERATIONS

We conducted the simulation experiment to evaluate the efficiency of the proposed method using PIAX[9]. The proposed method deal with high and low performance peers. Therefore, we evaluated high performance peer, low performance peer and a combination of low and high performance peers.

4.1 Evaluation of High Performance Peers

We explain the experiment to evaluate the efficiency of the proposed method with high performance peers. The number of peers in this experiment ranged from 500 to 4000. The number of high performance peers accounted for 1/3 of the whole peers. The keys of peers are random values from zero to the number of all peers.

Peers select the keys with random values and perform range search in the range 0 to 3. We conducted 50 times of the experiment as 1 trial. And, we measured the peers' number of forwarding until search completion. The average number of forwarding of the high performance and that of other peers using the proposed method for 10 trials are outlined in Fig.7. Figure 7 indicates the number of forwarding of high performance peers are larger than that of other peers regardless of the number of all peers. That means high performance peers are able to forward messages in advance. The reason for this is that the highest level of the high performance peers located over the average level remains stable.

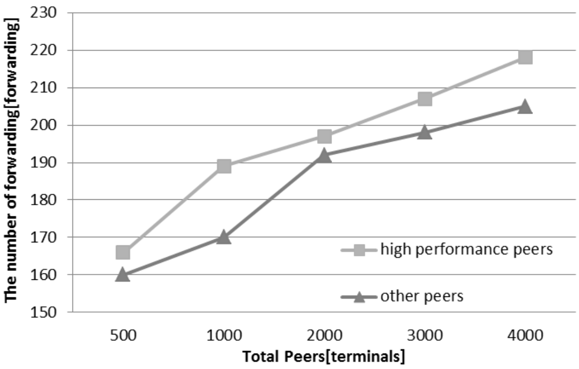


Figure 7: Number of forwarding using the proposed method for high performance peers

4.2 Evaluation and Consideration of Low Performance Peers

We explain an experiment to evaluate the performance of the proposed method with low performance peers. The experiment's setup and evaluation items are similar to those described in Subsection 4.1. The limit level in the experiment is 2 levels smaller than the average level. Figure 8 plots the average number of forwarding.

When there are more than 3000 peers, the number of forwarding of low performance peers decreases. That indicates the forwarding of messages of low performance peers is limited by the proposed method. However, when there are fewer peers than 2000 peers, the number of forwarding of the low performance peers are more than that of the other peers. The reason for this is that the number of peers on high levels is sparse because this experiment apply the proposed method to only low performance peers. Therefore, the search messages are not forwarded on high levels which can send message to peers with keys but forwarded on low levels. And the number of forwarding until search completion is increased because the forwarding on low levels is mainly sent messages to peers having near keys. When there are more than 3000 peers, the number of forwarding times of low performance peers is lower than that of the other peers. The reason for this is that the average level is difficult to shift with many peers because the average level is calculated by log*N*, and the forwarding messages on high levels are more than the case with fewer peers because the density of peers was higher on high levels.

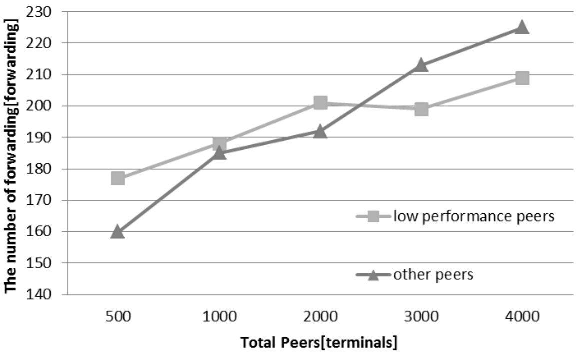


Figure 8: Number of forwarding using proposed method for low performance peers

4.3 Evaluation and Consideration of High and Low Performance Peers

We explain an experiment we carried out to evaluate the proposed method for the combination of low and high performance peers. The experimental setup and evaluation items are the same as those described in Subsection 4.1. In addition, we also evaluated searching hops. Figures 9 and 10 plot the results of the measurements.

Figure 9 indicates that when there are 2000 peers, the forwarding messages by high performance peers are more than that by the other peers, and the forwarding message by low performance peers are less than that by the other peers. Especially when there are 3000 peers, the number of forwarding of high performance peers are more than 10% of that of low performance peers. The reason for this is that the topology constructed by the proposed method raises high performance peers over the average level and low performance peers below the average level remain stable. However, we can find a problem in which forwarding by low performance peers is greater than that by medium peers when there are fewer peers than 2000. The reason for this is similar to the reason given in Subsection 4.2.

When the number of peers is from 500 to 2000 in Fig.10, there is a difference in hops between the general method and the proposed method. However, this is not a large difference and the number of hops to search by the proposed method is similar to that by the general method.

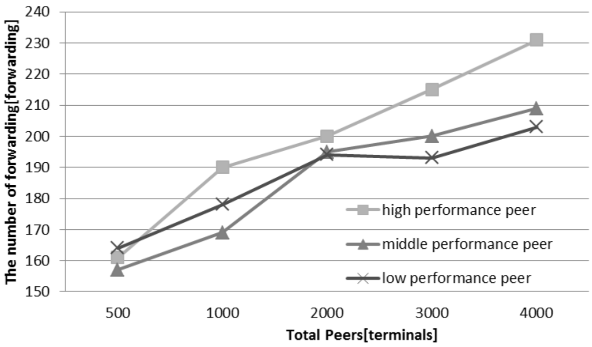


Figure 9: Number of forwarding using the proposed method for low and high performance peers

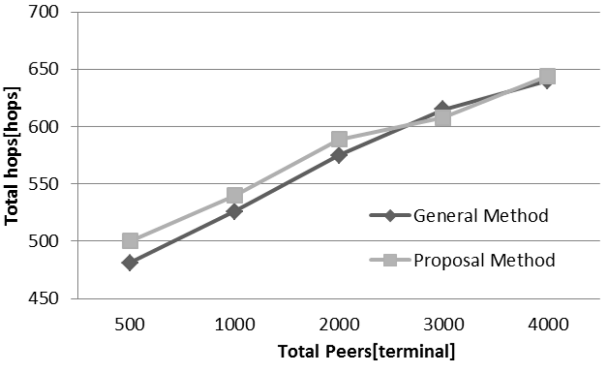


Figure 10: Number of hops using the proposed method

5 CONCLUSION

We considered communication speeds and the processing speeds of terminals and classified terminals into high, medium, and low performance peers. Our proposed method used different participation and reconstruction methods for each terminal to join SkipGraph according to this classification. And we proposed the method to let the terminals with the good communication environment such as high performance peers perform search processing more with precedence, and to limit search processing to the terminals with unstable communication such as low performance peers. We evaluated the efficiency of the proposed method by simulation experiment in varied communication environments. From the results of experiments, the proposed method is able to construct the efficient topology and to fix high performance peers over an average level and low performance peers under an average level by using limit level for all peers.

In the future work, it is necessary to examine the cases that mixture rate of terminals is an inclination to one of the terminal types and is real environment. Our current method decides the highest level of a certain peer based on the average level. For example, when there are too many low performance peers in all peers, too many terminals are located at low levels and very few peers are located at high level. The simulated experiments on other problems produced results in which the proposed approach was not efficient for a few peers.

In addition, low performance peers were assumed to be mobile terminals, which may be defective, have irregularities with SkipGraph, and cause many reconstructions because mobile terminals suffer from the unstable nature of electric waves. We also intend to reevaluate the proposed SkipGraph by addressing these problems.

REFERENCES

[1] Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S., ``A scalable content-addressable network," Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01), pp.161-172 (2001).

[2] Aspnes, J., and Shah, G, ``Skip Graphs," ACM Transactions on Algorithms, Vol.3, No.4, pp.37:1-37:25 (2007).

[3] Pugh, W, ``Skip lists: a probabilistic alternative to balanced trees," Communications of the ACM, Vol.33, No.6, pp.668-676 (1990).

[4] Tagami, A., and Ano, S. ``Design and Implement of AsymmetricP2P Network with Fixed Mobile Convergence Network, "Proceedings of the 2011 Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO2011), pp.121-126 (2011).(in Japanese)

[5] Garces-Erice, L., Biersack, E.W., Ross, K.W., Felber, P.A., and Urvoy-Keller, G., ``Hierarchical P2P Systems," Proceedings of ACM/IFIP International Conference on Parallel and Distributed Computing, pp.643-657 (2003).

[6] Zhang,C., Krishnamurthy, A., and Wang, R.Y., ``SkipIndex: Towards a Scalable Peer-to-Peer Index Service for High Dimensional Data," Technical Report TR-703-04, Technical Report, Department of Computer Science, Princeton University (2004).

[7] Konishi, Y., Yoshida, M., Takeuchi, S., Teranishi, Y., Harumoto, K., and Shimojo, S., ``An Extension of Skip Graph to Store Multiple Keys on Single Node," IPSJ Journal, Vol.49, No.9, pp3223-3233 (2008).

[8] Makikawa, F.; Tsuchiya, T.; Kikuno, T., "Balance and Proximity-Aware Skip Graph Construction," Networking and Computing (ICNC), 2010 First International Conference on , vol., no., pp.268,271, 17-19 Nov. 2010

[9] PIAX: http://www.piax.org

Formal Verification Technique for Consistency Checking between equals and hashCode methods in Java

Hiroaki Shimba[†], Takafumi Ohta[†], Hiroki Onoue[†], Kozo Okano[†]and Shinji Kusumoto[†]

[†]Graduate School of Information Science and Technology, Osaka University
{h-shimba, t-ohta, h-onoue, okano, kusumoto}@ist.osaka-u.ac.jp

Abstract - Java objects used with the standard collection should override both of its equals and hashCode methods. Both of them need to satisfy the consistency rules, or unexpected behaviors may cause faults that are hard to detect. One of previous studies checks whether an equals method satisfies part of the consistency rule. In order to avoid the unexpected behaviors, however, it is necessary to check both equals and hashCode methods satisfy the rules. This research proposes a method which checks the consistency between equals and hashCode methods in Java. We model Java source code and check whether both methods satisfy the rules using an SMT solver called Z3. We have applied our proposed method to some projects in practice. As results, we have detected some of Java source code violating the rules.

Keywords: Java, equals method, hashCode method, Formal Verification, Satisfiability Modulo Theories(SMT)

1 Introduction

In Java, an equals method should be rightly overridden in a class, if its objects are compared. In order to guarantee an appropriate behavior of the collection framework, when a class overrides its equals method, its hashCode method also be overridden[1]. Therefore, Oracle API document defines some rules for the methods in an Object class[2]. For example, an equals method necessary to satisfy reflexive, symmetric and transitive properties. A method violating the rules may cause faults. It is well known that these faults are hard to detect [1][3][4]. Rupakheti et al. [5][6][7] presented a checker called EQ which is designed to automatically detect such an equals method violating the rules. EQ models an equals method and performs model checking to check whether the equals method satisfies part of the rules. Since EQ checks only equals methods, it cannot detect the class may cause fault when such an object is interacted with the collection framework. Also, EQ uses a model description language called Alloy which cannot model the bit operations. Hence, EQ cannot model equals methods using bit operations. Therefore, in order to avoid the unexpected behavior, we propose a new method which checks inconsistency between equals and hashCode methods. We use an SMT solver called Z3[8] to manipulate an arithmetic operations and bit operations which are often used in a hachCode methods. Since implementation patterns of equals and hashCode method are different, we propose new implementation patterns of hashCode methods. Also, we propose a method which converts Java code to an expression in a model description language called SMT-LIB[9]. We have applied our proposed method to some projects in practice. As results, we have detected some of Java source code violating the rules. The rest of this paper is organized as follows. Section 2, Section 3, Section 4, Section5, Section 6 and Section 7 present the consistency rules for equals and hashCode methods, a details of Z3, a motivation example, how convert Java code to SMT-LIB, an evaluation of our proposed method and discussion and a conclusion of this paper, respectively.

2 Consistent rules

This section presents the rules which equals and hashCode methods must satisfy.

2.1 Java Object class

Java Object class is defined as “root of the class hierarchy. Every class has Object as a superclass. All objects, including arrays, implement the methods of this class.” by Oracle API document[2].

2.2 Consistent rules for equals methods

An equals method for Object class determines whether some other object supplied through its argument equals to this object. An equals method must satisfy the following four rules except a null object[2].

- reflexive: for any non-null reference value x, x.equals(x) should return true.
- symmetric: for any non-null reference values x and y, x.equals(y) should return true if and only if y.equals(x) returns true.
- transitive: for any non-null reference values x, y, and z, if x.equals(y) returns true and y.equals(z) returns true, then x.equals(z) should return true.
- For any non-null reference value x, x.equals(null) should return false.

The equals method for Object class is defined as follows[2]. “The equals method for class Object implements the most discriminating possible equivalence relation on objects; that is, for any non-null reference values x and y, this method returns true if and only if x and y refer to the same object ($x == y$ has the value true). Note that it is generally necessary to override the hashCode method whenever this method is overridden, so as to maintain the general contract for the hashCode method, which states that equal objects must have equal hash codes.”

```
public class Sample{
private int val;
private String str;
public boolean equals(Object obj){
if (obj == null)
return false;
if (this == obj)
return true;
if (!(obj instanceof Sample))
return false;
Sample that = (Sample) obj;
if (this.str == null){
return that.str == null;
}
return this.val == that.val && this.str.equals(that.str)
}
public int hashCode(){
return val + (this.str == null ? 0 : this.str.hashCode());
}
}
```

Figure 1: example of correct implementation of equals and hashCode methods

2.3 Consistent rules for hashCode methods

The hashCode method returns a hash code value for the object. This method is supported for the benefit of hash tables such as those provided by HashMap. The hashCode method must satisfy the following two rules[2]. In this definition, information implies the returned value from the method invoked by its equals method or a field value used in the equals method. Thus, if there are some inconsistency between equals and hashCode methods, rule violation occurs.

- Whenever it is invoked on the same object more than once during an execution of a Java application, the hashCode method must consistently return the same integer, provided no information used in equals comparisons on the object is modified. This integer need not remain consistent from one execution of an application to another execution of the same application.

- If two objects are equal according to the equals(Object) method, then calling the hashCode method on each of the two objects must produce the same integer result.

The hashCode method for an Object class returns a different integer value for each different instances. Figure 1 shows an example of a correct implementation of equals and hashCode. The sample class has val and str as the integer and String type field values. The equals method for sample class determines whether an argument is the instance of the sample class after determines whether an object passed as the argument is identical to itself. Next, if the field value str is null, the equals method checks whether the str in passed object is also null. Finally, it determines whether the value of val and the string of str are identical. The hashCode method for the sample class concatenates the value of val and hash value of str. The sample class satisfies the consistent rules for both of equals and hashCode method.

3 Related works

Researches about implementation and design of method in Object class proposed the method that automatically generate the equals and hashCode methods. Rayside et al. have proposed a method which automatically generates the equals and hashCode methods which match the user demands by using a annotation of classes and methods [10]. This study performs dynamic analysis of source code. Grech et al. solved the problem of Rayside research that it consumes long time verifying cyclic objects by analyzing source codes statically[11]. Also, Jensen et al. proposed an annotation which guides the user when user copying objects by using clone method[12]. Recently, researches using model checking, SAT solver and SMT solver gain the attentions. Anastasakis et al. proposed a conversion method that converts class diagrams of UML with OCL to Alloy [13]. This research helps the developer who would like to perform verification about Alloy without knowledge of Alloy. Liu et al. suggested scalable bounded model checking by representing object oriented languages as bit vector of SMT solver[14]. This research supports high speed verification. Balasubramaniam proposed a constraint solver MINION that has high scalability and equips many functions[15]. Also, they proposed a method that automatically generates a constraint solver optimized to each domain[16]. This research helps generating the domain specific constraint solver. Burdy et al. proposed the method that statically verifies Java source code[17]. This method specifies the code location that may cause exceptions such as a NullPointerException. Also, it can verify Java source code annotated with JML. It is able to check whether each method satisfy the its constraints base on JML.

3.1 EQ

EQ checks whether equals method in Java satisfies the consistency rules. EQ receives a type hierarchy and outputs whether equals method satisfies the consistency rules. Here after, a type hierarchy is a structure of classes and interfaces represented as a DAG (Directed acyclic graph). Except Object class, classes and interfaces which have an inheritance relationship are belonged into the same type hierarchy. EQ consists of the following four steps. 1) Perform path analysis for equals method. 2) Analyze the pattern of equals method. 3) Convert Java code to a model described as Alloy. 4) Verify the model by alloy analyzer. EQ has two problems. One problem is that EQ dose not check whether a hashCode method satisfies the consistent rules. The other is that since alloy cannot model bit operators, alloy cannot model equals methods using bit operators. In this study, in order to solve those two problems, we use Z3 not Alloy.

```
public class COSString extends COSBase{
    public byte[] getBytes(){
        …
    }
    public boolean equals(Object obj){
    return (obj instanceof COSString)&&
        java.util.Arrays.equals
        (((COSString)obj).getBytes(),getBytes())
    }
    public int hashCode(){
        return getBytes().hashCode();
    }
}
```

Figure 2: hashCode methods violating consistency rules in PDFBox of Apache

3.2 Z3

SMT（Satisfiability Modulo Theories) problem is a decision problem for logical formulas expressed in first-order logic. An SMT solver solves SMT problems automatically. The SMT solver determines if a given logic formula which combination of theories expressed in first-order logic is satisfiable. If theories are satisfied, the SMT solver outputs assignments for variables that makes given theory satisfied. SAT problems described as theories that consists of only propositional variables. On the other hand, SMT problems described as theories that consist of propositional which can be many types such as Int similar to types in programming language. Also, SMT problems can define and use functions. In this study, we determine if both equals and hashCode method satisfy the consistency rules by using the SMT solver called Z3 exhaustively[3]. Z3 can use the arithmetic operations, bit vectors, arrays and recode types. Since an SMT solver searches the answer in bounded space exhaustively, it can verify there are no assignment which violates the consistency rules.

4 The Motivative example

In this section, we motivate this study by showing an example.

EQ[7] detected equals methods violating the consistency rules by experiments for four open source projects. The class implemented such equals methods may cause fault that is hard to detect. If an instance of a class which implements its equals method violating the consistency rules is used in the standard collection, unexpected behavior might cause faults. For example, if an instance of class which has the equals method violating reflexive is used in standard collection, a contains method of standard collection cannot determines correctly whether collection contains such a instance. Since in order to check equivalence of instances, a contains method of collection such a List uses equals methods, an unexpected behavior may occur. Also, if equals methods judge two instances are equivalent but these two instances return different hash values, hashCode methods cannot perform correct behavior. For example, HashMap may contain two instances judged equivalent by equals methods. Figure 2 shows our motivative example. This example shows an implementation of the hashCode method violating the consistency rules in PDFBox of Apache[18].

```
public class ArEntry implements ArConstants{
    private String filename;
    public String getFilename() {
        return this.filename;
    }
    public boolean equals(Object it) {
        if (it == null || getClass() != it.getClass())
            return false;
    return equals((ArEntry) it);
    }
    public boolean equals(ArEntry it)
    if (this.filename == null)
        return (it.getfilename() == null);
    else
        return
            this.getFilename().equals(it.getFilename());
    }
    public int hashCode() {
        return super.hashCode();
    }
}
```

Figure 3: Conversion example of Java source code

PDFBox uses java.util.Arrays.equals as equals method of COSString class. Also, PDFBox uses the hashCode method of byte array as the hashCode method of COSString class. Hence, equals method checks if two arrays have the same number of the elements and all corresponding pairs of the elements in the two arrays are equal The hashCode method checks these two arrays have the same memory address. Therefore, if instances of arrays are different and these array have the same elements with the same order, the equals method judges these two objects are equivalent but the hashCode method returns a different hash value for each other. In this case, HashMap may contain two instances judged equivalent by the equals methods. HashMap must not contain many instances judged equivalent by the equals methods. Since many cases are can be thought, it is difficult to detect the fault. For example, an insert procedure has fault and collection has fault.

In order to avoid such unexpected behavior, we propose new method that check whether both equals and hashCode methods satisfy the consistency rules.

5 Our proposed method

Our proposed method analyzes the Java code and models behavior of both of equals and hashCode methods in a model description language called SMT-LIB. The model is checked by Z3. Our proposed method receivesthe type hierarchy of the code and then outputs whether each of equals method sat-

```
;Class information
(declare-datatypes () ((Type ArEntry ArConstants UnderARC Object Null)) )
...
(declare-datatypes () (( Ref(Rfield (eqnum Int) (hsnum Int) (pointer Int)) )) )
(declare-datatypes () ((ArEntry(Arfield (filename Ref)) )) )
(declare-datatypes () (( Object(Ofield (ar ArEntry)(pointer Int)(class Type)))))
(declare-const this Object)
(declare-const that Object)
(declare-const other Object)
(declare-const nobj Object)
...
;method information
(define-fun equalsRef ((r1 Ref)(r2 Ref)) Bool
   (ite (and (and (not (= (pointer r1) 0)) (not (= (pointer r2) 0))) (= (eqnum r1)(eqnum r2))) true false ))
(define-fun equalsMain ((o1 Object)(o2 Object)) Bool
   (and (=> (or (= (class o1) ArConstants) (or (= (class o1) UnderARC)(= (class o1) Object)))
         (= (pointer o1)(pointer o2)))
         (=> (= (class o1) ArEntry) (and (and (not(= (pointer o2) 0)) (= (class o1)(class o2)))
         (or (and (= (pointer(filename (ar o1))) 0) (= (pointer(filename (aro2))) 0))
         (equalsRef (filename (ar o1)) (filename (ar o2)))) ) )
   )
)
(define-fun hashCode ((o1 Object)) Int(pointer o1))
;equality check
...
(assert (not (equalsMain this this) ) )
...
(assert (not(iff (equalsMain this that) (equalsMain that this)) ) )
...
(assert (not(=> (and (equalsMain this that) (equalsMain that other))
(equalsMain this other)) ) )
...
(assert (not(=> (not(= (pointer this) 0)) (not(equalsMain this nobj))) ) )
...
;hashCode check
(assert (not(=> (equalsMain this that) (= (hashCode this) (hashCode that) )) ) )
...
```

Figure 4: Conversion example of SMT-LIB

isfies the consistency rules. Our proposed method consists of the following four steps. 1) It perform path analysis for equals method. 2) It analyzes the pattern of the equals method. 3) It converts a given Java code to a model described in SMT-LIB. 4) It verifies the model by the Z3. Path analysis generates a control flow graph, and performs data flow analysis. Data flow analysis specifies what class is referred by a reference variable at each position of the source code and specifies what methods are called. Then, specified methods are inlinined into equals or hashCode methods if it is needed. Equals or hashCode methods perform some types of procedure. Therefore, pattern analysis classifies each method into some patterns. It is difficult to directly convert the hashCode procedures which contain loops including arithmetic operation or library calls, we analyze this procedure using heuristics operations. After pattern analysis, we convert Java code to SMT-LIB based on information from pattern analysis. Also, in order to check the violation of the obtained consistency rules, we also give some constraints to the SMT-LIB model. It is very difficult to model the first consistency rule of hashCode method. Please recall that the rule is “Whenever it is invoked on the same object more than once during an execution of a Java application, the hashCode method must consistently return the same integer, provided no information used in equals comparisons on the object is modified. This integer need not remain consistent from one execution of an application to another execution of the same application.” In order to model this rule, it is necessary to model the concept of the time. However, since first-order logic cannot represent the concept of the time, an SMT solver cannot check the first consistency rule of hashCode methods. Therefore, in order to resolve this problem, we introduce more strict consistency rule which replaces the first hashCode rule. On the other hand, since the second consistency rule of hashCode methods is representable in the first-order logic, an SMT solver can check the second consistency rule of hashCode methods directly. The substituted consistency rule of hashCode method is as follows. We define the first rule below as the Subset rule and second one as the Equivalence rule.

- Subset rule: Set of the fields used in the hashCode methods must be subsumed by the set of fields used in equals methods.
- Equivalence rule: If two objects are equal according to the equals(Object) method, then calling the hashCode method on each of the two objects must produce the same integer result.

Figure 3 and 4 show that an example of convert a Java source code (Fig.3) to a model written by SMT-LIB (Fig. 4). In this example, there are three classes in a type hierarchy. That is, an ArConstants interface, ArEntry class which implements ArConstants and overrides equals and hashCode method, and a class implementing ArConstants but do not override equals and hashCode method (this class represented as UnderARC in Fig.4). Figure.4 represents the SMT-LIB model of the source code in the type hierarchy. Figure.4 represents a declaration of types by the class information, a definition of the method behavior by the method information and the constraints used in validation by equality check.

5.1 Path analysis

Path analysis is similar to that of [7]. At first, our method searches equals and hashCode methods. Our method traces the inheritance relationship for a class which does not override its equals and hashCode methods. If we detect the class which overrides equals and hashCode methods, we regard equals and hashCode method of its parent class as the equals and hashCode method of such class. If there are no overrides of equals and hashCode methods in a inheritance relationship, we regard equals and hashCode method of Object class as the equals and hashCode in such class. Next, we analyze Java byte code using Soot[19] and generate its control flow graph. This control flow graph is represented by Jimple. Jimple represents a Java source code as three-address code , each expression consists of one operator, two operand, and one variable which stores the result of operation. Hereafter, we analyze a Jimple code generated by the Soot.

Next, our method performs path analysis. At first, our method enumerates paths using the obtained control flow graph. Next, our method performs data flow analysis for each path, and specifies what class is referred from a reference variable at each source code location and what methods are called. By this information, our method performs inlining the method invocations in equals or hashCode methods. However, since there are very large number of method invocation, our method limits the inlining. Our method only inlines the method invocations only in the type hierarchy. Also our method does not inline a getter method which is modeled as refer directly the field values. Although, Our method does not inlines outer methods, it models methods of Object class, wrapper classes, Array classes and Collections, because our behavior of these method are already well-known.

Finally, our method trims the path which is unreachable and not necessary to our model. Since our method models equals method as returns true， we trim the path which returns false. Also, in order to avoid modeling the null pointer exception, our method trims the path which includes uninitialized reference variables. Our method enhances the performance by trimming the path which is not necessary to model.

5.2 Analyzing the pattern of methods

In this step, our method analyzes the pattern of the procedure in equals and hashCode methods. By referring the modeling rules for each pattern, our method converts Java source code to SMT-LIB. Also, beside the pattern analysis, our method checks whether subset rule is violated in this step.

5.2.1 Analyzing patterns of equals methods

EQ introduce the six pattern of procedure in equals methods. Our method analyzes what pattern matches the equals methods. Six procedure pattern are equivalence checking of array, equivalence checking of List, equivalence checking of Set, equivalence checking of Map, type checking and state checking. Type checking checks whether there are type checking by instance operator in if expression, typecast by cast operator, type checking by getClass method in Object class. State checking checks whether there are equivalence checking of field values and checking reference variable is not null. Equivalence checking of array, List, Set, and Map checks whether there are comparison of the elements in each structure by loop.

5.2.2 Analyzing patterns of hashCode methods

We introduce the pattern of procedure of hashCode methods and defies the rules of each procedure. The hashCode methods procedure pattern are converting to int, bit operation and arithmetic operation in loop. Converting to int checks whether there are type converting by cast operation and type converting by library method of wrapper class. Arithmetic operation in loop checks whether there are procedure of add operation in loop.

5.2.3 Checking of the subset rule

Our method performs checking of subset rule. Our method collects a set of field variable used in equals and hashCode method by analyzing the equals method and hashCode method, and checks whether set of field variable used in hashCode methods are subsumed by set of field variable used in equals methods. If hashCode method invoke the method of parent classes and other methods since path analysis inlines the method of parent classes and other methods in hashCode methods, set of field variable used in hashCode method contains field variables used in such method. If values of variables in method of parent classes and other methods are changed, the change affects the return value of equals and hashCode methods. Therefore, since it is necessary to consider such field values, we substitute subset rule for the first rules of hashCode methods. Two cases occur in the consistence rule of hashCode methods. One is hashCode methods use fields values used in equals method In this case, if field values used in

```
(declare-datatypes () ((Type ArEntry ArConstants UnderARC
   Object Null)) )
(define-fun subof ((t1 Type) (t2 Type)) Bool
   (ite (or (= t1 Null) (= t2 Null)) false
      (ite (and (= t1 ArEntry) (= t2 ArConstants)) true
         (ite (and (= t1 UnderARC) (= t2 ArConstants)) true
            false
         )
      )
   )
)
(declare-fun instanceof (Type Type) Bool)
(assert (forall ((x Type) (y Type))
   (=> (subof x y) (instanceof x y))))
(assert (forall ((x Type) (y Type))
   (=> (and (instanceof x y) (instanceof y x))
      (= x y))))
(assert (forall ((x Type) (y Type) (z Type))
   (=> (and (instanceof x y) (instanceof y z))
      (instanceof x z))))
(assert (forall ((x Type)) (= (instanceof Null x) false) ))
(assert (forall ((x Type)) (=> (not(= x Null)) (instanceof x
   Object) )))
(assert (forall ((x Type)) (=> (not(= x Null)) (instanceof x x) )))
(assert (forall ((x Type)) (=> (not(= x ArEntry)) (not(instanceof x
   ArEntry)) )))
(assert (forall ((x Type)) (=> (not(= x UnderARC))
(not(instanceof x UnderARC)) )))
```

Figure 5: Model of the instanceof operation

equals method are not changed, hash values also not change. The other one is hashCode methods use not only field values used in equals method but also field values not used in equals methods. In this case, nevertheless field values used in equals method not change, hash values possibly change. In order to check this case, it is necessary to check relationships of field value used in equals and hashCode methods. Since it is necessary to check all method which modifies field values, analyzing it consumes much resource.

5.3 Conversion of Java source code to SMT-LIB

This step consists of the the following two steps. 1) basic structure conversion converts methods, inheritance relationships, classes and field values to SMT-LIB. 2) procedure of method conversion converts the procedure of the method to SMTLIB by using information obtained from the step of analyzing the pattern of methods.

5.3.1 Basic structure Conversion

Our method represents classes and fields by records in SMT-LIB. Our method defines fields used in equals and hashCode methods. It converts all primitive values to Ints in SMT-LIB. Since equals methods perform only comparison, Int has enough power to represent the result of equivalence checking.

Table 1: Part of simple μ conversion rules

$\mu(n_1{+}n_2\ )$	=	+ $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1{-}n_2\ )$	=	- $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1{*}n_2\ )$	=	* $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1/n_2\ )$	=	/ $\mu(n_1)$ $\mu(n_2)$
$\mu(a_1{==}a_2\ )$	=	= $\mu(a_1)$ $\mu(a_2)$
$\mu(n_1{<}n_2\ )$	=	$<$ $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1{>}n_2\ )$	=	$>$ $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1{>=}n_2\ )$	=	$>=$ $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1{<=}n_2\ )$	=	$<=$ $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1! =n_2\ )$	=	not(= $\mu(n_1)$ $\mu(n_2)$)
$\mu(b_1 \Vert b_2\ )$	=	or $\mu(b_1)$ $\mu(b_2)$
$\mu(b_1\&\&b_2\ )$	=	and $\mu(b_1)$ $\mu(b_2)$
$\mu(!b_1\ )$	=	not $\mu(b_1)$
$\mu(a_1 instanceof a_2\ )$	=	instanceof $\mu(a_1)$ $\mu(a_2)$
$\mu(a_1$. getClass())	=	class $\mu(a_1)$
$\mu(T_1$. class)	=	$\mu(T_1)$
$\mu(b_1?a_1{:}a_2)$	=	ite ($\mu(b_1)$) ($\mu(a_1)$) ($\mu(a_1)$)
$\mu(n_1 \vert n_2\ )$	=	bvor $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1\&n_2\ )$	=	bvand $\mu(n_1)$ $\mu(n_2)$
$\mu(n_1\ \hat{}\ n_2\ )$	=	bvxor $\mu(n_1)$ $\mu(n_2)$

Although hashCode methods perform any types of arithmetic operations, since hashCode methods usually perform typecast to int type before arithmetic operations, our method always converts primitive types used in hashCode methods to Ints. Our method converts the enumeration field to the enum type in SMT-LIB. Since reference variables of enum types possibly refer null, our method models add NULL value to the identifier introduced by the enum type. Also, since the enum type of hashCode methods invokes a hashCode method of Object class, our methods models the enum type of hashCode methods as returning the different values for each identifier. Our method defines reference type fields by introducing new record Ref representing a reference type. Ref represents the object that is out of type hierarchy. Our method models such an object based on the hypothesis that such a method satisfies the consistency rules of equals and hashCode methods. Ref defines a field variable that represents reference of its object. It is used in equivalence checking as Int type field. Our method defines the equals methods of Ref when an Ref object is used. Our method does not define hashCode methods of Ref. It models this as a reference of the hash values. Our method models the data structure of Java by arrays and lists. Our method represents arrays, Sets, Maps using arrays of SMT-LIB. An array of SMT-LIB is defined by specifying the type of its index and its type of elements. For example, specifying the type of index as Int represents the array. Set is also represented by adding a constraint in which elements are differ from each other to this array. Our method represents the inheritance relationship of a class by nest of records. However, it cannot model the behavior of instanceof which checks whether a class has a inheritance relationship between other classes. Hence, our method introduces type named Type which enumerates the type of adds null to all class in the type hierarchy. Our method models instanceof operator by representing the relation ship of Type. Figure 5 shows an example of an instanceof operation model. Definition of Object class defines all class as field. Object class represents the runtime objects and defines pointer as Int type. Type defines a field representing where the instances comes from.

5.3.2 Conversion of the procedure of methods

Conversion of the procedure of method converts Java source code to SMT-LIB based on information obtained from step of analyzing the pattern of methods. First, our method generates expression trees for each expression represented as Jimple. Our method specifies the final expression returned by the return expression by tracing the expression tree and analyzing how valued of variables are calculated. The operation in expressions are converted by converting rules. Table 1 shows the simple converting rules of Java source code to SMT-LIB. The convert function converts Java source code to SMT-LIB, where bm and am represent an subexpression of boolean type and numerical type, respectively. Tm represents arbitrary types. Java represents an expression with infix notation while SMT-LIB represents expressions by prefix notation. Also, our method converts instanceof operator based on modeling previously described.

5.3.3 Conversion of equals methods

Our method converts equals methods based on six patterns obtained from the pattern analysis. Operations used in type checking are converted as shown in Table1. Since verification by an SMT solver is performed on the object level, cast operations used in equals method are not converted. Since statement checking compared values, the comparison expression is converted as Table 1. With regard to equivalence checking of arrays, Lists, Sets and Maps, our method models the method which performs comparison in the loop as performs comparison each element of an array. For example, lets' consider an instance of a class which has the array as the field, and performs equals method. Our method checks whether this equals method performs comparison of its field array with array of its argument by the same index. Next, our method checks whether a variable used in a loop header is used as index of array. If those two conditions are satisfied, our method determines it performs comparison. Most of loop operations in an equals match this pattern. Since other loop operations are rarely performed and SMT-LIB cannot evaluates statements dynamically, our method does not model such loop operations.

5.3.4 Conversion of hashCode methods

Our method converts hashCode methods based on the six patterns obtained from pattern analysis. Variables changed its type by cast operation or a method of Java class library is represented as Int type of SMT-LIB. Operands of bit operations are represented as 8bit type vector type,. Conversion results of the operations to Int types by applying the bv2int functions to the result. Although Int of Java is 32bit, if it models it as 32bit, modeling takes an enormous amount of time. Therefore, our method models it as an 8bit integer. Bit operations of hashCode methods operate two operands and not performs bit operations on specific one bit. Hence, our method can performs verification. Arithmetic operations in a loop are analyzed and our method determines what pattern matches the operations. Arithmetic operations in loop can be represented as expression, if the number of iteration is identical to the length of the array and arithmetic operations performed in loop do not contain nondeterministic values. However, the result of this operation is decided after the loop is terminated. Therefore our method limits the loop iteration. This is well used in bounded model checking. Our method calculates the result of the loop after 0 to 10 iterations. Our method cannot verify all cases but if our method decides a hashCode methods violate the rule, this decision is absolutely true. Similar reason of the equals method, our method does not model other loop operations.

```
unsat
  (error "line 74 column 17: model is not available")
unsat
  (error "line 80 column 22: model is not available")
unsat
  (error "line 86 column 28: model is not available")
unsat
  (error "line 92 column 22: model is not available")
sat
  ((this (Ofield (Arfield (Rfield 8 9 7)) 3 ArEntry))
  (that (Ofield (Arfield (Rfield 8 9 10)) 2 ArEntry)))
```

Figure 6: Results of verifying the code of Figure 4

5.3.5 Additional Constraints

Our method verifies the four consistency rules of equals methods and the equivalence rule of hashCode methods by an SMT solver. SMT solver solves the constraint and show assignments which is a set of values for the variables that satisfies all constraints. Therefore, in order to achieve an example of a type hierarchy which violates the consistency rule, our method introduces the negation of the consistency rules as the constraints.

5.4 Solving constraint by an SMT solver

Our method verifies the SMT-LIB expression which models Java source code using an SMT solver called Z3. In general, Z3 determines whether a given set of constraints is satisfiable or not. If it is unsatisfiable, it also outputs a counterexample which is a set of assignments of variables and interpretation of functions.

Since our method uses the negation of the consistency rules as the constraints in SMT-LIB, if Z3 outputs unsatisfiable, then we conclude that the source code does not violate the consistency rules. On the other hand, Z3 outputs satisfiable, we conclude that the source code violates the consistency rules In such a case, Z3 can output a set of assignments which makes the input true.

Figure 6 shows the results of verification by Z3 for the source code in Figure 4. The bottom line show the result of verifying the equivalence rule of the hashCode method and the other four lines are the results of verifying consistency rules of equals methods. Figure 6 shows that violation of the equivalence rule is detected. The optional outputs as as-

```
public class HCatFieldSchema implements Serializable {
   public enum Category {
      PRIMITIVE,ARRAY,MAP,STRUCT
   };
   String fieldName,typeString;
   Category category ;
   ...
   public boolean equals(Object obj) {
      if (this == obj)
         return true;
      if (obj == null)
         return false;
      if (!(obj instanceof HCatFieldSchema))
         return false;
      HCatFieldSchema other = (HCatFieldSchema) obj;
      if (category != other.category)
         return false;
      if (fieldName == null) {
         if (other.fieldName != null) {
            return false;
         }
      } else if (!fieldName.equals(other.fieldName)) {
          return false;
   }
      if (this.getTypeString() == null) {
         if (other.getTypeString() != null) {
             return false;
         }
      } else if (!this.getTypeString().equals(other.getTypeString())) {
         return false;
      }
         return true;
   }
   public int hashCode() {
      int result = 17;
      result = 31 * result + (category == null ? 0 : category.hashCode());
      result = 31 * result + (fieldName == null ? 0 : fieldName.hashCode());
      result = 31 * result + (getTypeString() == null ? 0 :
      getTypeString().hashCode());
      return result;
   }
}
```

Figure 7: A fixed HCatFieldSchema class

signments show that two ArEntry objects have the same field value but their references are differ.

6 Experiments

In this section, we evaluate our proposed method by experiment. We implement the verification function of the subset rule, a part of modeling to SMT-LIB and the verification function to our tool. We did not implement converting of bit operations and loops. These are one of the future work. Subsection 6.1 shows the results of applying our tool to some projects. The results show the effect of methods violating the subset rule. Subsection 6.2 shows the results that whether out tool can detect the violation of the consistency rules of equals methods. In the experiments, we firstly converted convert Java source code in practice to SMT-LIB manually. Then, we applied our tool to that model. Subsection 6.3 shows the execution time of our tool. Subsection 6.4 shows how often projects in practice violate the rules.

Table 2: Results of violation to the subset rule

Name	NumClass	Subset	Violation
Lucene	110	106	4

6.1 Evaluation of the subset rule

We applied our tool to Lucene4.6.0. Table 2 shows the results. Numclass represents the number of classes in which thier equals or hashCode methods are overridden. Subset represents the number of classes satisfying the subset rule. Violation represents the number of classes violating the subset rule.

We discuss about four classes which violate the subset rule. Two of four classes contain a field variable which stores the length of array and it used in only hashCode methods. The length of array can be calculated by the fields variable of array. Also, array is used in both equals and hashCode methods. Then, these classes are not completely violate the subset rule. Although these fields are declared with a keyword “final,” it guarantees that the reference variables refer always the same object, but it does not guarantee that the objects are not changed. Therefore, if the length of array changes, the field variable is not renewed and it does not store the correct value.

One of four class contains a field variable which stores the hash value already calculated for improvement of the performance. This class returns the hash value generated by converting memory address of object to integer value. Since this value does not change at runtime of application, the class does not completely violate the subset rule.

The last one class does not override its equals method, and invokes the equals method of Object class. Equals method of Object class does not use field values. However, this class overrides its hashCode method and it uses a field value. Therefore this class violates the subset rule.

6.2 Evaluation of the equivalence rule

We evaluated about the equivalence rule through the project in practice, HcatFieldSchema class of Apach Hive. This class receives a bug report which states that the class overrides its equals method but does not override its hashCode method at the past revision. This bug is fixed at the later revision. We manually modeled two revisions of this class. One contains the bug and the other fixed the bug. We can conclude our tool correctly work, if the following two conditions are satisfied. 1) Our tool detects that an unfixed class violates the consistency rules. 2) Our tool detects that a fixed class does not violate the consistency rules. Figure 6 shows that the source code of fixed class. This class does not have its parent class. An unfixed class does not override its hashCode method. If the hashCode method of the unfixed class is invoked, the unfixed class invokes the hashCode method of Object. The equals method of this class determines the equivalence of two objects by comparing field values. However, the hashCode method returns true if two objects are the same. Hence, this class violates the equivalence rule. Since the hashCode method of the fixed class returns a hash value by performing arithmetic operations about a field value used in the equals method, the fixed class does not violate the equivalence rule. We check

Table 3: Comparison of execution times

Name	Path length	Path analysis	Pattern procedure	analysis	Execution time
Lucene	16,970	12s	29s	1s	48s
Tomcat	257,590	38s	240s	2s	285s
JFreeChart	3,538,281	11,181s	11,491s	6s	22,689s

Table 4: The number of violated rules

Name	equals method				hashCode method		total
	reflexive	symmetric	transitive	null	subset	equivalence	
Lucene	2	0	0	0	4	1	7
Tomcat	11	3	4	3	14	7	35
JFreeChart	1	1	2	0	76	36	113

the violation of the equivalence rule by Z3. Z3 determines the unfixed class violates the equivalence rule, but the fixed class does not violates the equivalence rule. This result shows that our method can detect the implementation which violates the equivalence rule.

6.3 Execution times

In order to evaluate the cost of checking, we applied our tool to Lucene4.6.0, Tomcat8.0.1 and JFreeChart1.0.17. We compared the execution times. Figure 3 shows the results of this experiment. Path length, name of each step and time represent the total path length of each project, the execution time of each step and total execution time, respectively. Time represents the total execution time.

Theses result show that our proposed method is effective when checks it small or medium size projects. Our method can check large projects by limiting and reducing the search space. Execution time is approximately in proportion to the total pass length. We do not have an obvious answer to the cause of this result. Analyzing this cause is a future work. Also, analyzing procedure of a method and converting Java source code to SMT-LIB model consume over 50% of the total execution time. We can reduce the total execution time by improving the performance of these steps.

6.4 Evaluation of projects in practice

We evaluated how often projects in practice violate the consistency rules. We applied our tool to Lucene4.6.0, Tomcat8.0.1 and JFreeChart1.0.17.

Table 4 shows the results of this experiment. Each name of the rule column represents the number of implementations violating its rule.

We discuss about the cause of the violations of consistency rules. The causes of violating the rules of equals methods are those of [7]. That is, asymmetry null checking, invalid type checking at type hierarchy and miss typing. Also, we model the method invocations for fields as a nondeterministic function, and such modeling may generate wrong models. Three type hierarchies violating the rules caused by the wrong models. This problem can be solved by improving out tool. For example, we can solve this problem by using the information of method behavior from users for the method which is not inlined.

Regarding to the subset rule of hashCode methods, some classes contain a field variable which stores the hash value already calculated for improving the performance. This method returns the hash value generated by converting memory address of the object to an integer value. Since this value does not change at runtime of application, the class does not completely violate the subset rule. Also, regarding to the equivalence rule, there are many classes which override their equals methods but not override their hashCode methods, and violating this rule. This violation is only in JFreeChart and the other two projects do not contain such violation. Therefore, the policy of implementation of the project may affect this result. Consequently we claim that projects policy must contain the rule that if a class overrides the equals methods, then the class must override the hashCode methods. Also, there are two classes violate the equivalence rule of the hashCode methods. It is caused by their equals methods which violate the consistency rules.

7 Conclusion

In this paper, we proposed a method that verifies the consistency between both equals and hashCode methods. Also we have evaluated our method by experiments. Our method analyzes Java source code, and converts these code to SMT-LIB. Our method verifies whether the source code violates the consistency rules by using Z3. If they violate any of consistency rules, our method is able to output counter examples. Experimental results show that our method detects that some of real code includes a wrong method implementation which violates some of the consistency rules.

We will implement the functions which are not implemented our tool yet Also, we will evaluate the performance of our tool by applying our tool to many projects in practice. Experimental result shows that our method detects the inconsistency of some project, but does not shows that how many projects can be checked by our tool. We will apply our method to many projects and reveal it. These are future works.

Acknowledgments

This work is partially being conducted as Grant-in-Aid for Scientific Research S (25220003)

REFERENCES

[1] J. Bloch, “Effective Java,” Addison-Wesley, 2008.

[2] Oracle, “Java Platform, Standard Edition 7 API Specification,” 2013. http://docs.oracle.com/javase/7/docs/api/.

[3] D. Hovemeyer and W. Pugh, “Finding bugs is easy,” ACM SIGPLAN Notices Homepage archive, pp.92-106, 2004.

[4] M. Vaziri, F. Tip, S. Fink, and J. Dolby, “Declarative Ob- ject Identity Using Relation Types,” Proceedings of the 21st European Conference on Object-Oriented Program- ming, pp.54-78, 2007.

[5] C.R. Rupakheti and D. Hou, “An Empirical Study of the Design and Implementation of Object Equality in Java,” Proceedings of the 2008 conference of the center for ad- vanced studies on collaborative research: meeting of minds, pp.111-125, 2008.

[6] C.R. Rupakheti and D. Hou, "An Abstraction-Oriented, Path-Based Approach for Analyzing Object Equality in Java," Proceedings of the 17th Working Conference on Re- verse Engineering, pp.205-214, 2010.

[7] C.R. Rupakheti and D. Hou, "Finding Errors from Reverse- Engineered Equality Models using a Constraint Solver," Proceedings of the 28th IEEE International Conference on Software Maintenance, pp.77-86, 2012.

[8] L. deMoura and N. Bjorner, "Z3: An Efficient SMT Solver," Proceedings of the 14th international conference on Tools and algorithms for the construction and analysis of systems, pp.337-340, 2008.

[9] Clark Barrett, Aaron Stump and Cesare Tinelli, "The SMT-LIB Standard Version 2.0," 2010.

[10] D. Rayside, Z. Benjamin, R. Singh, J.P. Near, A. Milice-vic, and D. Jackson, "Equality and Hashing for (almost) Free: Generating Implementations from Abstraction Functions," Proceedings of the 31st International Con-ference on Software Engineering,, pp.342-352, 2009.

[11] N. Grech, J. Rathke, and B. Fischer, "JEqualityGen: Gen- erating Equality and Hashing Methods," Proceedings of the ninth international conference on Generative programming and component engineering, pp.177-186, 2010.

[12] T. Jensen, F. Kirchner, and D. Pichardie, "Secure the clones: Static enforcement of policies for secure object copy- ing," Proceedings of the 20th European conference on Pro- gramming languages and systems: part of the joint European conferences on theory and practice of software, pp.317- 337, 2010.

[13] K. Anastasakis, B. Bordbar, G. Georg, and I. Ray, "UML2Alloy: A Challenging Model Transformation," Proceedings of the ACM/IEEE 10th International Conference on Model Driven Engineering Languages and Systems, pp.436-450, 2007.

[14] T. Liu, M. Nagel, and M. Taghdiri, "Bounded Program Verification using an SMT Solver: A Case Study," Proceedings of the 5th International Conference on Software Testing, Verification and Validation, pp.101-110, 2012.

[15] I.P. Gent, C. Jefferson, and I. Miguel, "Minion: A Fast, Scalable, Constraint Solver," Proceedings of the 17th European Conference on Artificial Intelligence, pp.98-102, 2006.

[16] D. Balasubramaniam, C. Jefferson, L. Kotthoff, I. Miguel, and P. Nightingale, "An Automated Approach to Generating Efficient Constraint Solvers," Proceedings of the 20129oiokpjg International Conference on Software Engineering, pp.661-671, 2012.

[17] L. Burdy, Y. Cheon, D.R. Cok, M.D. Ernst, J.R. Kiniry, G.T. Leavens, K.R.M. Leino, and E. Poll, "An overview of JML tools and applications," International Journal on Software Tools for Technology Transfer, pp.212-232, 2005.

[18] Apache, "Apache PDFBox - A Java PDF Library," 2012. http://pdfbox.apache.org/.

[19] R. Vallee-Rai, L. Hendren, V. Sundaresan, P. Lam, E. Gagnon, and P. Co, "Soot a Java Optimization Framework," Proceedings of the 1999 conference of the Centre for Advanced Studies on Collaborative research, pp.125-135, 1999.

A study of the Product Development Process through Strengthened Fundamental R&D
- Based on a Case Study of Mobile Phone Businesses for Senior Users -

Tatsuo TOMITA*, Yoichiro IGARASHI*, Masao YAMASAWA**, Kaoru CHUJO**,
Masayuki KATO*, Ichiro IIDA* and Hiroshi MINENO***

*Fujitsu Laboratories Ltd., **Fujitsu Limited, *** Shizuoka University
{tomita.tatsuo, y-igarashi, kaoru.chujo, mkato, iida.ichirou}@jp.fujitsu.com,
m.yamasawa@cybersoken.com, mineno@inf.shizuoka.ac.jp

Abstract -We study a process that enables the accumulated results of proprietary basic research to be leveraged effectively in in-house development of that company's products, and propose a third synchronization process that links the two existing processes of basic research and product development. This synchronization process was devised based on modeling of undocumented discussions and decision flows observed on the frontlines of the two existing processes (basic research and product development), which are executed under different timelines. The authors also discuss a method for establishing this synchronization process in organizations and putting it into continuous practical use.

Keywords: fundamental research, product development, synchronization process, pilot team, scientific-level

1 INTRODUCTION

Manufacturers today have been exploring methods of new-product-planning using competitive in-house technologies. In the home appliance industry, for example, Japanese manufacturers founded after World War II established their position in the industry by using the strategy of emulating European and American companies in terms of their superior quality in manufacturing systems, while also applying the industriousness of Japanese workers.

After a while, the Japanese manufacturers began gradually shifting their mindset from *learning* to *creating*. This shift in strategy led Japanese manufacturers to focus on advanced research and development (R&D) activities, which resulted in a steady flow of attractive new products into the consumer market.

Economic prosperity led the Japanese economy to the stage of higher worker salaries. This reduced the cost-effectiveness of manufacturing somewhat, but Japanese companies continued their efforts by implementing cost-reduction strategies to balance out their competitiveness in the worldwide market.

At this point, the pace of development in the information technology field had been accelerating rapidly, and these changes led the manufacturers to start including business models in their product-dependent businesses. This changing situation led to great advances in the ways that products were developed. In particular, the importance of software has dramatically increased, and major functions that had once been based on hardware are now largely implemented with software.

Unfortunately, those changing trends led to a time of confusion for some companies. Moreover, some manufacturing companies turned their attention away from fundamental R&D to application research that contributes directly to commercialization.

Companies also began to promote various types of cost-reduction strategies; these strategies started in factories and were expanded to other departments such as the product development or administration departments. Under these circumstances, the model cycles have been gradually shortened, and this change led people to misunderstand the trend as evidence of improved R&D processes.

In fact, third-party technologies can be effective in developing new products. However, the trend of depending on a faster model cycle can lead to difficulties for companies that do not have a lot of experience with such cycles. Moreover, some of the resulting products lost their unique competitiveness. This situation narrowed the factors leading to success in the competitive environment down to lower production costs.

The objective of the authors is to take a second look at the role of fundamental R&D activities within manufacturing companies. To be more concrete, we should define and build a complementary process in which fundamental R&D is synchronized with product development.

However, there are great gaps between technologies and products. There are also some classic issues in technology management. One is known as the “Devil's River” [1]; this refers to the gap between the fundamental R&D and the product development processes. The “Valley of Death” [1] is another issue that lies between product development and commercialization. From the viewpoint of the marketing department, manufacturers must overcome the “chasm [2]” to reach their mainstream segment.

The authors set up a hypothesis that if we could build a mechanism to synchronize the two major processes of fundamental R&D and product development, the ideas that were generalized as a result could become the new basis of product development.

This paper discusses a new synchronization process based on the hypothesis; it discusses how to fill gaps between fundamental R&D and product development.

The rest of the paper is as follows. Chapter 2 describes existing product development processes. Chapter 3 extracts the success factors based on two case studies of mobile phone development. Finally, chapter 4 defines the proposed process and validates its consistency using the two case studies introduced earlier.

2 GAPS BETWEEN BASIC RESEARCH AND PRODUCT DEVELOPMENT

This chapter describes the current status of development models and the hidden problems.

2.1 Increased Complication in R&D Processes

Product development processes, especially for information communication technology (ICT) products, have become increasingly complex. In particular, the affordable development period for fundamental R&D has been becoming gradually shorter [3]. In contrast to the early days of this industry, it would be too late for a company to start and complete the R&D process in the required period if this process was started upon receiving a request for a product from a business unit.

Takeuchi and Nonaka [4] observed several Japanese manufacturing companies and found that they had been changing their product development style from "sequential" to "overlapping."

In practical situations and in standard specifications, manufacturers use a certain number of third-party components at affordable prices. However, it would be harder for manufacturers to develop a unique competitiveness using only third-party technologies.

Furthermore, many changes in the development process may be necessary, depending on the strategies used by competitors. Such changes may not only include an earlier completion deadline but also frequent specification changes. If the R&D teams are not able to keep up with the turbulent processes, they will lose a business opportunity in the next model cycle. More than ever before, the current R&D processes must be able to contend with such an uncertain situation.

2.2 Discussion on Time Scale

To highlight the current manufacturing situation, the authors observed the problems that exist in the actual management of technologies using classical process models. However, the observation results indicated that those reference models were not suitable for describing the authors' intentions. The details are as follows.

The major innovation process models that describe the relationship between R&D and product development are the *linear model* and the *linkage model*. [5] The former is a classic R&D model, which represents processes to be connected to process components in series. The "linkage model" [6]-[8] was proposed in the 1990s to improve the deficiencies of the linear model by adding the actual process in manufacturing fields.

According to the authors' observations, the core elements of the two models are based on a common concept; two of the major differences are the starting point of each process and the feedback of know-how flowing across the processes, even in the backward direction, to upstream processes such as the R&D process. However, in the practical planning and development phases, the primary requirement is to ensure sufficient time-to-market, which means the required time span to achieve commercialization. Thus, it is essential to observe the process in terms of management on a time scale. These existing models, however, lack the information flow or mechanism to converge the development processes into a concrete goal. Therefore, the authors carried out the study in order to prove the need for synchronization of timing to establish a concrete launch date.

In the early 1990s, when the two models previously mentioned were proposed, competition in the high-tech industry was not as fierce as it is today. In addition, widespread access to the Internet by consumers could be one factor that has exacerbated the competition. Fig. 1 illustrates the situation of the general relationship between fundamental R&D and product development.

In the consumer products category, the typical model cycle of Fujitsu's mobile phones takes approximately six months to one year. On the contrary, a study reported that the

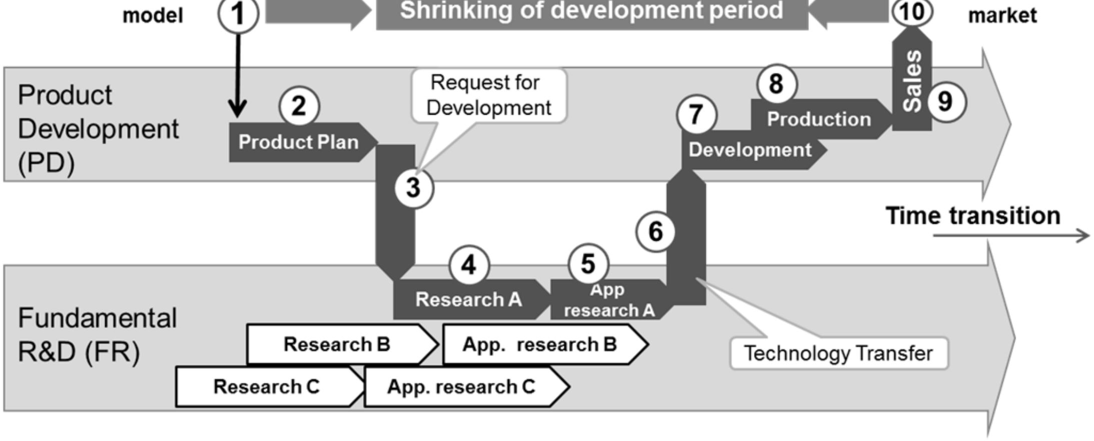


Fig. 1 Fundamental R&D and Product Development Processes

average period of R&D for a typical product in the electronics sector in Japan is 17.8 years [9]. This means that the time spent on R&D processes could be a severe burden or even a fatal factor in the entire cycle of product development. In this situation, manufacturers would probably decide to reform their organizational structure and adopt an accelerated product cycle in order to maintain competitiveness. The simplest solution would be to integrate the R&D department into the business unit. However, this kind of organizational reformation would lead to a decline in R&D activities, and finally, a collapse of the R&D process. Ultimately, the internal source of competitive technologies would be lost.

Another major discussion issue regarding the outcome of R&D is how to achieve product commercialization in the appropriate timing. This is described in related work in this area as *corporate technology stock (CTS) models* [10], [11]. Technology stock refers to all of the knowledge that exists in the R&D department such as technology documents, patents, and implicit knowledge. The value of technology stock is assumed to decrease with time. However, the authors surmise that the value of technologies accumulated through basic research could be maintained and increased by finding promising applications. The authors believe there is a mechanism to control the value on a time scale. In fact, discussions on time scales for manufacturing processes have recently been coming to light.

2.3 Issues in Current R&D: Discussions on Time Scale

In the product development process, items (4) and (5) in Fig. 1 should strongly focus on the fundamental R&D to maintain product competitiveness. However, the older-style organizational structures do not have the flexibility when it comes to the issue of how to provide the business units with expected technologies. A general approach to solving this problem involves two aspects, which must be well balanced. One is maintaining the independence of the R&D and business units; the other is determining the launch date and the expected quality of the product. The product development process should be carried out considering the issues listed below.

(1a) Business units should conduct a thorough investigation to screen technologies applicable to their products. This process should be conducted for both in-house and third-party technologies. An early start of the investigation would make it possible to obtain good results.

(1b) Promising technologies should be applied to multiple products with as few modifications or as little tuning as possible. Typical technologies are not fully ready to be adapted to new applications on demand. Therefore, organization-wide efforts to maintain the technologies are very important.

Measures for R&D departments are described below.

(2a) Product planning teams in business units expect the R&D teams to provide scientific knowledge. This should be done by not only following the classic *waterfall* style (Fig. 1), but also by exploring business opportunities where the technologies were not assumed in the planning stage. This sort of contingency to keep the possibilities open for different types of applications would increase the possibilities for discovering and applying innovations.

(2b) Applying rapid adaptation process to different fields
Typical research outcomes are tailored to particular applications that are described in the planning phase. Therefore, converting those outcomes so that they are adapted for different products requires a great effort to overcome the boundaries in technical and timing issues. Also, any issues concerning organizational structures and operation must be resolved in order to find new application areas where the technologies can be applied.

(3a) Issues in identifying applicable technologies
The R&D department should set up an information pool for “Research A, B, and C” (Fig. 1), in which the outcomes of R&D are stored and maintained.

Managing technical information using electronic media and autonomous management technologies will not cover the necessary functions to arrange matching of technology seeds and products beyond examples. Knowledge management of humans (research) is therefore essential.

(3b) Problems behind application of technologies
There are considerable obstacles in applying technologies in different areas. After the research process has reached the “Research (4)” or later phases in Fig. 1, conversion becomes different from earlier stages.

The authors have analyzed the above descriptions in order to outline the policy for considering new processes.

2.4 Applicable Strategies

The proposed strategies for product development departments are as follows.

(1a) Establish a protocol to discover promising technologies from a broad range of sources, and apply the technology to the product as a new application. To do this, processes to train engineers or researchers who can introduce new technologies and create a new value proposition for the products should be designed and implemented.

(1b) R&D teams should plan distinctive research themes based on expected emergent needs as quickly as possible. Precise market forecasting of mid- to long-term trends makes it possible to create practical research plans that may lead to marketing success. However, such decisions in practical management sometimes result in unintended strategies that reflect an estimated investment effect or an individual decision to set priorities among research themes.

In the fundamental R&D department, the following policies are applicable.

(2a) Adaptation process with the minimum impact:
In business units, the ideal goal of utilizing fundamental R&D is to reach the final stage of product development, where the technologies are completely adapted or tuned to the target products. From the perspective of R&D, a better strategy is to define an "intermediate stage" in the research processes and accumulate the R&D results. This new mechanism will make it easy to meet the requirements for adaptation even if a research theme is underway and not complete.

Finally, there is one strategy that is independent of any existing organizations.

(3a) Capabilities to combine technologies with unknown applications
New ideas for products and businesses are not produced using only the above-mentioned mechanism. Every researcher and engineer should have the ability to manage uncertainties in the project and to explore new combinations of technology seeds and markets.

3 CASE STUDIES

This chapter introduces two examples of product development involving cell phones designed for senior customers.

3.1 Background

At Fujitsu, the development of cell phones specialized for senior citizens started around 1998. The marketing department raised questions about entering this kind of market, since there was no significant marketing information to support the plan. The core members of the planning team convinced the others that there would be substantial market expansion. At that time, however, the cell phone market for younger generations was growing rapidly. People opposed to the idea of only developing phones for younger generations insisted that they should plan products for senior users by composing a product "subset" with essential components taken from the mainstream line for younger customers with a minimum impact on the development cost.

The product-planning team conducted an in-depth study on the market segmentation. These detailed processes are described in [12]. As a result of their discussions, they selected three core concepts to focus on while developing a phone for senior users: ease of listening, ease of looking at (e.g., numbers and letters are easy to see), and security (for personal health and safety, or other factors). These core concepts have been selected for use over the long term in the new category of cell phones for senior users.
In reference to the above core concepts, the following sections trace the development history of two key technologies: motion-sensing and voice signal processing [13].

3.2 Case 1: Motion-sensing Technology

The motion-sensing technology used in mobile phone handsets is designed to help digitize information on a user's activities, for example, exercises, sporting activities such as golf or running, or information beneficial to personal healthcare applications [14]. The original concept of this technology was created for application to HOAP-1, Fujitsu's humanoid robot released in 2001 [15]. The unique algorithm that supervises the motion of the humanoid robot was composed of self-learning algorithms inspired from biological nervous systems that could be described using mathematical models. The following sections describe the development history of motion-sensing technology in three stages.

Stage 1: Unfound needs and seeds
The first generation of the *Raku-Raku* phone was equipped with a pedometer. However, the pedometer function was based on a third-party company's technology. In this situation, the R&D team did not have a chance to meet the needs of the business units since the researchers were concentrating on commercialization of the humanoid robot.

Stage 2: Commercialization of in-house pedometer
The R&D team created a prototype pedometer customized for mobile phones. The business unit evaluated the prototype and approved it for commercialization in 2006. The commercial success of Raku-Raku phone [16], [17] was achieved because the different workflows (of the R&D and business units) were organized to obtain the appropriate timing for the product release. The R&D team was seeking business opportunities other than in the robotics industry, and the business unit was focusing on a new strategy to expand the cell-phone market. The most significant step that resulted in the commercialization was an idea the R&D team had. They redefined the original role of the algorithms (for the sensing motion of robots) into "sensors that identify human behavior."

The main factor for the R&D team was accumulating knowledge of this technology at the *generalized* level, which means understanding the technology through the fundamental scientific basis that led to the technology. At the same time, the business unit was investigating in-house technologies that were applicable to a pedometer on a cell phone. However, the true intent of the business unit was to reduce costs by introducing in-house technologies. In the product development process in the business unit, the R&D team worked on their original process to tune the algorithm for the pedometer. The team completed the process within a month, an exceptionally short period, using their tuning technique. This achievement was outstanding in terms of the development of a new function, which often contains many uncertainties.

In stage 2, the following factors led to a successful development.

- A common goal of the two organizations emerged: commercialization of an in-house pedometer
- The R&D team cultivated an outstanding ability to apply a new technical challenge to new categories of products that they did not have experience in.
- The tuning technology accelerated the time-to-market and resulted in additional value for the business unit.

Stage 3: Expansion of product lineup

The broad utilities of the motion-sensing technology had come to the attention of the business unit during the process of developing the new cell phone. The R&D team started to upgrade the technology in their efforts to develop devices that could monitor human activity. The new activity monitor contained a gyroscope and an accelerometer, which enabled detection of human activity (e.g., walking, jumping) with the cell phone. The outcome of the development was demonstrated in a prototype cell phone displayed at a tradeshow in around 2008. In this demo, the prototype synchronized the motion of animation with a character in the virtual space of an application.

Next, the business unit initiated the development of new and high-value-added applications utilizing this technology in a shorter period than expected. These applications were based on the original tuning feature, which enables algorithm programmers to develop algorithms for new applications without requiring specialized technical or scientific knowledge. The motion-sensing technology was adaptable to particular sports that have a certain pattern of motions such as golf, walking, and running.

In stage 3, the following factors contributed to success.

- The supplemental technologies helped in customizing the advanced motion-sensing algorithm.
- The principle of the technology as described in advanced and complicated mathematical expressions was understood and shared with other team members by a key person in the business unit. This in-depth sharing of information accelerated the commercialization of the pedometer.

The achievement of this technology was the creation of the new function category of *motion sensing* in cell phones.

In 2012, Fujitsu released a new pedometer for dogs, which was achieved with the motion-sensing technology [18]. The applicable industries for this technology have been increasing.

3.3 Case 2: Voice-processing Technologies

This section describes the voice-emphasis technologies that are essential for increasing the clarity of the voice in conversations conducted on cell phones.

Stage 1: Start of development in a virtual organization

Fujitsu Laboratories has been developing voice/audio processing technologies since the 1980s.

The mobile phone business unit decided to plan a new model cell phone with competitive voice-emphasis features. The decision was prompted by the fact that the business unit was able to grasp the progress of R&D themes in Fujitsu Laboratories in terms of completion rate and expected commercialization timing. That information-sharing simplified the decision. The business unit had been organized under policies that encouraged the use of in-house technologies in their products. Those policies had continued in subsequent generations of mobile phones. That atmosphere fostered a strong relationship between the R&D department and the business unit.

The development team was organized into a cross-functional *pilot team*, which consisted of members from both the R&D team and the business unit. The pilot team discussed voice quality on cell phones based on concrete data of the measured frequency response for each model. The practical atmosphere encouraged positive and creative discussions, which resulted in new technology solutions. Through the results of discussions, the pilot team finalized the specifications for the voice emphasizing functions by combining a voice codec, digital filter, and voice signal modeling, and then commercialized the new functions that adjusted the phone’s volume adaptively to the surrounding noise.

The factors that led to success in this stage are as follows.

- The information on technologies was shared across the organizations at the generalized (scientific) level. This accelerated the coordination of the specifications.
- The “clear voice” function was planned by selecting and combining information from the in-house technology pool.

Stage 2: Adaptation of basic research to commercialization

The first generation of the voice emphasis function was introduced in the Raku-Raku phone III, released in 2003. The R&D team that joined the product development process grasped the requirements for the technologies for the next-generation model. These cross-sectional activities gave the R&D team an opportunity to join the product planning phases in the business units.

The next-generation Raku-Raku phone released in 2007 added a new function for automatic adjustment of the receiver volume. The model in 2008 was equipped with adaptive volume control in noisy environments with improved real-time processing performance. The model in 2009 improved the tolerance to non-stationary (bubble) noise such as the type that occurs in human speech.

The R&D team continued in this aspect into the final stage of product development. In the later stages, the new voice-processing algorithm required tuning to enable digital signal processors (DSPs) to be installed in cell phones, which require expertise to get past the constraints of hardware such as the affordable memory usage, programming steps, and power consumption. The R&D team conducted these tuning processes and was able to create a concrete image of the next models in the early R&D process. In other words, the R&D team was able to "synchronize" their efforts to the time scale of the business unit.

In the later stages, the R&D department and the business unit set up an official meeting based on the activities of the pilot team.

The success factors in stage 2 were as follows.

- The participation of the R&D team in the pilot team enabled an earlier start time of the R&D process.
- The R&D team collected feedback from the business unit in the pilot team. This feedback was used to plan the next model cycle.
- The unique organizational management in the official collaboration process between R&D departments and business units.

3.4 Summary of the Success Factors

This section summarizes the factors that led to the successful development.

The first key was the *generalized technologies*. Technological knowledge is accumulated and stored and is independent of particular applications. The generalized technologies were easily diffused and understood across departments, which have different areas of expertise. The second key was the utilization of pilot teams. Several departments began collaborations from within the pilot team to achieve in-depth sharing and understanding of both technologies and markets. The third key was the flexible management practices of Fujitsu Laboratories. There were a few key persons among the researchers who had advanced abilities and knowledge of the technologies, as well as the connections, action, foresight, and capability to find undiscovered problems. The flexible management contributed to Fujitsu discovering "accidental matches" among technologies and products.

In case 1, the objective of the business unit in adopting the in-house pedometer was to reduce costs. Fortunately, however, the motion-sensing functionalities brought unexpected competitiveness to the product-planning team. In the mobile phone business unit, the motion-sensing technology was a disruptive technology [19].

The next chapter describes the mechanism to convert and accumulate the technologies to be "generalized."

4 SYNCHRONIZATION PROCESS

This section discusses a new process to connect the fundamental R&D and product development.

4.1 The role of the Synchronization Process

On the basis of the case studies described in Section 3, the authors composed a mechanism for determining the success factors of the Raku-Raku phone; this mechanism is shown in Fig. 2. The difference from Fig. 1 is the third process, which synchronizes the information on technologies between the existing R&D processes (FR) and product development (PD).

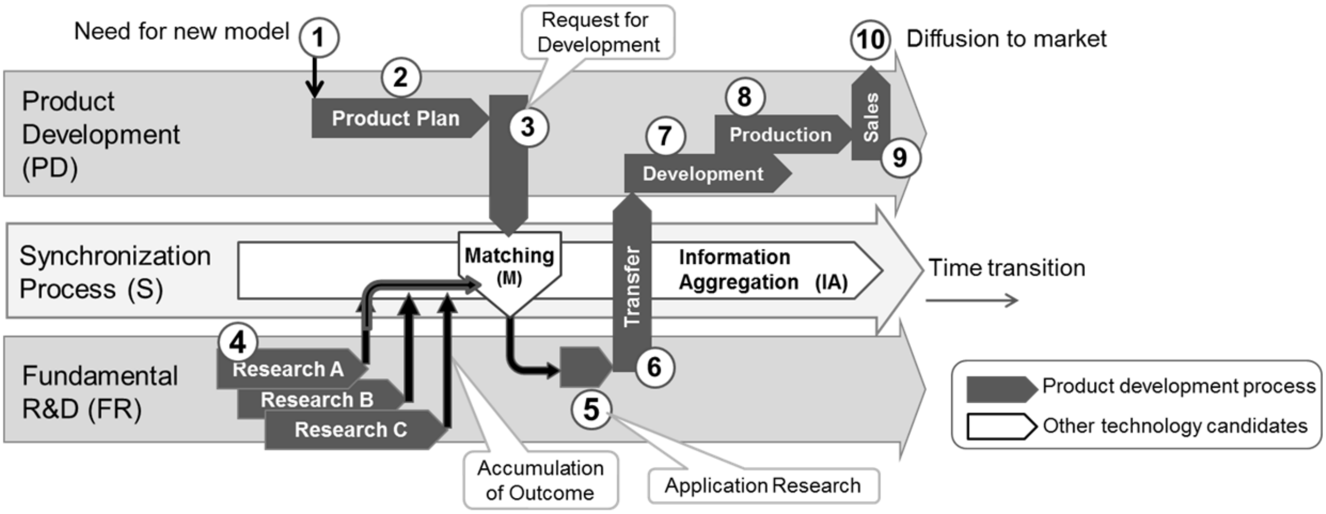


Fig. 2 Core concept of synchronization process

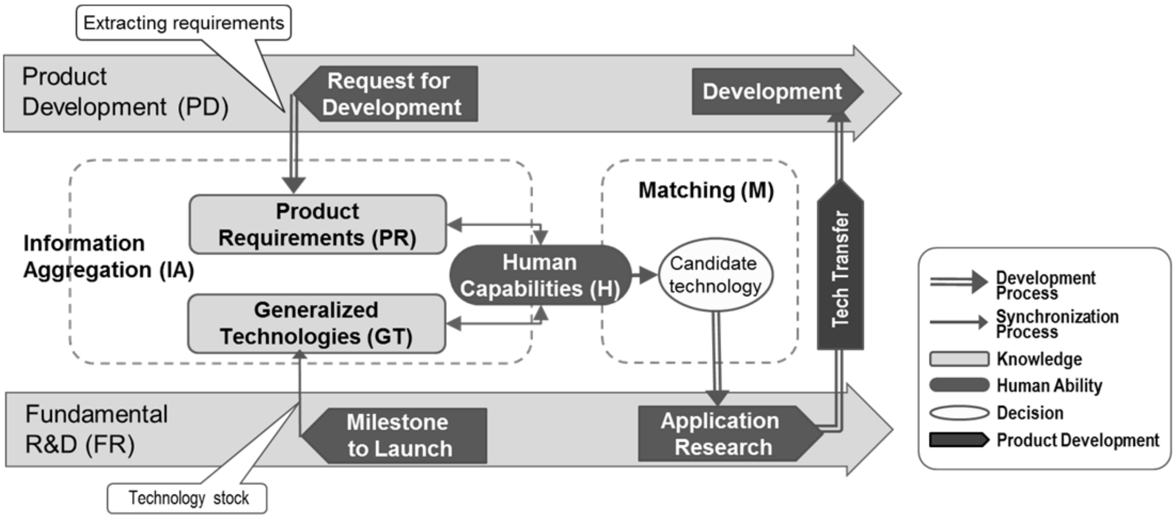


Fig. 3 Internal Components of Synchronization Process

The information aggregation (IA) step is where the R&D outcomes are collected. The fundamental R&D process (FR) consisting of Research A, B, and C is the stage where the progress of each research project and the outcomes of the projects are accumulated.

In Fig. 1, the product development process (PD) is the stage in which a *request for development* is sent to a particular R&D project team that can provide concrete new technologies and realize the planned product.

The synchronization process in Fig. 2 follows different steps. The *development request* (3) is connected to IA. In the IA step, the specifications are received, and *Research A* is identified, which matches the request of the business unit.

Then, in the *application research* (5) step, the algorithm is tuned to fit the target product. If there is no technology to match the request, the business unit would look for third-party technology.

4.2 Initiation of the Synchronization Process

In the case studies, the synchronization processes were initiated without any particular management objective or systematically described procedure. To drive the model described in Fig. 2 in a practical situation, the authors developed an internal construction of the process and operation. This construction is shown in Fig. 3.

Details of the functions described in Fig. 3 are as follows.

- Generalized Technologies (GT): Features of technologies described at a “scientific level”
- Product Requirements (PR): Concrete requirements or specifications for a product in basic research
- Human Capabilities (H): The practical driving force of the “synchronization process,” which includes the capabilities of researchers and engineers, as well as knowhow and implicit knowledge

The generalized technologies (GT) are based on a wide variety of researchers’ knowledge. In other words, (GT) is kept in hot-standby status, ready to be connected with different unplanned applications. The potential applications

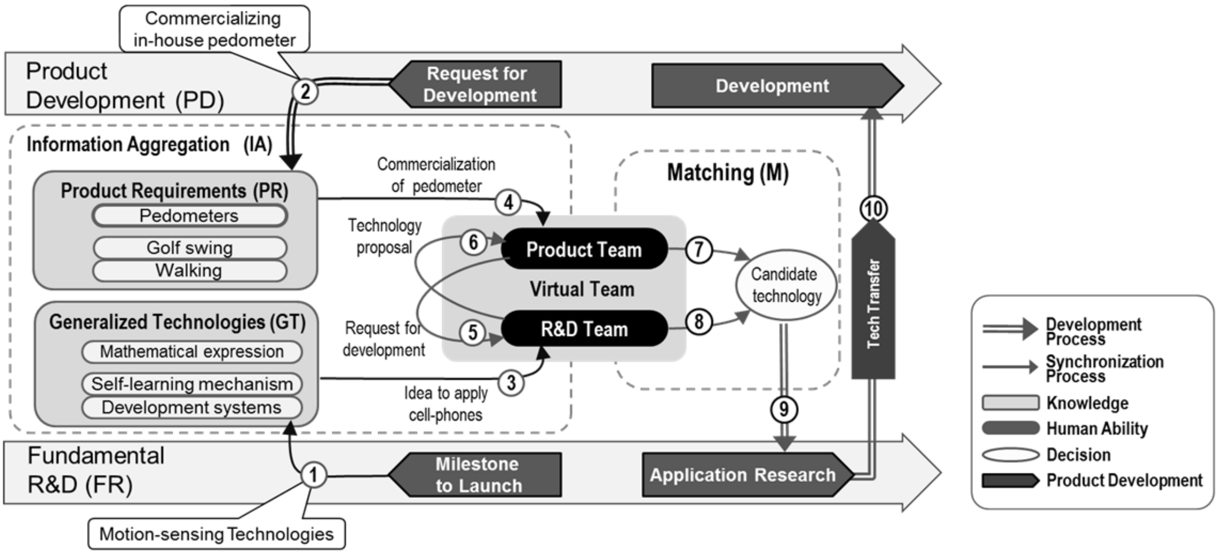


Fig. 4 Synchronization process in motion-sensing technology

in different industries could not be found only by using the technical documents, data, and patents of each R&D outcome.

To overcome the barrier differences of application categories, the (GT) should be described in a common language in order to connect the seeds and needs.

4.3 Operation of the Synchronization Process

This section describes how the synchronization process in Fig. 3 is initiated.

a. Case 1: Motion-sensing technology

Fig. 4 depicts the project history of motion-sensing technology along the components of the synchronization process.

The synchronization process in this case is based on a short term (around five years). Each step is carried out in the following order.

(1) The fundamental R&D process (FR) accumulates the motion-sensing technologies describable at a scientific level into the generalized technologies (GT).

(2) The product development process (PD) registers the request for a cost-reduction strategy into product requirements (PR).

(3) The R&D team selects the appropriate technologies from the GT pool.

(4) The request for the R&D team is composed of concrete specifications.

(5) The PD requests the R&D team to develop the pedometer.

(6) The R&D team completes the development of a prototype and proposes it to the product development team.

(7) The product team sends the final candidate for the product plan to the matching team (M).

(8) The R&D team also sends the candidate technologies to M.

(9) The matching team (M) decides the final product plan and starts the product development.

(10) The fundamental R&D (FR) transfers the completed research outcome to the product development (PD) process.

A summary of the process (Fig. 4) is as follows. In the early stages of the R&D, the primary target was the robotics industry. However, the technology ended up being applied to digital cell phones. The effort of the R&D team in trying to find new application areas was the driving force in converting the technology to the new role.

As a result, this technology was selected for the product plan of the mobile-phone business unit for the new in-house pedometer and was then synchronized to the tight schedule with their sophisticated tuning technology for the motion-sensing algorithm.

b. Case 2: Voice-processing technologies

The next case involves the voice-processing technologies; the development was based on a long-term plan (several decades) in order to fully develop the synchronization process. The process illustrated in Fig. 5 was carried out as follows.

(1) In the fundamental R&D (FR) process, signal-processing technologies were accumulated as Generalized Technologies (GT).

(2) The Product Development (PD) team plans a new voice-emphasis function as a product requirement (PR).

(3) The PD determines the specifications of the new function.

(4) The R&D team selects suitable technologies from the GT to develop the voice-emphasis functions.

(5) The R&D team finishes the plan and sends it to the matching section (M).

(6) The product team finalizes the requirements and sends the finalized specifications to M.

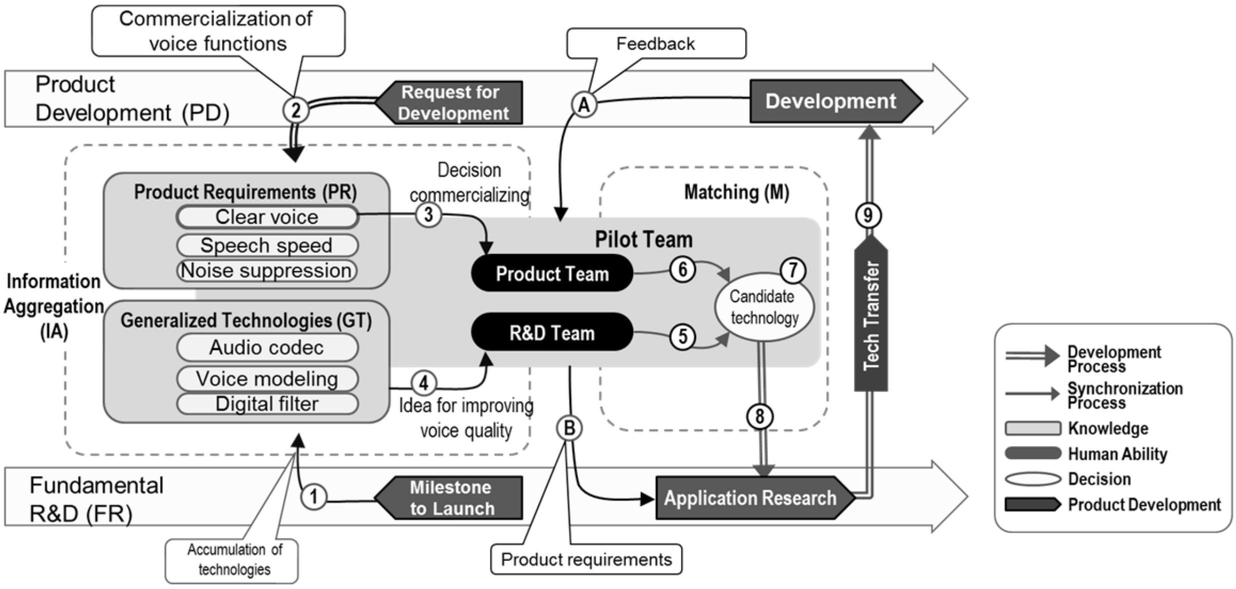


Fig. 5 Synchronization Process: Voice-processing technologies

(7) The matching section (M) finalizes the product specifications of the new voice-emphasis function.
(8) After that, the R&D team conducts application research to adapt the hardware.
(9) The functions ready to be commercialized are transferred to the PD in the business unit.

In addition to the above processes, the pilot team received feedback on the product development process after the technology was transferred (items A and B in Fig. 5).

(A) The knowhow obtained by the R&D team in the tuning stage (8) was able to be used as reference information in next generation products.

(B) The application research phase was improved by utilizing the knowhow obtained in the development of earlier generation products.

The process in Fig. 5 is summarized as follows.

The generalized technologies (GT) consist of the accumulated outcomes of voice emphasis technologies that have been compiled over the long term (since the 1980s). There was implicit knowledge of technologies; the R&D teams in this area were easily able to join the daily discussions as a cross-functional team.

Those cross-organization discussions motivated the researchers to identify new ways to improve the next product plan, which connects the technology to the product requirements. The flexible collaboration between the two organizations—the R&D team and the business unit—shortened the time-to-market of product development compared to the existing situations.

4.4 Discussion

a. Operation of the synchronization process

Fig. 6 illustrates the authors' challenge in implementing the synchronization process into daily management. The most important requirement is to form pilot teams between the departments involved in the project. At this point, the exchange of persons between departments is the basic strategy. In particular, the authors have been implementing the three measures and policies itemized below.

The first item is the rotation of researchers inside the R&D department. Researchers who represent their research areas belong to the strategy department and take part in on-the-job training to gain an overview of a broad range of technologies to give them the ability to match the technologies and potential target products (item 1 in Fig. 6).

The second item is the rotation of researchers and engineers between the R&D department and business units. This measure is aimed at improving the ability to form pilot teams (item 2 in Fig. 6).

The third item is the promotion of flexible management. Self-governing of researchers encourages a novel approach to help them meet their challenges or carry out actions based on the researchers' confidence.

The authors defined a new job title of *Innovation Director* in order to promote the policies that achieve the synchronization process.

b. The role of pilot teams

The two cases described in this paper have some differences in the relationships between departments. However, there is a common point that is initiated in the synchronization process: the pilot teams. In the "voice signal-processing" case, the planning was started within the officially organized pilot team. In contrast, the "motion-sensing technology" case had no official team at the starting point. However, there was a virtual (unofficial) pilot team consisting of two key persons who respectively belonged to the two departments of the R&D and the business unit. They shared in-depth knowledge of the technology. Practical pilot teams do not require physical rooms to communicate; they can do so using online media such as social networking services.

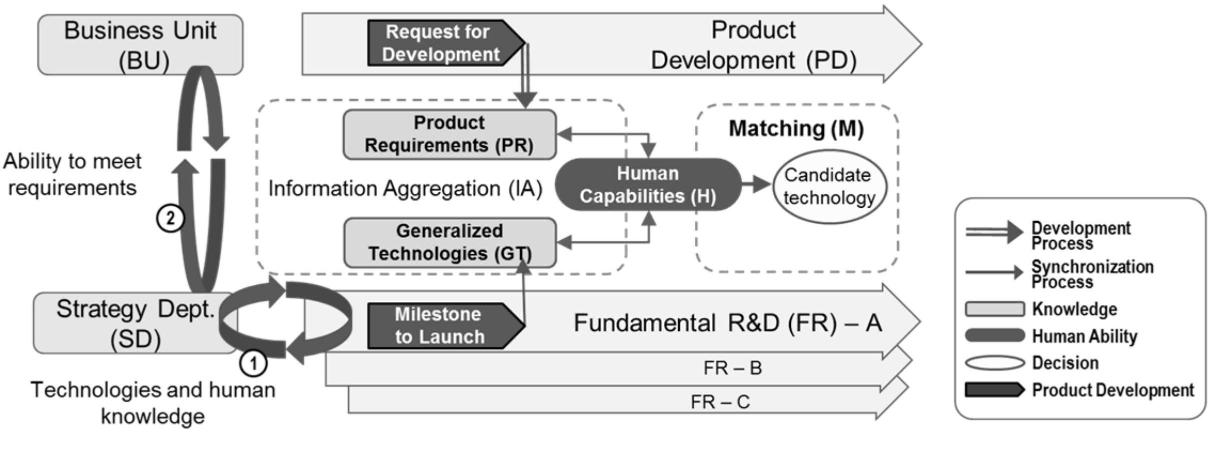


Fig. 6 Synchronization Process: Operation Model

5 CONCLUSION

This paper discussed the classic issue occurring in high-tech industries: how to synchronize the R&D process and the constantly changing product plan of the business unit. In particular, this important issue has roots in the difference in time-scales between R&D and product development.

The authors address this issue by proposing a synchronization process to manage promising candidate technologies and concrete product plans. The latter part described the authors' challenge in upgrading the current R&D department to consolidate the synchronization process. The authors expect that this synchronization process is a continuous process that may enable a company to stay in a competitive position.

Another promising benefit from the synchronization process is that it enables the inclusion of sales and marketing knowledge, which tends to be omitted in discussions in the early stages of R&D. The driving force of this new process is the knowledge of the researchers' technologies from the viewpoint of "scientific levels," not the benefit at an application level.

The scope of this paper is the improvement of the commercialization processes within a company-wide organization. However, this concept is not limited to the user-led communities described in [20]. The authors believe these leading edge technologies can encourage such user communities to initiate innovation processes with the technologies and that the synchronization process described in this paper can contribute to the development of competitive products in a turbulent market.

REFERENCES

[1] "Between Invention and Innovation: An Analysis of Funding for Early-Stage Technology Development," National Institute of Standards and Technology, NIST GCR 02–841, (2002)

[2] G.A. Moore, "Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers Collins Business Essentials," HarperCollins, (2002)

[3] The White Paper on *Monodzukuri* (Manufacturing) 2007, Ministry of Economy, Trade and Industry (METI), (2007)

[4] H. Takeuchi, I. Nonaka, "The new product development game," Harvard business review 64.1: p.137-146, (1986)

[5] B. Godin, "The Linear model of innovation the historical construction of an analytical framework, "Science, Technology & Human Values 31.6: p.639-667, (2006)

[6] S. J. Kline and N. Rosenberg, "An overview of innovation." The positive sum strategy: Harnessing technology for economic growth 275: 305, (1986)

[7] S. J. Kline, "Innovation is not a linear process." Research management 28.4: p.36-45, (1986)

[8] S. J. Kline, "Innovation Styles in Japan and the United States: Cultural Bases: Implications for Competitiveness: the 1989 Thurston Lecture," Stanford University, Department of Mechanical Engineering, Thermosciences Division, (1990)

[9] H. Sakuma, et al, "A study of Research and Development Period in Industry Technologies," The journal of The Development Engineering Society of Japan, 30.1: p.45-52, (2010)

[10] A. Kameoka, S. Takayanagi, "A "Corporate Technology Stock" Model -Determining Total R&D Expenditure and Effective Investment Pattern-," PICMET '97: Portland International Conference on Management and Technology, Publication Year: 1997, p.497 – 500, (1997)

[11] A. Kameoka, S. Takayanagi, "A Corporate Technology Stock Model: Financially Sustainable Research and Technology Development, " PICMET '99. Portland International Conference on Management and Technology, Publication Year: 1999, p.397 – 401, vol.2, (1999)

[12] H. Saso, et al, "A study of Commercialization Process to Enhance Productivity -Based on Case Study of Cellular Phones for Middle-Aged and Elderly People-," TECHNOLOGY And ECONOMY, (2011-5): 43-50, (2011)

[13] K. Hayashida, et al, "Development Concept and Functions of Raku-Raku PHONE," FUJITSU, Vol. 61, No.2, p.184-191, (2010)

[14] K. Chujo, et al. "Human Centric Engine and Its Evolution toward Web Services," FUJITSU Scientific & Technical Journal, 49.2: p.153-159, (2013)

[15] "Fujitsu Introduces Miniature Humanoid Robot, HOAP-1," Fujitsu press release on 10 Sep. 2001, URL: http://pr.fujitsu.com/en/news/2001/09/10.html

[16] "Fujitsu and Orange Partner to Deliver Smartphones to the Rapidly Growing Senior Market in Europe," Fujitsu press release on February 19, 2013, URL: http://www.fujitsu.com/global/about/resources/news/press-releases/2013/0219-02.html

[17] "Fujitsu Special Campaign Celebrating Sales of 20 Million Raku-Raku Phones," Fujitsu press release on September 15, 2011, URL:http://www.fujitsu.com/global/about/resources/news/press-releases/2011/0915-01.html

[18] "Cloud Service Launched for Wandant Dog Pedometer," Fujitsu press release, Nov. 27 2012, URL: http://www.fujitsu.com/global/about/resources/news/press-releases/2012/1127-01.html

[19] C. Christensen, "The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail", Harvard Business Press, (1997)

[20] E. Von Hippel, "Democratizing innovation," MIT press, (2005)

Session 3:
Social Systems
(Chair：Kozo Okano)

Learning Material Recommendation Service with E-Learning Database

Yuji Wada[*], Ryuhei Kurihara[*], Jun Sawamoto[**], Hiroyuki Sato[**] and Hisao Fukuoka[*]

[*]Tokyo Denki University, Japan
[**] Iwate Prefectural University, Japan
[*] yujiwada@mail.dendai.ac.jp

Abstract - Today, online learning content, known as e-learning material, is growing at a fast pace. E-learning allows user information and studies, viewed content and other data to be saved on an e-learning server as log data. By using the log data, the system can provide users with content suited to them. However, a problem occurs with cold starts, in which the system cannot make suitable user recommendations due to insufficient log data. To solve this issue, this paper proposes a new recommendation approach to provide user content based on few data points and attributes. By reporting the experimental data and results, it is verified that this approach is actually beneficial and that the approach can provide useful content to users.

Keywords: recommender system, cold-start problem, e-learning, collaborative learning system

1 INTRODUCTION

Recently, there has been much research on recommendation systems. Recommendation systems save user browsing data in a database to infer which content from the database will be suitable or of interest to the user. The system then recommends or provides such content to the user on subsequent visits or visits from other users.Amazon.com [1] system is an example that infers products of interest from user data and then recommends products. Applying a recommendation system for learning content in e-learning could also provide user-optimized learning content. This would allow it to deepen and accelerate user understanding.

A method for recommending e-learning materials is shown in Figure 1. The recommender works by using log data on what kind of e-learning content has been viewed. Based on the log data, the system selects a recommender with tendencies similar to the recommendee to offer content recommendations.

2 EXISTING RESEARCH AND RESEARCH OBJECTIVES

2.1 Existing Research

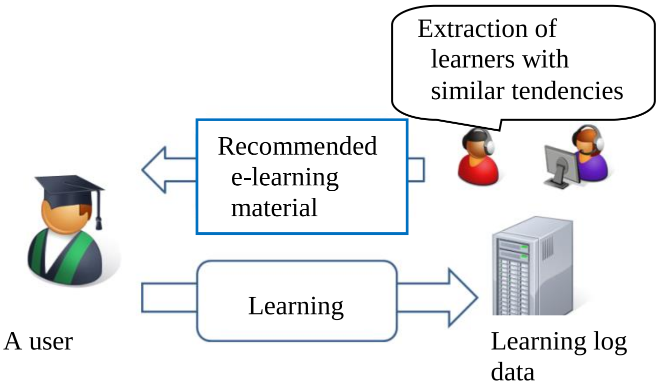

Extraction of learners with similar tendencies
Recommended e-learning material
Learning
A user
Learning log data

Figure 1: E-learning material recommendation.

Recommendations using log data appear in various scenarios [2], [3], [4]. In one particular case, an e-learning recommendation system in actual operation has been effective to a certain extent [5]. However, there are concerns about cold starts, in which a user, such as a new user, without log data on the server accesses the system. In this case, the system cannot offer appropriate content as it cannot identify or determine user preferences through the normal recommendation system.

2.2 Cold Starts

A large issue overall with recommendation systems is what is commonly known as a cold start. As mentioned before, recommendation systems use one of two approaches for making recommendations: one is user-based approach and the other item-based. Some form of user log data is a key element in both of these approaches. Thus, when the system has little or no user history, there may be cases in which the system cannot make appropriate recommendations to the user. This is the cold start problem.

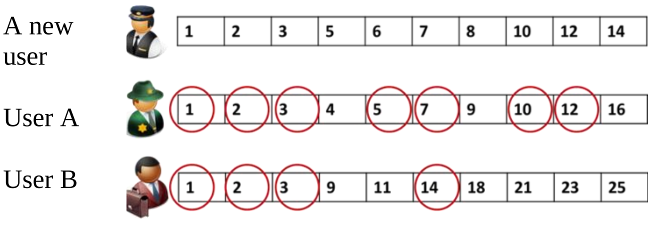


Figure 2: Case with sufficient user history.

Figure 2 gives an example where there is sufficient user history. When the user history is compared to that of User A

This study was made possible through support of the "Research on Knowledge-Infused Technology in Ubiquitous Databases," Grant-in-Aid for Scientific Research Basic Research (C) topic number 24500122.

and User B, there are more matches with User A. Thus, the results show that the user is more similar to User A than to User B.

Figure 3 gives an example of insufficient user history. Based on the information given, when the user history is compared to that of User A and User B, there are three matches for both. Thus, the results show that the user is just as similar to User A as to User B. If the history reaches the point given in Figure 3, the system could make incorrect recommendations until it stops finding User B to be similar.

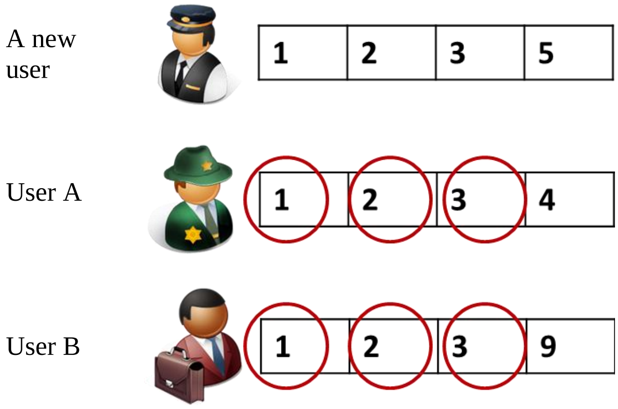


Figure 3: Case with insufficient user history.

2.3 Research Objectives

This study proposes an approach to offer satisfying content optimized for the user with minimal information to combat cold starts and to show whether this approach can actually offer content that users find useful.

First, Proposed Method 1, which uses only a few attributes, is verified through an experiment. Next, the alternative, Proposed Method 2, is verified in similar fashion. Then, another approach is given that solves the problems found in Proposed Method 2

3 PROPOSED METHOD 1

3.1 Overview

When devising the approach of this study, each person's ability was scrutinized first. People have different reasons for using learning materials. One user might be trying to expand their field of expertise, while another is studying about a field they never learned before. Therefore, information on the user's academic major and past studies should be useful in making recommendations.

To demonstrate the effectiveness of this approach, experiments were conducted to verify whether the content the individual has previously studied, as well as their current major, translated to different levels of comprehension for the recommended content.

If a person studying within their major comprehends recommended content better than someone from outside that field, the past learning is an important factor in recommending learning material. Conversely, if the two people comprehend content equally or the person studying outside their major comprehends the content better, the previously studied content is not that important in recommending learning material.

Proposed Method 1 suggests content based on an individual's major attribute.

An explanation of the method progression is given by using Figure 4. First, a user (recommendee) accesses the e-learning material server and registers his information. In the example in Figure 4, the major of the user is information systems. Next, the system collects information on other people in information systems from the usage history on the server and extracts content to offer the user based on this information.

There are a couple advantages to this approach. First, it puts a small load on either the user or server by only requiring the user to provide his major. Second, it allows the system to make inferences and offer content recommendations immediately upon registration.

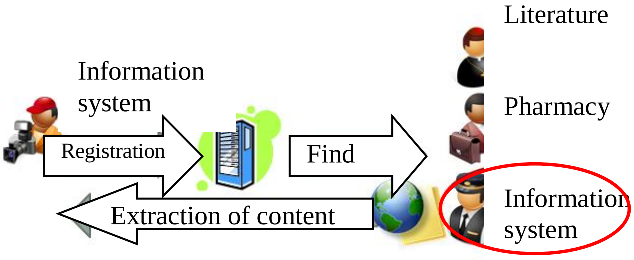


Figure 4: Proposed Method 1.

3.2 Verification Experiment

First, the learning material is prepared with an associated course set. For this exercise, the system was set to recommend relational algebra from the TDU individual review support system named AIRS [6].This content was recommended for the following reasons:

- The database course is not a field studied in primary years
- Little chance to touch on the subject outside of an information systems major
- Does not require using highly specialized knowledge to understand content

Relational algebra covers the nine topics listed below. Figure 5 shows the actual learning material content.

- Selection
- Projection
- Sum
- Intersection
- Difference
- Cartesian product
- Join
- Natural join
- Division

The related major was set to the TDU School of Information Environment.

Test participants were first asked whether this was their major. Next, they studied the learning material for 15 minutes.

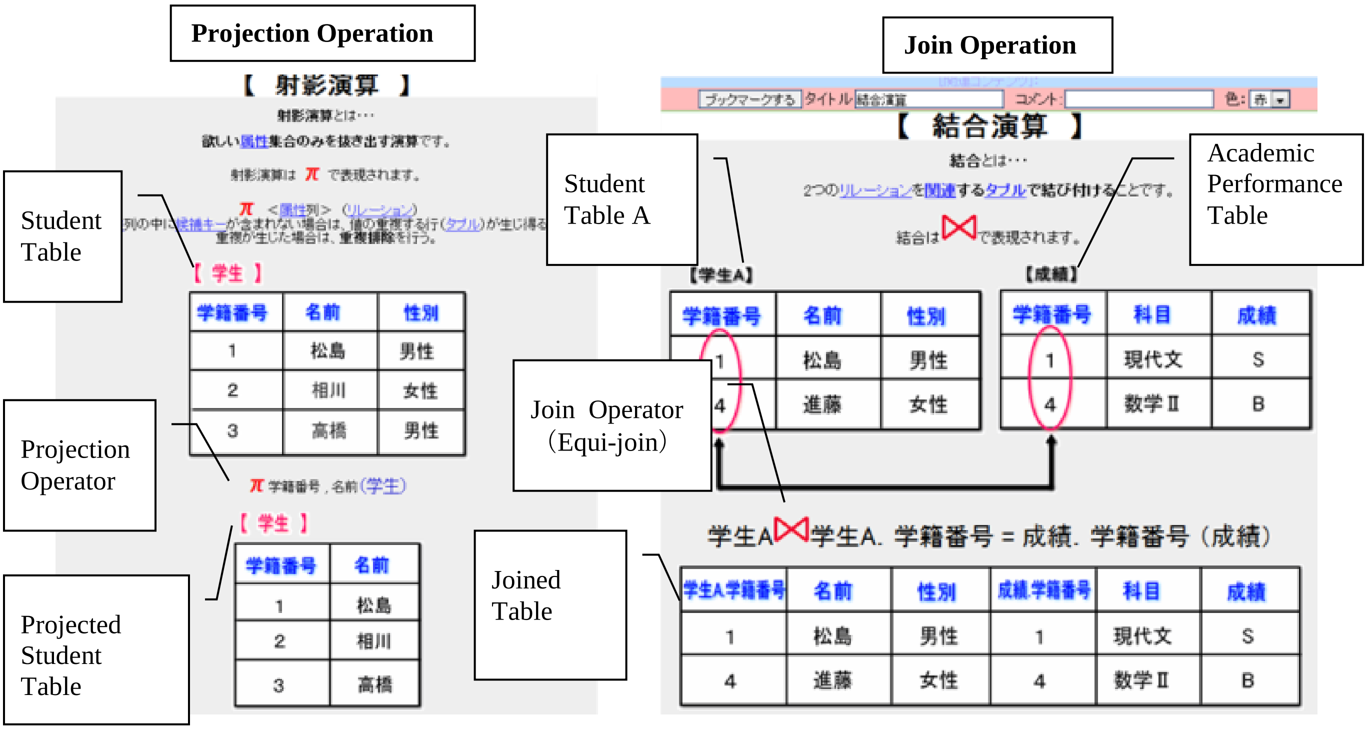
After studying, the test participant took a comprehension check for 25 minutes to gauge whether they correctly understood the learning material. The following two approaches were used in the comprehension check:

(1) The test checks whether the participant can correctly explain the learning material content (topics listed for relational algebra in AIRS: selection, projection, sum, intersection, difference, Cartesian product, join, natural join and division).

(2) Using the learning material studied, participants were asked to create actual problems and give the solutions.

The above methods were used to increase verification reliability [7]. Test participants who did not understand the learning material would have more trouble explaining the topics and creating problems than they would by simply solving prepared problems. Results were scored on a 10-point scale, adding points for correct answers. The main criteria for giving points included correct explanations (2-3 points), explaining with figures (3 points), and so on.

Next, the extent of test participant comprehension on the studied content was measured to compare results between both test participant groups.



【 射影演算 】

射影演算とは・・・

欲しい属性集合のみを抜き出す演算です。

射影演算は π で表現されます。

π <属性列> (リレーション)

列の中に候補キーが含まれない場合は、値の重複する行(タプル)が生じ得る

重複が生じた場合は、重複排除を行う。

【 学生 】

学籍番号	名前	性別
1	松島	男性
2	相川	女性
3	高橋	男性

π 学籍番号，名前(学生)

【 学生 】

学籍番号	名前
1	松島
2	相川
3	高橋

【 結合演算 】

結合とは・・・

2つのリレーションを関連するタプルで結び付けることです。

結合は ⋈ で表現されます。

【学生A】

学籍番号	名前	性別
1	松島	男性
4	進藤	女性

【成績】

学籍番号	科目	成績
1	現代文	S
4	数学Ⅱ	B

学生A⋈学生A. 学籍番号 = 成績. 学籍番号 (成績)

学生A学籍番号	名前	性別	成績.学籍番号	科目	成績
1	松島	男性	1	現代文	S
4	進藤	女性	4	数学Ⅱ	B

Figure 5: Learning material content (partial).

3.3 Results

Results for the verification are given in Table 1. Both those participants not majoring in information environments (Groups A and B) and those majoring in information environments (Groups C, D and E) scored higher than 70% overall in terms of explaining the content correctly. All groups were also able to create problems that made appropriate use of the learning material content within a limited timeframe.

However, looking at the learning topics in thirds, as given in Table 2, all groups gave almost perfect explanations for earlier topics but tended to falter with incomplete explanations for later topics.

Table 1: Verification results

	Explanation of content	Creation of problems	Provision of correct answers
Group A	◎	○	○
Group B	○	○	×
Group C	○	○	○
Group D	○	○	×
Group E	◎	○	○

◎:more than 80%　○:more than 70%
×:not possible

Table 2: Detailed results for study content explanations

	First three	Middle	Last three

	items	three items	items
Group A	◎	○	△
Group B	◎	○	△
Group C	◎	△	×
Group D	○	○	○
Group E	◎	◎	○

◎:more than 80% ○:more than 70%
△:more than 60% ×:less than 60%

Results for the verification showed that explanations for later learning content tended to be insufficient. Impacts from outside and past knowledge require examination.

As before, the test participants study the learning materials and then submit to a comprehension check. However, the additional verification experiment only verifies results for the later learning material topics for which explanations tended to be insufficient: join, natural join, and division. Also, the earlier verification had set time restrictions for studying and the comprehension check. For the additional verification, the time restrictions were lifted so that test participants could study until they said they understood.

In the additional verification, more than 90% of test participants from all groups could explain the material, and they not only created the problems but also gave the solutions.

3.4 Discussion and Conclusion

From the verification results, recommendees are at least capable of understanding the learning material content even if they study the recommended materials, regardless of outside and past knowledge.

Further, from the results of the additional verification, possible reasons for lower comprehension of the latter half of the learning material are as follows: 1) too much learning material was used in the verification, or 2) the study time given was too short for the amount of learning material.

This verification has shown that outside and past knowledge do not have much impact when recommending learning materials. Thus, the conclusion was that Proposed Method 1 was not very useful in recommending learning material to users.

4 PROPOSED METHOD 2

4.1 Overview

Results showed that determining and offering content recommendations based on one's major, as proposed in Proposed Method 1, is not very effective. Thus, another approach was investigated. One characteristic of Proposed Method 1 was that it searched for contents to recommend based on attributes centered on the inherent abilities of the individual. The proposal for this new approach also centers on individual abilities.

One of the possible individual abilities is reading speed. The approach proposed for Proposed Method 2 centers on this attribute of reading speed.

Here is an explanation using Figure 6. First, the user is prompted to view text or content provided from the e-learning material to infer user reading speed. Based on this information, the log data server is searched for others who read at similar speeds and proposes content previously viewed by such people.

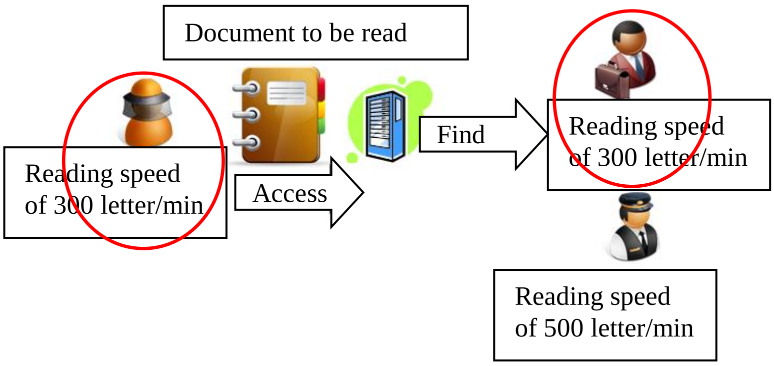

Document to be read
Reading speed of 300 letter/min
Access
Find
Reading speed of 300 letter/min
Reading speed of 500 letter/min

Figure 6: Proposed Method 2 overview.

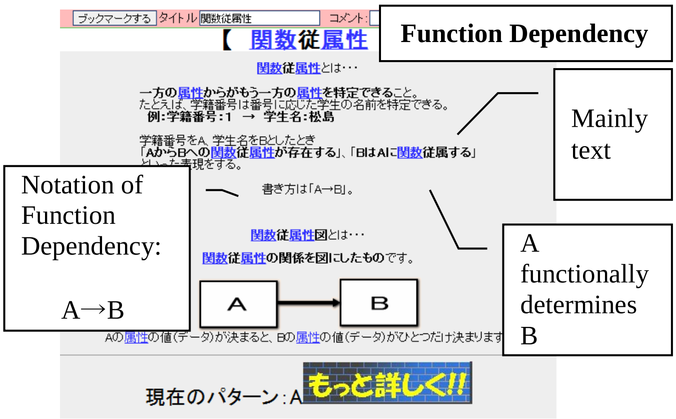
The advantage of this approach is that it can provide content recommendations quickly, as reading speed can roughly be inferred by reading only one page. Also, making the required attribute reading speed eliminates the need for registering personal information. This would allow content recommendations to be provided to temporary guest users.

For Proposed Method 2, experiments were conducted to verify whether there was a connection between reading speed and easily understood content. If a difference can be found in content understanding between fast and slow readers, this approach will be effective.

4.2 Verification Experiment

First, the reading speeds of the test participants were measured. Test participants read three texts of similar length but different content, with the average saved as their reading speed attribute.

Next, they studied for 20 minutes on AIRS. The content studied and content types were as follows:



【 関数従属性

関数従属性とは・・・

一方の属性からがもう一方の属性を特定できること。
たとえば、学籍番号は番号に応じた学生の名前を特定できる。
例:学籍番号:1 → 学生名:松島

学籍番号をA、学生名をBとしたとき
「AからBへの関数従属性が存在する」、「BはAに関数従属する」
といった表現をする。

書き方は「A→B」。

関数従属性図とは・・・

関数従属性の関係を図にしたものです。

Aの属性の値(データ)が決まると、Bの属性の値(データ)がひとつだけ決まります

現在のパターン:A

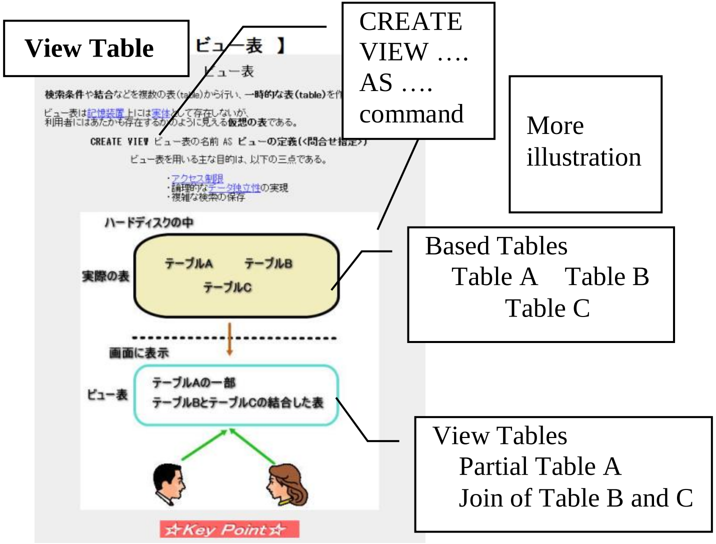
Figure 7: Content Pattern A.

(1) Content with text and readable figures (Pattern A)

As given in Figure 7, the content in Pattern A is mainly composed of text with minimal explanation. Some content included illustrations, but others contained text only.

(2)Content with more detailed figures included (Pattern B)

As given in Figure8, Pattern B had more illustration than Pattern A with more detailed explanations. Almost all Pattern B content shared these features. The condition for selecting Pattern B content for use in the verification experiment was that it had more detailed illustration than Pattern A content.


View Table
CREATE VIEW …. AS …. command
More illustration
Based Tables Table A Table B Table C
View Tables Partial Table A Join of Table B and C
ビュー表 】
ビュー表
ハードディスクの中
実際の表
テーブルA
テーブルB
テーブルC
画面に表示
ビュー表
テーブルAの一部
テーブルBとテーブルCの結合した表
☆Key Point☆

検索条件や結合などを複数の表(table)から行い、一時的な表(table)を作

ビュー表は記憶装置上には実体として存在しないが、
利用者にはあたかも存在するかのように見える仮想の表である。

CREATE VIEW ビュー表の名前 AS ビューの定義(<問合せ指定>)

ビュー表を用いる主な目的は、以下の三点である。

・アクセス制限
・論理的なデータ独立性の実現
・複雑な検索の保存

Figure 8: Content Pattern B.

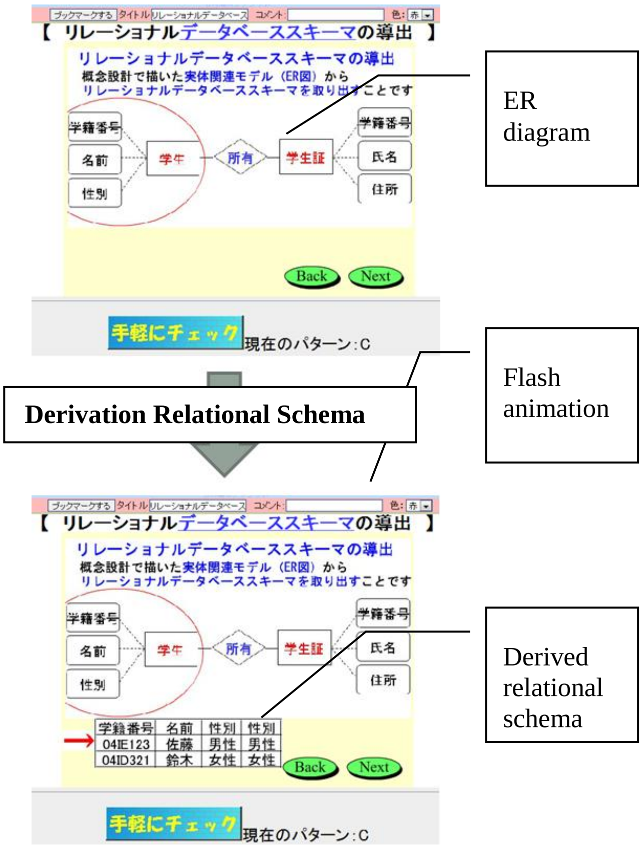
(3)Content with explanations in Flash format (Pattern C)

In Pattern C, content is explained with Flash animation.

As Pattern C content contains animation, it emphasizes viewing more than reading. Also, the animation does not advance automatically, but rather requires the test participant to press the next button to advance, as given in Figure 9. Thus, test participants can advance through the animation at their own pace of understanding without the animation moving on before they understand.

Test participants studied for 20 minutes, and were then given a five minute break. After the break, they were asked to explain the content of the learned material to measure their comprehension level in a comprehension check. As with Verification Experiment 1 (as described in section 3.2), this approach was used to increase reliability, as test participants who did not understand the learning material would have more trouble explaining the topics than they would have by simply solving prepared problems. Also, as in Verification Experiment 1 (as in 3.2), results were scored on a 10-point scale, adding points for correct answers. The main criteria for giving points included correct explanations (2-3 points), explaining with figures (3 points), and so on.

After the comprehension check, reading speed and check results were compared to examine which types of content were well understood at each reading speed and whether reading speed affected which content types were easy to understand.



リレーショナルデータベーススキーマの導出
概念設計で描いた実体関連モデル（ER図）から
リレーショナルデータベーススキーマを取り出すことです

学籍番号	名前	性別	性別
04IE123	佐藤	男性	男性
04ID321	鈴木	女性	女性

Figure 9: Content Pattern C.

4.3 Results

For this experiment, verification involved three groups of test participants: students taking classes in the TDU database (Nos. 1 and 2), a foreign exchange student (No. 3), and students not studying information systems (Nos. 4 and 5).

As can be seen in the results for content A1 in Table 3, the test participants who read 500 or more Japanese characters per minute tended to have a firm understanding of text-intensive content. The test participant who read slower than this did not understand everything in text-intensive content and thus could not explain it well.

Meanwhile, as seen in the results for content C2 in Table 3, all test participants gave explanations for content given in Flash format at roughly similar levels.

Table 3: Verification results

No.	Speed (letter/min)	A1	A2	B1	B2	C1	C2
1	511	10	0	2	6	10	5
2	626	7	2	0	6	3	6
3	309	2	0	5	3	0	5
4	883	8	2	4	0	7	5
5	1610	6	0	2	7	6	4

4.4 Discussion and Conclusion

As shown in the results of Table 3, test participants reading at a speed of 500 to 1,000 Japanese characters per minute had a good understanding of Pattern A content. When recommending content, this segment would not find text-intensive content to be a hindrance. The test participant whose reading speed was less than 500 characters may not have been able to fully understand the content provided for text-intensive content. Thus, text-intensive content should not be recommended to slower readers. On the other end of the spectrum, the test participant whose reading speed was higher than 1,000 had similar results to those reading between 500 and 1,000 characters per minute. Thus, recommending text-intensive content would also probably not hinder fast-reading recommendees. As all the test participants for the verification experiment with reading speeds of 500 characters or higher were native Japanese speakers, using a native language attribute may also help somewhat in deriving content recommendations.

As seen from the results for content C2 in Table 3, all test participants' comprehended content shown in a Flash format, such as in Pattern C, at roughly the same level. Further, given that content with Flash-formatted images scored high overall compared to other patterns in Table 3 and produced steady scores for all test participants, such content appears to be highly effective, regardless of reading speed. A look at C1, however, shows that this pattern is not necessarily true. A great many of the test participants thought that content C1 was poorly illustrated and hard to understand, which likely affected the results.

There were a few content issues, possibly due to the fact that questions were selected at random. After the verification experiment, a great many of the test participants were of the opinion that the first question for Pattern B was hard to understand. Also, results for the second question were likely skewed as the content the question is based upon assumed knowledge on databases.

As can be seen from Table 3, test participant scores showed the same trends for all participant groups. Thus, there were no changes in performance due to differences in major for this verification.

This verification experiment showed that the reading speed attribute could be effective in recommending e-learning material. However, this experiment did not account for the element of individual memory ability. It remains to be verified whether results would be the same if this element was taken into account.

5 IMPROVED APPROACH

5.1 From Results of the Experiment for Proposed Method 2

The verification experiment for Proposed Method 2 did not account for elements such as individual memory ability or IQ. As such, it remained to be verified whether results would be the same if these elements were taken into account. The need had arisen to prove that content could be recommended without accounting for individual memory ability and IQ.

5.2 Background on Improvement

Despite attempts to devise a way to recommend content irrespective of individual ability and IQ, these attributes appear to be equivalent to inherent, innate individual abilities, as with reading speed. Furthermore, the question arises of whether the recommendations will really be useful to users if such attributes are excluded. This led to an improved two-dimensional recommendation approach using attributes for both individual ability and IQ as well as reading speed.

5.3 Verification Experiment

Experiments were conducted to verify the impact that the abilities of “reading speed” and “individual memory ability and IQ” have on comprehension of content recommendations. The experiment also verified what kind of content is preferred by test participants and what content they found easy to understand, regardless of their comprehension level.

Content sections are segmented into a first half and second half. The first half of the content is more basic, and the second half is applied content using knowledge from the basic content.

5.3.1 Overview

First, reading speed was measured. The same method for measuring reading speed was used from the verification experiment in Proposed Method 2: test participants read three texts of similar length but different content, with the average saved as their reading speed attribute.

Next, IQ was measured using an online IQ test. The IQ calculated in the online test is a rough figure, and not necessarily the test participant’s accurate IQ.

Next, the participants studied the learning content for 20 minutes on AIRS. This verification experiment followed two patterns: 1) allowing test participants to view the content of their choice without any set learning content, and 2) having them view the same content as in Proposed Method 2 for a direct comparison of experiment results.

After studying for 20 minutes, test participants were then given a five-minute break. After the break, a comprehension check was performed to verify comprehension levels of the learning content. The comprehension check was the same format as in Verification Experiments 1 and 2 (as described in section 4.2), with test participants explaining the material.

After this, test participants were asked to create problems using the studied content. Given that it is difficult to create problems without understanding the learning content, the problem creation portion was conducted to grasp how well participants understood what they were studying.

Finally, test participants were asked to complete a questionnaire on the content, including which contents were easiest to follow.

5.3.2 Results

The verification results showed that those with higher than average IQs (120+, with the average being 100) had high comprehension for second half content as well as first half content. Many people used content from the first half when creating problems. In the content questionnaire as well, participants responded that content in the first half of the section was easier to understand. The questionnaire also asked about display patterns, and Pattern B was seen as the best.

Table 4: Verification results

No.	Speed (letter /min)	IQ	A	B	C	First half content	Second half content
1	515	104	9	4	5	7	3
2	516	120	7	2	5	7	6
3	510	102	6	4	6	6	3
4	893	126	8	2	5	9	7
5	1590	103	6	3	4	7	3

5.4 Discussion and Conclusion

It was found that people with high IQs could remember and understand broad content, including that in the second half of the section, within a limited time span. Conversely, those with average IQs have trouble remembering or understanding much content in a limited time. Thus, offering the first half of the content when making recommendations could produce uniform results. Given the characteristics of users with no log data, one possibility could be to only offer users basic first half content at first. It could be effective to offer basic content and content on practical applications in quicker succession for persons with high IQs than those with average IQs.

Also, it was found that users see well-illustrated content with text being central as easier to understand and follow than content with too much text or animations. Still, user evaluations of content and actual comprehension results did not always directly correspond. Thus, it would be difficult to determine what content to provide based solely on user evaluations.

It was confirmed that improving Proposed Method 2 would allow the system to make better recommendations to users. However, this approach does have its problems. Unlike reading speed, IQ cannot be measured by using existing content for general e-learning material. This leaves only two options: installing a system for measuring IQ into the e-learning material system, or having users self-assess their own IQ.

6 OVERALL CONCLUSION

The effectiveness of a recommendation system for e-learning was confirmed. Proposed Method 1 was run through verification experiments, but was not found to provide effective recommendations. Based on this, Proposed Method 2 was proposed and run through the same verification experiments. This approach proved to be more effective than Proposed Method 1. Hence, Proposed Method 2 is modestly effective as a new approach for recommending e-learning material. Improving Proposed Method 2 by adding an IQ attribute increased the likelihood of providing users with useful content. Therefore, results show that useful content recommendations can be made to users by using attributes for IQ and reading speed. At the same time, cold starts were re-confirmed as being a great hindrance to recommendation systems, and result in lower rates of useful content provided with fewer attributes.

The verification experiment for the present study was small in scale with few test participants. With no validation data for large numbers of test participants, conducting the same verification experiment with more test participants could produce different results.

REFERENCES

[1] Amazon.com, http://www.amazon.com/

[2]T.Ono, H.Asou, and Y.Honmura, Technologies and research issues for recommendation, IEICE Vol.94, No.4, pp.310-315(2011-04).

[3]Y.Tarui, Recommendation System of Tourist Site Using Collaborative Filtering Method and Contents Analysis Method, Journal of Faculty of Management Information Sciences Jobu University, Vol.36, pp.1-14(2011-12).

[4]Y.Matsubara, T.Nagata, and S.Tamaki, Proposal of the Web Search Technique for Lessons of the Elementary School Education using Collaborative Filtering, IPSJ SIG-CE 2007(101), pp. 69-74(2007-10-05).

[5]Y.Wada, S.Matsuzawa, M.Yamaguchi, and S.Dohi, Bidirectional Recommendation Technology for Web Digital Texts, Journal of Digital Information Management, Vol. 8, No. 4(August 2010).

[6]Y.Wada, Y.Hamadume, S.Dohi, and J.Sawamoto, Technology for Recommending Optimum Learning Texts Based on Data Mining of Learning Historical Data, International Journal of Information Society IJIS, Vol.2, No.3, pp.78-87(2011).

[7]A.Inoue, Verification of learning effect in PBL information education, IPSJ SIG-IS 2007(25), pp. 123-130(2007-03-14).

A Proposal of a Care Worker Support System Using Care Worker's Act of Voicing "Koekake"

Jun Sawamoto*, Chikataka Sato*, Eiji Sugino*, Norihisa Segawa*, Hiroshi Yajima**,
Manabu Kurosawa**

* Graduate School of Software and information Science, Iwate Prefectural University, Japan
** School of Science and Technology for Future Life, Tokyo Denki University, Japan
sawamoto@iwate-pu.ac.jp

Abstract –Care service at a care facility of aged people has a predetermined and plus adjusted-for-the-user procedure, and a procedure consists of a series of operations. The operations and procedure of actual care work are varied depending on the individual workers and some care taking actions induce unexpected accidents during the care taking. In this research, we automatically collect the operations and procedures of the care service which each care worker is performing, then accumulate into a structured human functioning database. We aim at developing a system which supports to warn a new worker of doing the work which tends to induce an accident, and recommends better procedures of care service.

Keywords: Care service, Care worker, Structured human functioning data, Action of voicing, Voice data, Android terminal.

1 INTRODUCTION

In the care industry for aged people in Japan, there is a serious lack of human resources, it has a chronic labor shortage, and the burden of each care worker tends to increase steadily [1]. Therefore, in the care industry, even if a new employee joins a care facility, it is difficult to give him/her a sufficient off-the-job training, and the newcomer cannot but learn operations and the procedure of a service at the care spot directly in work after a short training period, and cannot utilize know-how of the senior workers in many cases [2]. A care service has a predetermined and plus adjusted-for-the-user procedure, and a procedure consists of a series of operations. It is required for a care facility of aged people to provide uniform and high quality service to the user. However, the operations and procedure of actual care work are varied depending on the individual workers and some care taking actions induce unexpected accidents during the care taking [2].

In this research, we automatically collect the operations and procedures of the care service which each care worker is performing, then accumulate into a structured human functioning database. We aim at developing a system which supports to warn a new worker of doing the work which tends to induce an accident, and recommends better procedures of care service.

2 RELATED WORKS

2.1 System Sara

Concerning the measure for improving the quality of the care service, Care house Sara [2] developed and introduced a care record and assessment system Sara[3]. The care house Sara has published a book [4] in which the know-how for creating an individual care procedure depending on the body condition and the state of the illness for each user is presented, and is building the system Sara based on the know-how of such care procedures. The system Sara operates on a mobile phone, and asks whether the right care procedure is performed in the nursing care service which the care worker actually performed, each care worker replies to it and the system evaluates each care worker's care service. The system is tackling about the quality of the care service like our research.

In order to evaluate each care worker's care service, it is necessary to determine the standard model of care services in advance and to input to the system Sara. In this point, Sara is similar to our research, namely in our research the contents of the nursing care services are analyzed from voicing data and in order to judge whether they are the right care services, we prepare the right care service procedures beforehand. Compared to Sara, we automatically collect the operations and procedures of the care service which each care worker is performing more finely than Sara through action of voicing "Koekake" of each care worker and reduce the time and effort of an input to the system as much as possible.

2.2 Communication by Voice Tweets

Torii et al. [5] has been developing an information assisting system based on smart voice messaging communications in a virtual field of nursing and care-giving space. In order to improve nursing-care services, they collect location information and voice twittering of nursing-care staff and integrate record of nursing-care utterance and nurse call. They verified that the system is effective in improving nursing-care services by conducting a field experiment. Utilizing voice twittering and location information enabled nursing-care staff to grasp the situation, which had been otherwise difficult for them. They confirmed it is efficient that nursing-care staff share up-to-date information of users and that they recognize and evaluate their services each other objectively.

Our approach is similar to this research in aiming improvement in the quality of nursing-care service at the care spot by using voice data, performing new communication between the staff, and visualizing and evaluating services with a little burden. However, our approach is not proposing a new communication tool by voice twittering, but we propose to utilize the action of voicing "Koekake" by each care worker to evaluate and urge an improvement of the quality of the care service.

3 PROPOSED SYSTEM

The action of voicing "Koekake" by a care worker to the user as shown in Fig. 1, such as "Let me take your left hand" or "Hold on the bar" preceding to real actions. "Koekake" is defined as the voicing action to inform the care recipient of the next care action to be taken by a care worker. "Koekake" is thought as a very important technique to improve mutual attention and coordination between the care recipient and the care giver.

This "Koekake" is recorded by speech recognition and used for the input to the system and analyzed and structured. Analysis of the actual contents of care is possible from the action of voicing in which the contents of the actually performed care are included. Extra effort for the worker of inputting data is reduced and the reliability of the data is improved compared to the off-line data input or record taking.

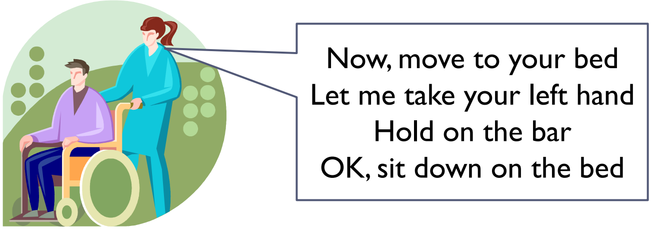


Figure 1: The action of voicing "Koekake" by a care worker.

3.1 Outline of the system configuration

The outline of the proposed system is shown in Fig. 2. Each care worker is working with an Android terminal in her/his chest pocket. And the system is constructed combining Android terminals and a server system. By the Android terminal, acquisition of the voice data of "Koekake" is performed. And the voice data saved in each terminal is transmitted to the server once a day while a terminal is charged in the middle of the night.

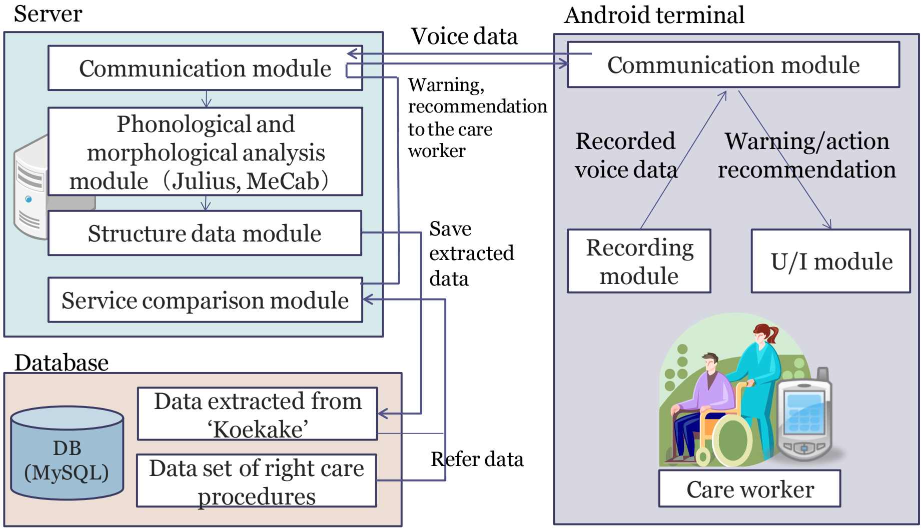
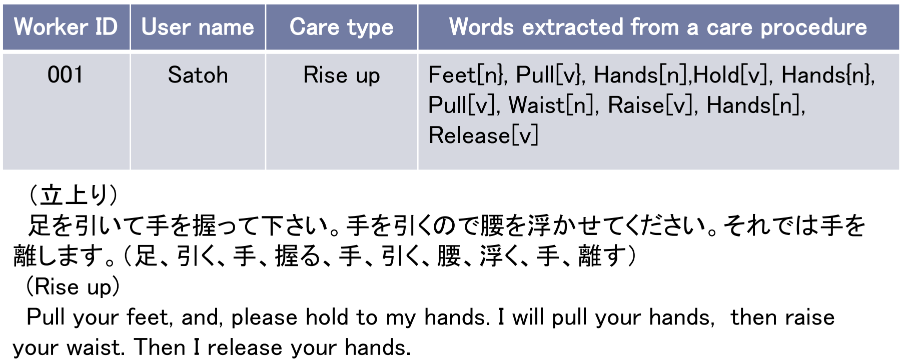

Server
Communication module
Phonological and morphological analysis module（Julius, MeCab）
Structure data module
Service comparison module
Voice data
Warning, recommendation to the care worker
Save extracted data
Refer data
Database
DB (MySQL)
Data extracted from 'Koekake'
Data set of right care procedures
Android terminal
Communication module
Recorded voice data
Warning/action recommendation
Recording module
U/I module
Care worker

Figure 2: Outline of the proposed system configuration.

In the phonological and morphological analysis module in the server, Julius [6] and MeCab [7] were used for the conversion of voice data to text data. In the morphological analysis using MeCab, the care term dictionary was provided in order to extract care related terms required for the system (i.e., nouns and verbs) and the end of a word unification of verbs of extracted terms is performed. Then, in order to unify synonyms which happen frequently in spoken language and to decrease the ambiguity of language, Weblio synonym dictionary [8] is used and synonyms are unified.

A series of words extracted from "Koekake" corresponding to a care procedure is saved in the human functioning database in a structured manner as shown in Fig 3. And, as for the accumulated data of care procedures, analysis is performed by the comparison module of the service, and the distance of the performed care procedure from the right care procedure is measured by normalized Levenshtein distance (NLD) [9] (normalized by the number of words) to recognize the assumed right care procedure. The "right care procedure" is prepared by the facility as a data set of safe and appropriate care procedures as the guide to care workers [9].

After recognizing the right care procedure and the evaluation of the carried out care service in terms of the distance from the right one, the system produces warning/action recommendation to each user (care worker). Each user uses an Android terminal and refers to his own evaluation and warning/action recommendation results.



Worker ID	User name	Care type	Words extracted from a care procedure
001	Satoh	Rise up	Feet[n}, Pull[v}, Hands[n],Hold[v], Hands{n}, Pull[v], Waist[n], Raise[v], Hands[n], Release[v]

（立上り）
足を引いて手を握って下さい。手を引くので腰を浮かせてください。それでは手を離します。（足、引く、手、握る、手、引く、腰、浮く、手、離す）
（Rise up）
Pull your feet, and, please hold to my hands. I will pull your hands, then raise your waist. Then I release your hands.

Figure 3: A set of word data extracted from a care procedure (a case without errors).

3.2 Service comparison module

3.2.1 Search of a target care service

In searching a target right care service, right care procedures are prepared in the database and the right care procedure which should be compared with the performed care procedure extracted from "Koekake" is first distinguished as a target care service. This right care procedure is a procedure of target care service in a nursing facility, it is safe and it is a care procedure without danger. In this proposed system, right care procedures are created based on the books such as "Ultimate Practical Care" [10] which introduces right care procedures and useful tips of care giving, and are input into the system in advance. These right care procedures are classified into six categories, meal, bathing, change of clothing, body posture, rise up, transfer, and are input into the database, and they are used in the search of target care services and the comparison of contents of care services.

In order to discover target nursing care services, the normalized Levenshtein distance (edit distance) [9] is used. It is usually used when calculating the similarity between character strings. This time the normalized Levenshtein distance is implemented to calculate the similarity of the input care procedure and the right care procedure which are constructed with care terms of the care procedures. When there are two series of terms, one for right care procedure "hand, soap, rub, towel, wipe" and the other for input care procedure "hand, wash, towel, wipe", first "rub" is inserted to the latter series then "wash" is substituted by "soap" in order to transform the input care procedure to the right one. In this example, one insertion and one substitution are performed and Levenshtein distance is 2. Since 6 terms appear in two series in total, the normalized Levenshtein distance of this example is calculated as 2/6 (0.33).

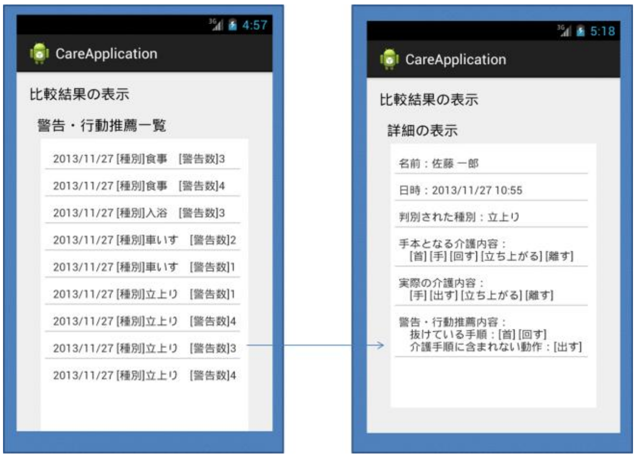
3.2.2 Comparison of contents of care services

In the comparison part, a lack of operation is checked in the actually performed care procedure compared to the right care procedure which is opted by "Search of a target care service". When the contents of the actually performed care procedure are included in the contents of the right care procedure and there is shortage of operations compared with the right care procedure, the portion which runs short compared with the right care procedure, the portion which is not included in the right procedure are recorded in the database. Then warning and action recommendation are performed to each care worker.

3.3 Warning and action recommendation

A care worker receives warning/action recommendation results based on the data recorded by "Comparison of contents of care services" when she/he views the evaluation of her/his care services with an Android terminal.

Fig. 4 shows the display image of a warning/action recommendation result on a mobile terminal. In the displayed image, the left side image shows a list of warning messages for the user and the right side image shows the content of one of the warnings for "rise up" procedure. The example points out some missing operations, "neck" and "hold" and also redundant operation "put", compared with the targeted right care procedure.



比較結果の表示
警告・行動推薦一覧
2013/11/27 [種別]食事　[警告数]3
2013/11/27 [種別]食事　[警告数]4
2013/11/27 [種別]入浴　[警告数]3
2013/11/27 [種別]車いす　[警告数]2
2013/11/27 [種別]車いす　[警告数]1
2013/11/27 [種別]立上り　[警告数]1
2013/11/27 [種別]立上り　[警告数]4
2013/11/27 [種別]立上り　[警告数]3
2013/11/27 [種別]立上り　[警告数]4

比較結果の表示
詳細の表示
名前：佐藤 一郎
日時：2013/11/27 10:55
判別された種別：立上り
手本となる介護内容：
[首] [手] [回す] [立ち上がる] [離す]
実際の介護内容：
[手] [出す] [立ち上がる] [離す]
警告・行動推薦内容：
抜けている手順：[首] [回す]
介護手順に含まれない動作：[出す]

Figure 4: A warning/action recommendation result on a mobile terminal.

4 PRELIMINARY EXPERIMENT AND ITS RESULT

In evaluating the proposed system, two preliminary experiments were conducted in this research. First, an experiment which evaluates the whole process from voice data analysis to "Search of a target care service" implemented in the system is carried out. Evaluation about the ability of the proposed system to recognize correctly in actual environment is performed here. Second, the performance of the "Search of a target care service" itself is evaluated. When converting a voice input into text data, under the influence of the surrounding noise etc., there are deficits in the series of terms extracted from "Koekake" of a care procedure. Under how much deficits, the system can identify the right care procedure is evaluated.

In this experiment, 49 data of six care types, i.e., meal, bathing, change of clothing, body posture, rise up, transfer, were used as the right care procedures of care services.

4.1 Experimental result of the whole process

Voice data of six subjects (students of our University) was prepared for the experiment as the input. After explaining the right contents of care procedures of three patterns of meal, bathing, and transfer, an Android terminal was put in the chest pocket of each subject and voice data of "Koekake" of above-mentioned three patterns without noise and with noise supposing the real care spot was recorded.

4.1.1 Experimental result

In this experiment including voice recognition, the rate which was able to extract the required care term with noise was 30% (42 out of 138 terms) and without noise 67% (92 out of 138 terms). The precision rate of search of a target care service with noise was 33% (6 out of 18 cases) and without noise 60% (11 out of 18 cases).

4.1.2 Consideration

In this experiment, when voice data of "Koekake" was given as an input, the problem that the accuracy of voice recognition had largely influenced by the quality of input sounds, such as the surrounding noise, was confirmed. It becomes a problem that various noises, such as sound of television and sound of the surrounding conversation, are made when this system is utilized at the actual care spot. In the future, experiments should be repeated considering the usage of a directional microphone, and it is important to improve both the quality of an input sound and also voice analysis accuracy.

Moreover, even in the situation without noise, the difference has appeared in the voice recognition rate with the volume of the voice. Since it is not so realistic to perform extremely small "Koekake" to the user who receives care services, we can verify the recognition rate when the analysis of voice is restricted to "Koekake" over certain prefixed level of volume. Furthermore, it is necessary to examine the noise level where an effective analysis result could be still obtained by the system.

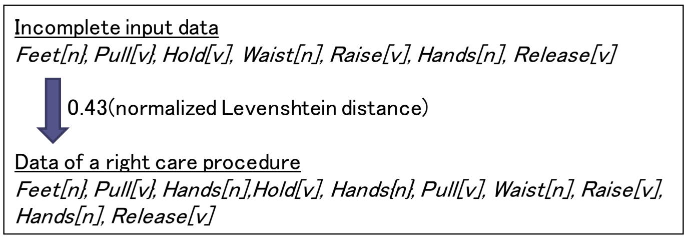
4.2 Experimental result for the evaluation of search of a target care service

Some accumulated incomplete or erroneous data of care procedures excluding some operations from the right care procedures were prepared, and input to the comparison module of the service. This serves as an input supposing there being some deficits in a speech recognition from situations, such as the surrounding noise, when speech recognition is performed at the actual care spot.

We evaluated the percentage of correctly retrieved answer (precision) as the ability of search of a target care service by the comparison module.

4.2.1 Experimental result

Three incomplete data were created for each selected right care procedure from 6 care types respectively. Experimental results were sorted in NLD. Fig. 5 shows an example of care type rise up. Input data is missing some care actions, and the data for right care procedure which is closest from the input data (NLD = 0.43) is retrieved. As an experimental result, among 18 test cases, 12 cases are correct (Precision = 67%). Furthermore, when the care type is given beforehand and considered, 13 cases are correct (Precision = 72%).



Incomplete input data

Feet[n], Pull[v], Hold[v], Waist[n], Raise[v], Hands[n], Release[v]

0.43(normalized Levenshtein distance)

Data of a right care procedure

Feet[n], Pull[v], Hands[n],Hold[v], Hands{n}, Pull[v], Waist[n], Raise[v], Hands[n], Release[v]

Figure 5: An example of data matching and normalized Levenshtein distance.

4.2.2 Consideration

In the experiment of the search precision, it is shown that the precision value is over 60% in average. However, the right search results were not obtained most of the time when the care procedure given as an input consists of only two or three care operations. In general, when the right care procedure includes many operations, and few operations is given as an input, it turned out that the precision of the search result becomes quite low.

Moreover, if normalized Levenshtein distances of the experimental results are examined closely, the largest value of normalized Levenshtein distance at the time of successful search is 0.6. And the minimum value of normalized Levenshtein distance when mistaken search results have come out is 0.5. Therefore, as a result, the normalized Levenshtein distance that exceeds 0.6 cannot be considered to suit with the right search result. However, even if the normalized Levenshtein distance is less than 0.5, it is unable to say that it suits with the right care procedure. Because an erroneous care procedure may induce an accident and affect human life, we need more precise way to judge the operations of the care worker.

Therefore, to raise the precision of search results further, it is required to devise some methods to consider characteristics which are not covered by such normalized Levenshtein distance.

5 CONCLUSION

We proposed a system which automatically collects the operations and procedures of the care services which each care worker is performing via voice data, then accumulates into a structured human functioning database. With preliminary experiments about speech recognition and the search of target care services, we confirmed the feasibility of the basic function of our method. About search of target care services, since what has the shortest distance in normalized Levenshtein distance is not always an intended candidate, and in order to raise the precision of matching, it is thought that it is necessary to consider comparison of the similarity which considers the meaning of the care procedures. We intend to develop the system further and evaluate the total system function from voice data of "Koekake" to warning and action recommendation.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 24500243.

REFERENCES

[1] Cabinet Office, Government of Japan, White Paper for Aged Society. http://www8.cao.go.jp/kourei/whitepaper/w-2013/gaiyou/index.html

[2] Dayhouse Sara. http://www.ito-pharmacy.jp/sara.html

[3] The actual situation of the care record system which aims to increase efficiency of home helper work and to improve the quality of the care service,Home visit nursing care service, No. 3/4, 2007, nissoken, http://www.ito-pharmacy.jp/zassi-houmon1.htm

[4] Itoh, M., How to make individual care recipes which nestle close to the heart, Miwa shoten (2007/10) .

[5]Torii,K.,Uchihira,N., etal. 2012.Service Space Communication by Voice Tweets in Nursing. In: Proc. of AHFE 2012.

[6] Julius. http://julius.sourceforge.jp/

[7]MeCab: Yet Another Part-of-Speech and Morphological Analyzer. http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

[8] Weblio. http://thesaurus.weblio.jp/category/wrugj

[9] Levenshtein distance. http://en.wikipedia.org/wiki/Levenshtein_distance

[10] Ohta, H., Miyoshi, H., 2013.Ultimate Practical Care, Kodansha. (in Japanese)

Operator-Assisted e-Government System and its Application Domain

Yoshihiro Uda, Kazuhiro Yoshida and Yoshitoshi Murata

Graduate School of Software and Information Science, Iwate Prefectural University, Japan
g236i002@s.iwate-pu.ac.jp, kazuhiro.iwk@gmail.com, y-murata@iwate-pu.ac.jp

Abstract -Japan is ranked 18th in the world in the 2012 United Nation's e-Government database. This relatively lackluster showing is due to poor system usability stemming from bureaucratic wording (jargon) and non-universal interfaces. To improve the e-Government system experience, we have developed an operator-assisted e-Government system in which operators at a call center assist applicants by providing application process guidance as well as by taking over keyboard operation for people with low IT literacy. Jargon is a major obstacle in Web site usability, and so we have also developed a feasibility analysis tool for current Web sites based on phrase difficulty indices. Evaluations using a prototype system indicated that operator-assisted applications could be completed 20% faster than conventional solo applications and with only negligible errors. The process times and error rates calculated by the proposed analysis tool were in good agreement with the experimental results. The proposed system will greatly help accelerate system usage by people who are confused by jargon in e-Government Web sites.

Keywords: e-Government, call center, operator-assisted application, jargon, Web site usability..

1 INTRODUCTION

The Japanese government launched its e-Government system in 2006. However, acceptance of the system has not been as high as expected. According to the United Nation's e-Government database in 2012, Japan ranked 18th in the world, while the Republic of Korea has remained on top for years [1]. Many studies on e-Government system usability have been carried out and major issues have been identified, most notably bureaucratic wording (jargon) and the lack of uniform interfaces among systems. In the view of economy and speed, instead of fully redesigning the current e-Government systems, improving the usability of the current system is desired.

To improve the user interface, we propose an e-Government system with call center-based operator assistance for applicants. The operator talks with applicants over the phone and helps them with the application process, also taking over keyboard operation for applicants who are not familiar with PCs and/or the Internet if necessary. We examined the effectiveness of the proposed system by implementing a prototype and measuring the process time and application content errors.

We have also developed a phrase difficulty rank evaluation tool for a usability analysis of current Web sites. This tool is useful for determining if we can improve usability by adding the call center functionality to the current e-Government systems.

The rest of this paper is organized as follows. In Section 2, we describe related work on e-Government systems and their usability issues. The functions and configuration of the proposed system are discussed in Section 3, and our experiments using the prototype system are explained in Section 4. In Section 5, we discuss the proposed system's application domain by analyzing the experimental results with our feasibility analysis tool. We conclude in Section 6 with a summary and a brief mention of future work.

2 RELATED WORK ON E-GOVERNMENT SYSTEM USABILITY ISSUES

The Web site usability issues facing the e-Government system have been discussed for years [2], [3]. In the early stages of system implementation, both governments and citizens expected the systems would be quickly and widely accepted. However, the adaption rate of the e-Government systems is lower than most e-commerce and other Web-site-based systems in Japan.

In an earlier study, Fuchs clearly pointed out substantial differences between e-commerce and e-Government systems, notably that the latter have no competition for the same services, lack a uniform outlook, and have subdivided territorial levels due to broad public administration scopes [4]. These findings explain why the current e-Government systems' one-stop interfaces cannot be used for different applications (tax payment, passport application, etc.) and why system usability is typically lower than that of e-commerce systems [5-7].

Bureaucratic wording (jargon and technical terms) is a significant indication of low usability of e-Government Web sites [8]. The basic problem is that applicants are forced to take extra time to learn this jargon. If jargon absolutely must be used on the pages, sufficient explanation should be included. Also, application guidance for applicants should be provided. We found that, while some e-Government Web sites have links to operation manuals, most applicants would not willingly take the time to read the huge manuals in detail, and often they did not even notice the manuals were there in the first place.

Gauvin et al. found that age is a markedly higher demographic determinant of Internet usage than education, income, gender, or urbanity [10]. Thus, assisting the elderly with the Internet and e-Government system operations has become an important issue for governments to better serve their citizens in view of e-inclusion [11-13]. Kim reported that "mass digital literacy campaigns" for several tens of millions of elderly citizens, government officials, and housewives were carried out as part of Information Technology (IT) education programs in the Republic of

Korea—a country which has been consistently at the top of the e-Government system usability rankings. These campaigns are one of the key reasons for the success of the Korean systems [14]. However, such widespread educational campaigns are not necessarily applicable to massive populations of senior people such as found in Japan. Moreover, in the past few years, the Internet access platform has expanded from fixed line communication by PCs to smartphones and tablets. Preparing educational programs which cover all the different access platforms takes time, and once completed, the programs may already be out of date as IT networks continue to evolve. In this work, we propose a new user assistance system which is more suited to the unique situation in Japan than is mass user education

3 PROPOSED E-GOVERNMENT SYSTEM

3.1 Proposed System Functions

Senior citizens and other people with low IT literacy require assistance to use the e-Government system, and call centers have been set up to address this issue [15], [16]. Call center operators assist applicants by helping with the application process and answering any questions related to the jargon. These operators are trained in a flexible manner to allow for access platform evolution. Our proposed system extends the call center operator function from merely providing applicants with verbal assistance to taking over keyboard operation for them, as well [17], [18]. This would greatly reduce the burden on applicants with low IT literacy applicants. Our objective with this system is to reduce the application process time and to minimize operation and application errors.

As discussed in the previous section, e-Government systems which contain a lot of jargon result in poor Web site usability and the high possibility of making errors in the application process. Solo application takes longer on a difficult Web site than operator-assisted application and is also likelier to include more errors. From this viewpoint, we believe that e-Government Web site usability can be measured based on process time and application error rate. We explored these assumptions through experiments, as discussed in Section 4.

We also propose a numerical analysis tool to measure jargon difficulty in our usability evaluation of current Web sites. Jargon difficulty index definitions are provided for the major phrases used on the Web pages and the difficulty levels are then applied to process time calculations for applications with and without operator assistance. The error rate estimation for the solo application is carried out by weighted calculation of the difficulty levels and the number of conditional branches. By using the analyzed results (expected processing times and error rates), e-Government system owners can predict the call center effectiveness for current e-Government systems.

3.2 The System Configuration

In the proposed system, operators located at call centers for e-Government systems talk with applicants over the phone (Fig. 1). During the application process, an applicant and an operator share the Web pages of the application system through their respective PC displays (Fig. 2). The page sharing system is implemented by the following three processes.

1) Applicant identification. The applicants are given customized identification numbers (e.g., phone numbers) to correspond with the operator over the phone and Web pages.

2) Web page sharing between applicant and operator. Two Web page sets are prepared from the current e-Government system database: one for the applicants and one for the operators. During the operator-assisted application process, the operator works on the PC keyboard and fills in the application form on the operator Web page based on their conversation with the applicant. The application content is then stored in the database at the call center.

3) Confirmation of application content. When the operator finishes the keyboard work, the Web page is displayed on the applicant's PC so that the applicant can confirm that the content is correct. After confirmation, the applicant clicks on the register button on the Web page and finalizes the application form to be sent from the call center database to the e-Government system.

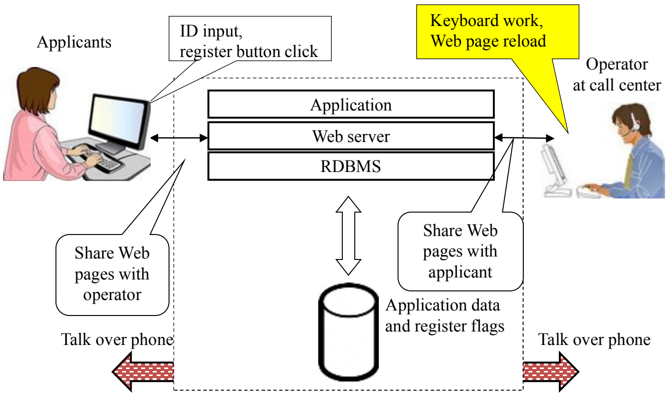
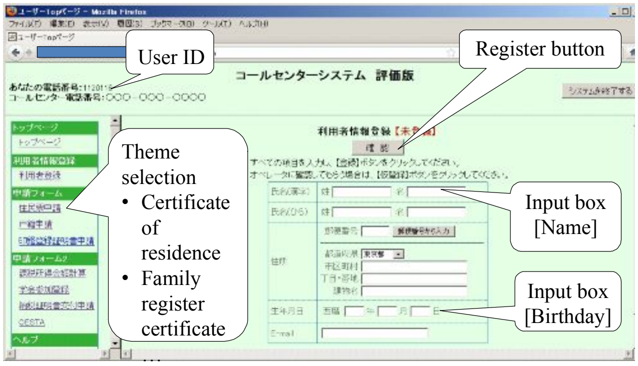


Figure 1: Proposed operator-assisted e-Government system.



ユーザーTopページ - Mozilla Firefox

コールセンターシステム 評価版

あなたの電話番号:112011
コールセンター電話番号:〇〇〇-〇〇〇-〇〇〇〇

トップページ
トップページ
利用者情報登録
利用者登録
申請フォーム
住民票申請
戸籍申請
申請フォーム2
ヘルプ

利用者情報登録【未登録】
確認

氏名(漢字) 姓 名
氏名(ひらがな) 姓 名
住所 郵便番号 都道府県 東京都 市区町村 丁目・番地 建物名
生年月日 西暦 年 月 日
E-mail

Figure 2: Example of operator's page on operator's display.

3.3 Phrase Difficulty Ranks for e-Government Web Site Jargon Analysis

There are many methods to determine the difficulty level of kanji (the Chinese characters used in the Japanese language), mainly prepared for non-Japanese speakers.

Basic Japanese words are also classified for Japanese students and foreigners in "Kanji 2100" [19]. However, most of the jargon found in e-Government Web sites is not included in the current basic word classifications and thus no difficulty levels are available. E-Government jargon requires a much higher reading level than the average citizen possesses, because the terms are not common in textbooks, newspapers, and magazines.

We propose an analysis tool to determine the phrase difficulty rank for jargon appearing in e-Government Web sites. The difficulty rank definition is given by assigning unique difficulty indices as extended ranks in the Balanced Corpus of Contemporary Written Japanese (BCCWJ) [20]. The BCCWJ covers a wide range of popular phrases found in books, magazines, newspapers, and Web sites.

The index assignments are ranked as follows.

Rank 1: Phrases found in "Kanji 2100." This rank corresponds to basic words at the reading level of Japanese junior high school students.

Rank 2: Phrases not listed in "Kanji 2100" but which have 100 or more search results in the BCCWJ.

Rank 3: Phrases which have 10 to 99 search results in the BCCWJ.

Rank 4: Phrases which have less than 9 search results in the BCCWJ or which contain only some of the words found in the BCCWJ.

We perform our analysis of Web site usability as follows.

Step 1: List all the phrases used on an e-Government Web site.

Step 2: Assign phrase ranks to the listed phrases per the above-mentioned indices.

For example, in the Web site of the National Tax Agency in Japan, the following paragraph is given as the income tax deduction guide of medical expense [21]. Underlined phrases from A to H in Fig. 3 are jargon and require high reading levels for the applicants:

If you pay premiums or premiums which are withdrawn
Phrase A
from your salary for health insurance,
Phrase B
national health insurance (tax), national pensions insurance,
Phrase C ***Phrase D***
national pension fund, medical insurance for the old-old
Phrase E ***Phrase F***
and nursing-care insurance, etc. of yourself, your spouse or
Phrase G
relatives living in the same household as you (refer to page 23) , you may claim this deduction.
Phrase H

Figure 3: Examples of jargon phrases in a medical expense tax deduction guide.

Table 1: Phrase difficulty ranks and time factors for the jargon phrases in Fig. 3

Phrase	A	B	C	D	E	F	G	H
Difficulty rank	2	1	2	2	3	4	2	2
Time factor	1.3	1	1.3	1.3	1.6	1.9	1.3	1.3

In step 2, phrase difficulty ranks are assigned to each phrase as summarized in Table 1. Time factors are defined per phrase difficulty rank. The time factor is a parameter to denote the search time ratio during the application process between the easiest phrase (rank 1) and a phrase of certain rank (rank 2-4). A phrase of rank 4 would require 1.9 time longer search time than a rank 1 phrase. The time factors are optimized by fitting the measured times in the experiments to time factor parameter sets.

The phrase difficulty ranks and time factors are applied to the expected process time calculations by solo and operator-assisted applications as well as solo-application error rate calculations, as described in Section 5-3.

4 EVALUATION EXPERIMENTS

4.1 Application Themes

Evaluation experiments were carried out to determine if the proposed system could improve system usability, as discussed in Sections 2 and 3. The parameters we measured to compare the usability of the current application systems and the proposed one were processing time and number of application errors (error rates).

Table 2: Numbers of ranked phrases.

	Theme A	Theme B	Theme C	Theme D	Theme E	Theme F
Rank	Registration	Residence certificate	Family register	Conference	Income tax	Tax certificate
1	5	0	0	0	6	4
2	2	3	2	2	7	9
3	0	2	2	8	11	13
4	0	1	0	5	11	13

Six application themes based on current e-Government and e-application systems were prepared on a Web server as a set of Web pages and databases. These themes ranged from simple to complicated processes as well as those which require a good understanding of both the application process and the jargon (Table 2). The design concepts of the themes are listed below.

Theme A: Registration of applicant profile (applicant's name, address, etc.). This theme focuses on correct inputs.

Theme B: Certificate of residence. Applicants are guided to register a new bicycle, and one of the requested documents is the certificate of residence. This theme examines if the applicant can choose the proper document required for bicycle registration.

Theme C: Family register certificate. Applicants are guided to change the legal domicile for their new passport. This theme determines if the applicant mixes up the old and new domiciles and/or current living address (note: in Japan, the legal domicile and the living address may not be the same).

Theme D: Technical conference registration. Applicants are requested to fill in the registration form for a technical conference. This theme examines if the correct options have been selected to calculate the registration fee under given conditions (member discount, etc.).

Theme E: Income tax calculation. A simplified tax calculation system is provided. Applicants are requested to input the total income as well as life insurance, social insurance, and medical expense deductions. This theme examines if the applicant can understand the deduction system described with a large amount of jargon and make correct calculations for the related deduction items.

Theme F: Tax payment certificate. The applicants need the tax payment certificate for a housing loan refinancing. The certificate application form is complicated and difficult to understand due to the jargon. The design concept of this theme is similar to Theme E in which it tests information access and correct calculations.

4.2 Experiment setup

Each test participant was given an individual identification number and then assigned to either a solo application group or an operator-assisted application group. The solo application group simulated the use of conventional application systems while the operator-assisted application group was established to determine the advantages of the proposed system.

Test participants and operators had their own PCs and displays but could look at only their own displays (Fig. 4). Applicant profiles (name, address, birth date, etc.) were fictitious and were commonly applied to all participants.

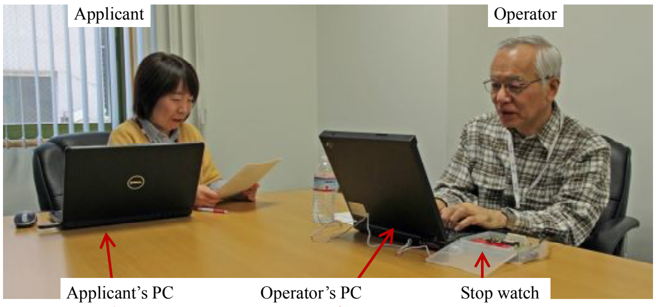


Figure 4: Experiment scene of operator-assisted application.

4.3 Experiment process

Applicant action steps are summarized in Table 3.

Participants in the solo application group were instructed to complete all themes by themselves. Each participant read a set of instructions to understand what information and actions were necessary to complete the application. If a participant did not understand the technical terms or jargon in a theme explanation, he/she had to use Internet search engines for guidance.

Participants in the operator-assisted application group spoke directly with an operator (one of the authors) instead of making a phone call. After choosing a theme, the participant asked the operator for guidance and provided the operator with information as guided. The operator looked at the application Web page and worked on the keyboard. After finishing the input process, the operator told the participant to update the Web page and confirm if the correct information appeared on the application form. After the participant provided this confirmation, he/she clicked on the “register” button on the Web page and the application was completed.

5 EVALUATION OF EXPERIMENT RESULTS

5.1 Test Participants

Thirty-seven participants (20 students and 17 office workers) were divided into two groups: solo (27 applicants) and operator-assisted (10 applicants). The operator-assisted group was smaller than the solo-application group because the preliminary experiments showed highly consistent results for both process time and errors in the operator-assisted applications. The participant profiles were classified according to age and years of PC experience (Fig. 5).

5.2 Experimental results

Both time and error comparisons demonstrate the effectiveness of operator assistance during the application process.

The experimental results of the average processing times for both solo and operator-assisted applications are shown in Fig. 6. With the latter three themes (D, E, and F), solo applications took 20% longer to finish than operator-assisted ones. These three themes contain a lot of highly ranked jargon, and it took the solo participants a lot of time to search for the meaning and understand the process.

We also compared the error rates of the solo and operator-assisted applications (Fig. 7). Operator-assisted applications had a significantly lower number of errors compared to solo applications for all themes. With theme C, a few errors occurred in operator-assisted applications because some of the participants confused the living address with the legal domicile and provided the operator with incorrect information. For solo applications, significant errors were observed when a lot of jargon and several conditional branches appeared in the application form.

According to questionnaires filled out by participants after the tests had been completed, the operator-assisted application obtained good favor, especially among seniors. More than 90% of participants claimed that the operator-assisted e-Government system would greatly reduce barriers to system entry.

Table 3: Applicant action steps

Step	Item	Action for solo application	Applicant's action for operator-assisted application
1	Read and understand a theme	Perform Web search for jargon and information on application process.	Asks operator to explain application process and gets advice on phone.
2	Fill in input box	Types in requested content in input box of Web page.	Answers operator's question and allows operator to fill in input box.
	Above steps are repeated untill last input box of Web page is filled.		
n	Confirm	Checks input results and clicks "Confirmation" button to proceed.	Shares Web page with the operator. Clicks "Confirmation" button after the input results have been confirmed.

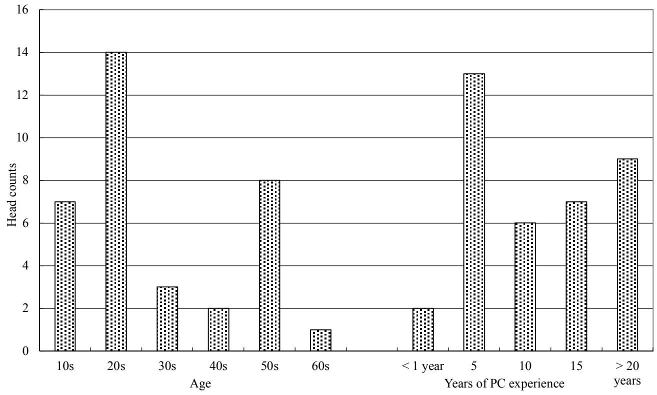


Figure 5: Test participant profiles (Age and years of PC experience).

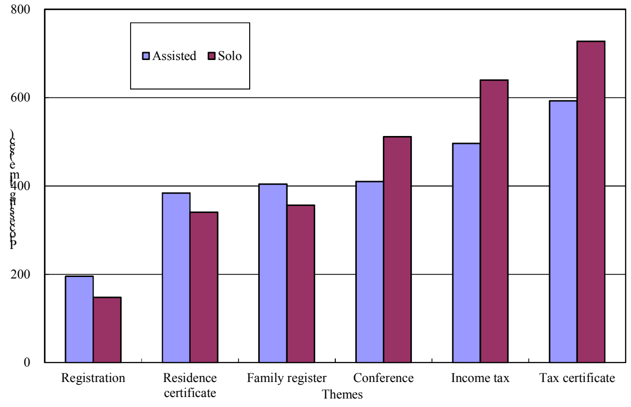


Figure 6: Experimental results of average processing times for operator-assisted and solo applications.

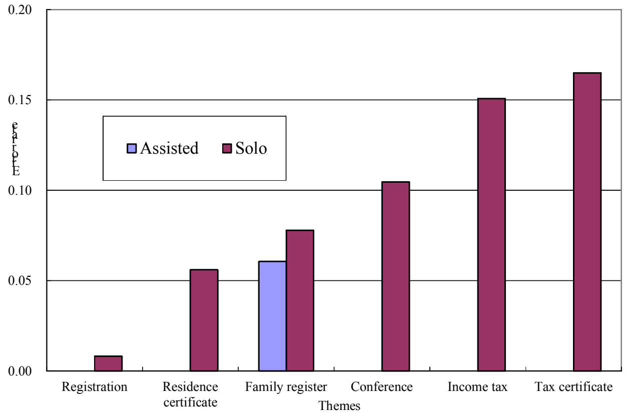


Figure 7: Experimental results of error rates for operator-assisted and solo applications.

5.3 Feasibility analysis tool and its evaluation results

The process time calculation was carried out using the proposed analysis tool for each experiment theme. Table 4 shows part of the analysis table for Theme E (income tax calculation). In this example, the first input box is the medical expense tax deduction amount. Prior to filling in this box, the applicant has to understand the meaning of jargon such as “Medical expense tax deduction”, “Medical insurance supplementation”, and “Hospital expense grant” by using the Internet search engines. The phrase ranks for the jargon are given based on the rules explained in Section 3.3. The time factors per phrase ranks are listed in the next column of the table. The expected jargon search time and the time taken to understand the instructions for solo application is calculated by multiplying the time factor sum (9.6 in this example) and a unit time (10 seconds) which denotes the average of measured time for jargon of rank 1 during the pre-experiment. Then, 10 second keyboard operation time is added to the search and understand time, resulting 106 seconds to fill the first input box. This calculation process is repeated for each input box on the theme table and the total calculated processing time is then examined.

For the operator-assisted applications, jargon search is not necessary and no phrase difficulty ranks are examined, because the operator provides the applicants with adequate advice. On the other hand, they must communicate each other over the phone, thus we allocate 30 second understand time and 20 second keyboard operation time to start at the first input box. Keyboard operation time for the operator-assisted application is longer than one for the solo application due to the conversation between the applicant and the operator to relay the information such as an amount of medical insurance supplementation before filling in the box. After the operator finished filling in the input boxes, he/she had to ask the applicant to check the box contents and verify the application form. This process is not necessary for solo application.

Thus, we defined unique unit times for solo and operator assisted applications referring to the pre-experiment results.

The calculated times for the six themes are shown in Fig. 8 and are compared to the experimental results both for solo and operator-assisted applications. The calculated times were in good agreement with the measured results in the experiments.

Table 5 shows the process used to calculate the solo application error rate for the same part of the process time calculation in Table 3. In Table 5, the error rate calculation results based on the phrase difficulty indices were smaller than the measured error rates. This indicates that some other calculation parameters should be added to explain the measured results. We analyzed the experiment process and found that the conditional branch in the application process caused solo application errors. Specifically, some applicants chose an incorrect branch, which resulted in a certain number of errors. We therefore added the number of conditional branches to the calculation parameters. After a consideration of the measured error rate, we chose 0.02 as a

Table 4: Part of analysis table for Theme E (Income tax calculation)

Application step	Jargon	Phrase difficulty rank	Input item	Time factor	Time factor sum	Solo application calculated time (sec)	Operator-assisted application calculated time (sec)
Read theme and understand/search	Medical expense tax deduction	2		1.3			
	Medical insurance supplementation	4		1.9			
	Hospital expense grant	4		1.9			
	Major medical expense	2		1.3			
	Family medical expense	3		1.6			
	One-off maternity benefit	3		1.6	9.6	96	30
Fill in input box			Medical expense tax deduction amount			10	20
Establish verification between applicant and operator						—	10
Calculated processing time for an input item						106	60
Read theme and understand/search	Social insurance tax deduction	3		1.6			
	National pension	1		1			
	National health insurance	2		1.3			
	Nursing-care insurance	2		1.3			
	Unemployment insurance	3		1.6	6.8	68	30
Fill in input box			Social insurance deduction amount			10	20
Establish verification between applicant and operator (Note 1)						—	10
	Note 1: For operator-assisted application						

Table 5: Part of error rate calculation table for Theme E (Income tax calculation)

Jargon	Phrase difficulty rank	Input item	Weighted difficulty index	Difficulty index sum	Error rate coefficient	Calculated error rate	Measured errors	Measured error rate
Medical expense tax deduction	2		1.3					
Medical insurance supplementation	4		1.9					
Hospital expense grant	4		1.9					
Major medical expense	2		1.3					
Family medical expense	3		1.6					
One-off maternity benefit	3		1.6					
		Medical expense tax deduction amount		9.6	0.02	0.192	9	0.2727
Social insurance tax deduction	3		1.6					
National pension	1		1					
National health insurance	2		1.3					
Nursing-care insurance	2		1.3					
Unemployment insurance	3		1.6					
		Social insurance deduction amount		6.8	0.02	0.136	17	0.5152

phrase difficulty index coefficient and 0.04 as a conditional branch coefficient. For theme E, the calculated error rate was 0.158, which agrees with the measured error rate of 0.151. Figure 8 compares the measured and calculated error rates for all themes with the same coefficients used for theme E. All rates were in good agreement, particularly those for the difficult themes (D, E, and F) which contain a lot of jargon and conditional branches.

These results lead us to conclude that the proposed processing time and error rate calculations are useful for examining current e-Government Web sites in terms of the effectiveness of call center operators to obtain better user interfaces.

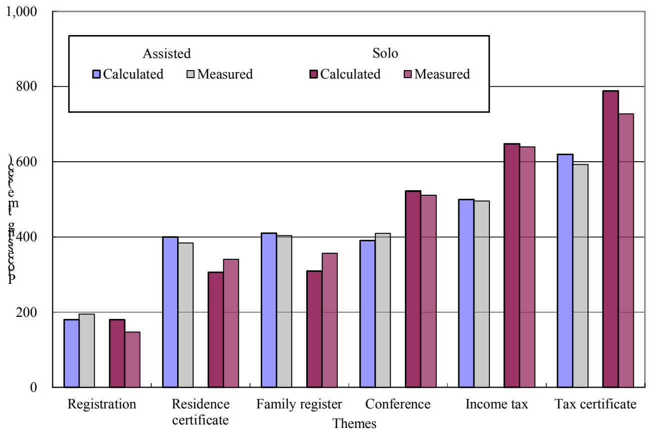


Figure 8: Comparisons of calculated and measured processing times for both operator-assisted and solo applications.

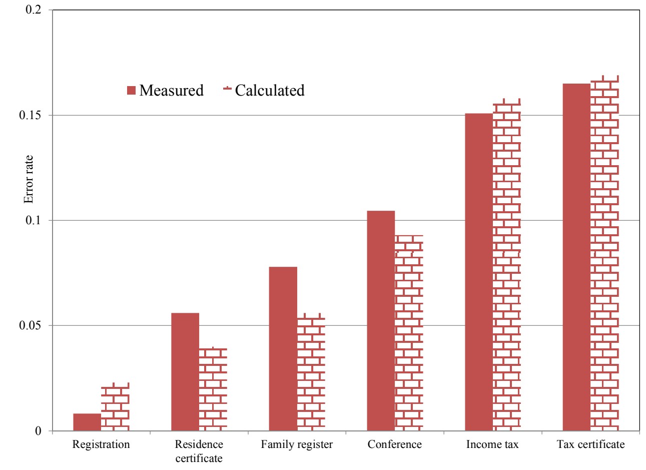


Figure 9: Comparisons of calculated and measured error rates for solo applications.

6 CONCLUSION AND FUTURE WORK

We proposed extended call center functionality for use with the current e-Government systems in Japan. Call center operators talk with applicants over the phone and assist them by helping with the application process and taking over keyboard operations for applicants who are not familiar with PCs and/or the Internet. The effectiveness of the proposed system was demonstrated through experiments with a prototype system in which we measured the process time and application content errors for both solo and operator-assisted applications.

The experimental results suggest that the process time can be shortened by as much as 20% by operator-assisted application compared to solo application in case of themes with full of jargon and conditional branches. Also, errors which occurred in solo applications were negligible in operator-assisted applications. We also proposed a tool for analyzing the feasibility of current Web sites based on phrase difficulty ranks and the number of conditional branches in Web site pages. Analysis results were in good agreement with the processing time and the error rate results measured in the experiments.

These results, along with positive feedback from test participants, indicate that the proposed e-Government system has great potential to accelerate system usage by people with low IT literacy. The proposed system has particular promise for e-Government Web sites which contain much bureaucratic wording (jargon) and lack uniform interfaces among systems.

For future work, we intend to determine Web site usability for users with a different mother tongue than the language on the site.

REFERENCES

[1] United Nations Public Administration Network, “2012 Global E-Government Survey,” 2013 on http://www.unpan.org/egovkb/global_reports/08report.htm [retrieved: June, 2014].

[2] J. Nielsen, “Usability 101: Introduction to Usability,” 2012 on http://www.nngroup.com/articles/usability-101-introduction-to-usability/ [retrieved: June, 2014].

[3] J. Iio and H. Shimizu, “Evaluation improvement method for business system usability,” Research Papers, Mitsubishi Research Institute, Vol. 50, pp. 30-53 (2008), (in Japanese).

[4] G. Fuchs, “Lost Youth? Attitudes Towards and Experiences Withe-Government: The Case of Germany University Students, “ Proc. of 12th. European Conference on e-Government, pp. 251-258 (2012).

[5] R. Schwester, “Examining the Barriers to e-Government Adoption,” Leading Issues in e-Government Research, pp. 32-50, Academic Publishing International Ltd., (2011).

[6] D. Evans and D. C. Yen, “E-Government: Evolving Relationship of Citizens and Government, Domestic, and International Development,” Government Information Quarterly, Vol. 23, pp. 207-235, (2006).

[7] S. Elling, L. Lentz, M. De Jong, and H. Bergh, "Measuring the Quality of Governmental Websites in a Controlled versus an Online Setting with the 'Website Evaluation Questionnaire'," Government Information Quarterly, Vol. 29, pp. 383-393, (2012).

[8] Z. Khabaziyan, H. Teimori, and M. Hekmatpanah, "Planning E-Citizen: A Step toward E-Society", World Academy of Science, Engineering and Technology, Vol. 59, pp. 2590-2593, (2011).

[9] P. Jaeger and M. Matteson, "e-Government and Technology Acceptance: The Case of the Implementation of Section 508 Guidelines for Websites," Leading Issues in e-Government Research, pp. 231-252, Academic Publishing International Ltd., (2011).

[10] S. Gauvin, K. Granger, M. Lorthiois, and D. Poulin, "The Shrinking Digital Divide – Determinants and Technological Opportunities, " Proc. of 12th European Conference on e-Government, pp. 259-267, (2012).

[11] C. W. Phang, J. Sutanto, A. Kankanhalli, Y. Li, B. C. Y. Tan and H-H Teo, "Senior Citizens' Acceptance of Information Systems: A study in the Context of e-Government Services," IEEE Trans. on Engineering Management, vol. 53, No. 4, pp. 555-569, (2006).

[12] T. Molnar, "Best Practices for Improved Usability of e-Government for the Ageing Population, " Proc. of 12th. European Conference on e-Government, pp. 493-501, (2012).

[13] E. Mordini et al. "Senior Citizens and the Ethics of e-inclusion," Springer on http://link.springer.com/article/10.1007%2Fs10676-009-9189-7#page-1, (2009) [retrieved: June, 2014].

[14] S. Y. Kim, "Korea ICT/e-Gov History, Best Practices and Lessons," presented at Georgian Cyber Security and ICT Innovation Conference 2011, (2011).

[15] A. K. Singh and R. Sahu, "Integrating Internet, Telephones, and Call Centers for Delivering Better Quality e-Governance to All Citizens," Government Information Quarterly, vol. 25, pp. 477-490, (2007).

[16] F. Bao and F. Zhao, "Study on the E-Government Call Center System Based on SOA," Computer and Information Science, Vol. 4, No. 4, pp. 120-122, (2011).

[17] Y. Murata, Y. Sato, T. Takayama, and N. Sato, "E-Government System Using an Integrated Call Center System and WWW," Proc. of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology – Vol. 03, pp. 199-202, (2008).

[18] Y. Uda, K. Yoshida, and Y. Murata, "Proposal on Operator-assisted e-Government Systems," Proc. of the Eighth International Conference on Digital Society, pp. 27-32, (2014).

[19] Y. Tokuhiro, "Kanji (Chinese characters) 2100, Listed according to Frequency and Familiarity," Sanseido Co. Ltd., (2008). (in Japanese).

[20] Japanese Corpus project, "Shonagon," 2012 on http://www.kotono.ha.gr.jp/shonagon/ (in Japanese) [retrieved: June, 2014].

[21] "Income tax and special income tax for reconstruction guide for foreigners," on https://www.nta.go.jp/taxanswer/english/12004.htm [retrieved: August, 2014].

Basic Design of Smartphone Application for Childcare Support

Ayaka Nishiwaki[*], Katsuhiro Naito[**], Takaaki Hishida [**], Tadanori Mizuno[**]

[*]Graduate School of Business Administration and Computer Science,
Aichi Institute of Technology, Japan
[**]Faculty of Information Science, Aichi Institute of Technology, Japan
2ayaka4nishi8@gmail.com

Abstract - The birthrate declining in recent years has attracted attention more and more. The local governments of Toyota city and Aichi prefecture in Japan have tried to support families with small children to increase the birthdate. The current support provides some special services for childcare in stores and facilities. However, practical difficulty to obtain the information is still an issue because only static information is an available for citizens. This paper proposes a basic design of childcare support system using smartphone applications. The applications can provide childcare information more visually and adaptively to families with small children. Additionally, the proposed system can manage word of mouth for some facilities like as restaurants, parks, communal facilities etc. The information can be used for selecting more adequate facilities for childcare support and improving services in these facilities. In the evaluation, we show that our developed application can provide a management function for word of mouth and search function for facilities providing childcare.

Keywords: childcare, word of mouth, map, smartphone application, visualization of childcare information

1 INTRODUCTION

Japanese industries, companies and governments have recognized the declining birthrate in recent years because it will probably have a severe impact on the Japanese economy. National census shows that the number of births in 2013 year decreased to 1,029,800, which is less than 69% in 1983. Local governments like as cities and prefectures have also recognized these issues. Therefore, some local governments have started to support families with small children to increase the birthrate. Toyota city and Aichi prefecture in Japan also started the support for childcare from 2008 because the population ratio of the younger than 24 years old decreased from 46% in 1975 to 27% in 2010. The support provides some discounts and special services in cooperate facilities likes as restaurants, facilities etc. However, the difficulty to obtain the information about these services is still an issue due to static based information.

57.1% respondents said the importance of supporting families with small children in the survey by the he Cabinet Office, government of Japan. However, the respondents in their thirties also show the slightly negative for direct communication to neighbors. The survey may show families with small children require supports for raising children from a local society with indirect communication. This mood can be found in anther survey about the issues in the raising children. The following is the quotes from the survey.

- Internet is the only way to obtain the new information when they move to a new city.
- They don’t have enough time to access Internet due to raising children.
- Information from Internet is not summarized.
- The best way is to obtain real information from friends directly.
- They don’t have enough chance to communicate with person from another generation.

On the Internet that much information flies about, it takes time to find necessary information, but can suppose that information is necessary as the child who cannot take its eyes off it all day to know it from the answer searching at bedtime of the child. Furthermore, we discovered that word of mouth is effective we can get necessary information as needed from friends.

Described the about childcare support carried out in two chapters to date as follows in Toyota city, speaks the development to a human probe in five chapters about the construction method of the software in four chapters about a basic function of "MAP for childcare" in three chapters as follows.

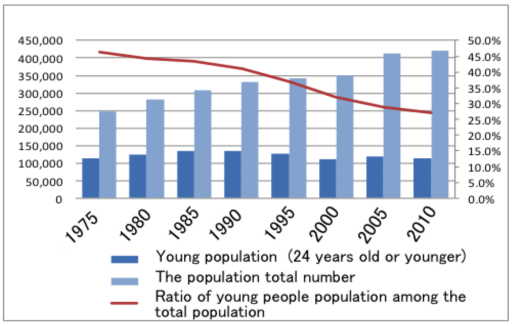


Figure 1: Young people population of Toyota city.

2 CHILD CARE SUPPORT OF TOYOTA CITY

Toyota city made a public nursery school, a public kindergarten a named "child garden" in Toyota city in 2008 and decided to do integral use. Toyota city offered uniform childcare in a public nursery school and a public kindergarten and decided to do it with equal childcare charges. We nominated it for an example and compared Toyama city, Toyama where the population total number was about the same with Toyota city. In Toyota, 67 places of public child garden of Toyota city were installed [5]. While, the number of public kindergartens installed was 55 places in comparison with public nursery schools in Toyama city [6]. Toyota city understands that there are many public child-care facilities by this result in comparison with Toyama city and understands that Toyota city lays emphasis on the setting of the child-care facility as a part of the childcare support. Furthermore, Toyota city carry out some plans about the childcare support. We give three about the childcare support plan that Toyota city carries below out.

(1) HAGUMIN card (childcare home special treatment business)

Business [7] that various scores that each store sets originally by distributing a courtesy card (HAGUMIN card) to homes with children under the age of 18 and the pregnant moms with the childcare home special treatment business, and showing this card when they used the shopping in a support store, facilities (special treatment shop of HAGUMIN) of the prefecture are received. This business is childcare support project that Aichi carries out in cooperation with cities, towns and villages exactly; Toyota city also cooperates in this project. A support store, facilities are it in 209 stores in total in Toyota city and carry out various privileges.

(2) The Toyota city child Ordinance

For the purpose of realizing the community which children who carried the future of Toyota city could live for happily, Toyota city established " The Toyota city child Ordinance " by guaranteeing the right of the child, and assisting the breeding of the child in Great Society [8]. Toyota city pushed forward various actions for environmental maintenance to surround a child so far in Toyota city, and while there are plans to make changes, there are no concrete actions in place to make a difference. In addition, in regards to rights abuses of the child including the bullying unless the facts of an apparent injury are seen until now; was groundless. In contrast, Toyota city come to have legal binding force in this Toyota city establish the child regulations that included the duties of the way and the person concerned with childcare, the child breeding support. In other words both policies are examined from "the viewpoint of the child", and the examination result comes to be reflected by contents of facilities, the business that Toyota city works on. The child regulations stand in none of "viewpoint of the child" in a policy about support raised in the right security of the child and child care, a child in this way, and it can be with a continuous general base pushing forward systematically.

(3) Toyota City child comprehensive plan (new Toyota child smile plan)

Because Toyota city guaranteed the right of the child, and the town which was kind to a child pushed forward making it generally and premeditatedly, Toyota city devised "a child synthesis plan" to prescribe it in child regulations Article 26 [9]. Toyota city intend for 0 to 20 year olds by the Toyota City child comprehensive plan and am intended to support independence from the birth of the child. The action policy can be divided into eight different principles.

- Construction of consciousness enlightenment and the relief support system of the right of the child
- Creating a safe and reliable environment to birth and raise a child
- Harmony of improvement of childcare, preschool education and work and life of the parent
- Reinforcement of the pro-breeding power in the home
- Improvement of the area power to support childcare
- Improvement of child independence
- Environmental maintenance and promotion made with an open-space school that a child is brought up and learns
- Promotion of young healthy upbringing and the support until independence to take the next era on

Toyota city carrys these out in Toyota city. However, it is the present conditions that approximately less than 40% citizen does not know less than 30% as for the protector about the cognitive situation of the child regulations [9]. From this, they can suppose that the recognition of the childcare support of Toyota city may be low. We mentioned in a previous section, that the action of all cities extracts it with nature and is disclosed because the things that a citizen obtains information includes are few. Therefore something providing information specialized in childcare may be necessary. Childcare support in Toyota city that we spoke in this chapter is only a part of the whole support. In addition, "MAP for childcare" plans visualization of childcare support performed in the city because there are a store and facilities performing childcare support in individuals other than the action of the city in the city. We speak of the function in the following chapter.

3 BASIC FUNCTION

We introduce a main function of "MAP for childcare". The functions that we can perform include three areas.

- Contribution of the word of mouth information
- Indication of childcare support facilities
- Indication of the support store of HAGUMIN.

We will have a users write and store information, the detailed information of facilities on a map and make up the map details by users. The word of mouth information of childcare uses will include restaurants that have child menus, easy to enter with children, and other features. A store and facilities are available and come to be gone to the town of childcare easily in peace by reading a note. Furthermore, we can do reviews of quality and store in itself of service going for by an evaluation by the word of mouth because we can watch word of mouth as the store and facilities side. It is in this way organization will be able to offer a pulling in of customers effect to a store and the facilities (figure 2). We squeeze it to an smartphone application full-time homemaker and think about structure of the smartphone application.

We will make the smartphone application to promote Toyota city going out of the user mainly on map application and two services of the store information service demanded among smartphone users having a child of the childhood period in particular. First, we display a map of Toyota city and appoint the place where a user is on a map or the place where we have been to and accumulate the name of the facilities, an average budget, an evaluation for the human log. We call a word of mouth database with a database as follows. When a user uses it as a map, we add the human log that accumulated from a database on a receipt, a map. Specifically, we find the longitude of the spot that registered from human log data, latitude and display a pin there. We display the human log data of the spot by tapping the pin. Not only we know the place of a certain store and facility on a map, but also the less dense information such as a budget or an impression comes to be provided by doing so it. We show a figure of summary in figure 3.

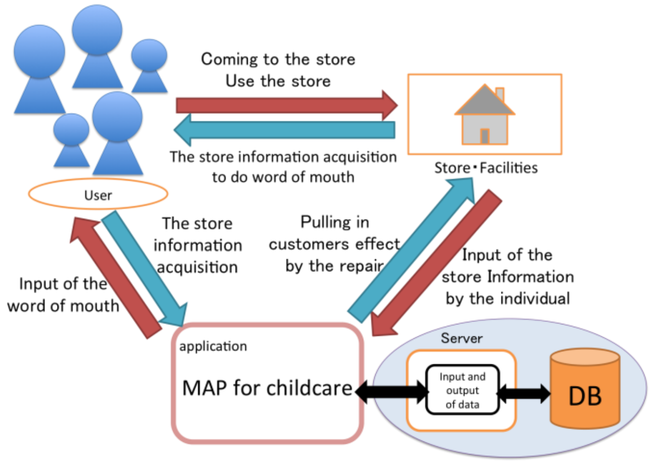


Figure 2: "MAP for childcare "constitution

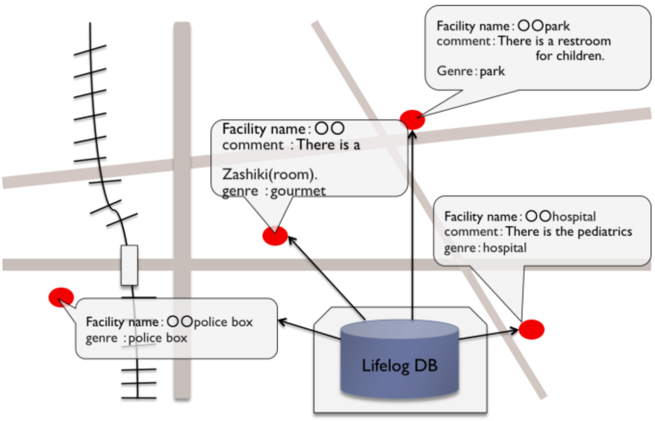


Figure 3: "MAP for childcare "summary

3.1 Interface

We show a figure of the screen transition in figure 4 when we use "MAP for childcare". It changes to "a map indication screen", "favorite indication", "the indication of the support store of HAGUMIN", " a list of age-specific information", "a list of genres" than a home screen and performs each function. It changes from a screen displaying a map to a screen outputting a screen inputting human log. Citizens input the data in the human log. In addition, only the pin which enrolled as a favorite from a home screen shows and changes to the function that only the pin of the specific spot that registered the kinds (genre) of facilities such as a hospital or a restaurant with human log displays.

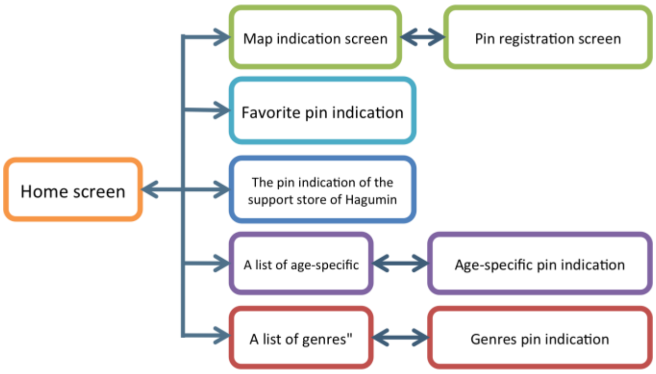


Figure 4: "MAP for childcare "screen transition

3.2 Home screen

When the application starts, it displays the home screen first (figure 5). We locate "a map", a "favorite" "list of genres", "support store of HAGUMIN", "list of age-specific information" button on a screen. We located the button and the text using the Linear Layout method of the xml language. It changes to each screen by tapping this button.

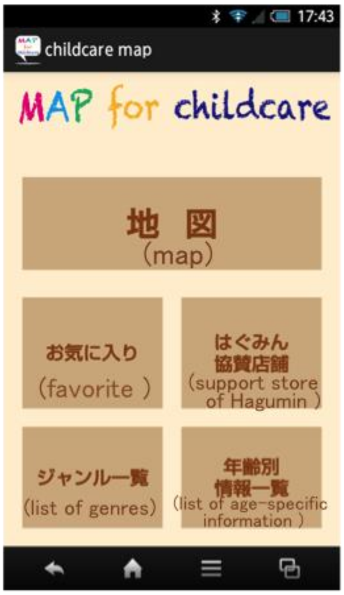


Figure 5: Home screen

3.3 Map indication screen

We let you display the map, which limited a range to Toyota city on a screen. We use Google Maps Android API v2 for indication. Google Maps Android API v2 is API to bury Google Maps on the smartphone application that Google provides, and specifications are largely changed by former Google Maps Android API v1 and can turn on a map when it is round with two fingers on a map and swipe. In addition, we came to be able to have 3D indication [28]. We show a map indication screen in figure 6. That a pin is displayed on a map, and change to the page that the detailed information of facilities can read when tap this pin; is functioned. When long shot taps any spot when one wants to send information, it changes to a screen inputting the pin information of the spot.

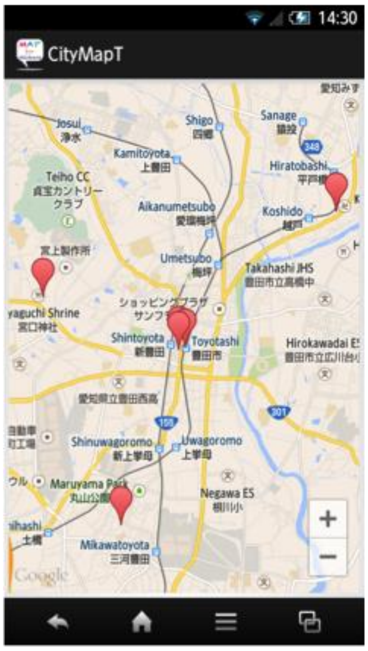
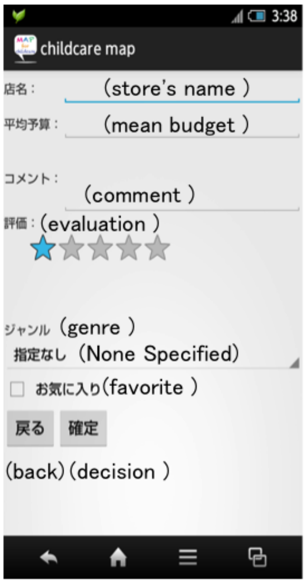


Figure 6: Map indication screen

3.4 Pin registration screen

This is the screen which inputs detailed information when we register a pin (figure 7). The information to input is "a store's name", "a mean budget", "comment", "an evaluation", "a genre", "a favorite" from the top. A genre includes "a hospital", "a police box", "a restroom", "a park", "a gourmet", "a baby outfitter". In addition, detailed information does not yet have a column choosing "age-specific information" now, but is going to add it in future. In the age-specific information, we choose the age of the child whom the information that we input is necessary for. It is in this way that we arranged the width of the choice to a user to become easy to do choice of the information more than a genre.

After filling each in, and having chosen it, a pin sticks in the spot that return, long shot tapped to a map indication screen when they tap a "decision" button (figure 8). It changes to the page that the information that we input with a pin registration screen some time ago by tapping this can read.


3:38

childcare map

店名：（store's name）

平均予算：（mean budget）

コメント：（comment）

評価：(evaluation）

ジャンル（genre）

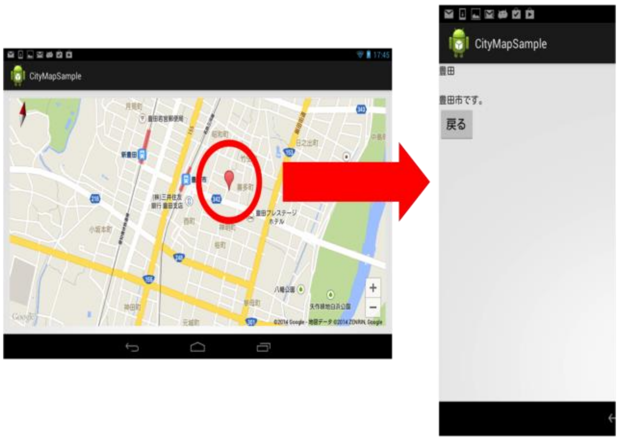
指定なし（None Specified)

□ お気に入り(favorite）

戻る 確定

(back)(decision）

Figure 7: Input screen of the word of mouth information


CityMapSample

豊田

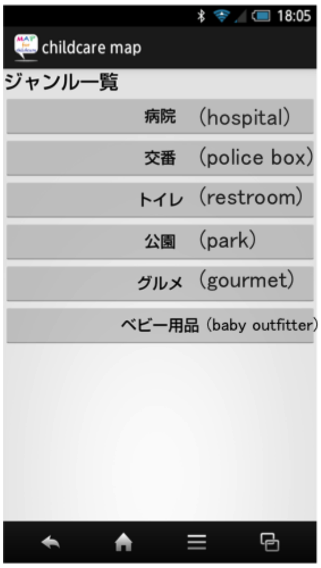
豊田市です。

戻る

Figure 8: Indication screen after the registration

3.5 Other functions

"A genre", the column of "a favorite" is established in the above-mentioned pin registration screen and can register the facilities of the spot. "a list of age-specific information " lets the function like " a list of genres " last. The facilities which we registered change to the map screen which we limit only the pin of the spot and can display by tapping the "a favorite" "list of genres" button of the home screen. We display a map to limit the pin of the facilities which we registered with the genre, and displaying only the spot that checked to "a favorite" with a pin registration screen when user tap a "favorite" button, and changing to a list of genres screen when we tap a "genre" button, and tapping each genre button (figure 9).



18:05

childcare map

ジャンル一覧

病院 (hospital)

交番 (police box)

トイレ (restroom)

公園 (park)

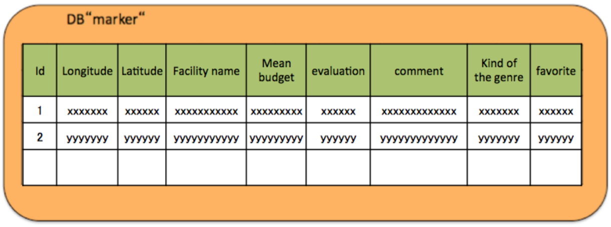
グルメ (gourmet)

ベビー用品 (baby outfitter)

Figure 9: List of genres screen

4 THE SOFTWARE CONSTRUCTION METHOD

We will explain the software construction method in this chapter. We assumed that we implemented it for the human log this time and stored information as a database instead in SQLite. SQLite is a lightweight database put on Android as standard equipment and can use it without servers easily [10]. We make database named "marker" storing id, longitude, latitude, facilities name, a mean budget, an evaluation, word of mouth, a genre number, pin information called the favorite presence using this SQLite (figure 10). When it changed, this database is called by a map indication screen and receives id of each pin, longitude, latitude. We let the cause display a pin at longitude and latitude. When we let you display it, we make a hash. The hash is the class, which can deposit and withdraw one value as a key with a certain value. When we put data, it stores a value to become a value and the key, which we want to store away together. The value that it stored by giving the value of the key when we want to give data is got. We show a brief summary in figure 11. It stores id of the pin as a key at the longitude of the pin, latitude using this hash. And when the pin was tapped, we send a key for the longitude of the pin, latitude and receive id. We send the id to the facilities information reading screen and change. We receive id data from a map screen if it changes to facilities information reading screen and am connected to the database. We take data of the store information as id from a database and compare the id data which we took as id of the database side from a map screen and output the store information of the line of agreed id. A string can post the store information data of a pin and the database displayed by a map screen by doing it this way. On the contrary, when we pushed the decision button with a pin entry screen when we register a pin, we send the information of the pin to the human log and start a map indication screen from 1 once again.



DB“marker“

Id	Longitude	Latitude	Facility name	Mean budget	evaluation	comment	Kind of the genre	favorite
1	xxxxxxx	xxxxxx	xxxxxxxxxxx	xxxxxxxxx	xxxxxx	xxxxxxxxxxxxx	xxxxxxx	xxxxxx
2	yyyyyyy	yyyyyy	yyyyyyyyyyy	yyyyyyyyy	yyyyyy	yyyyyyyyyyyyy	yyyyyyy	yyyyyy

Figure 10: Human log database

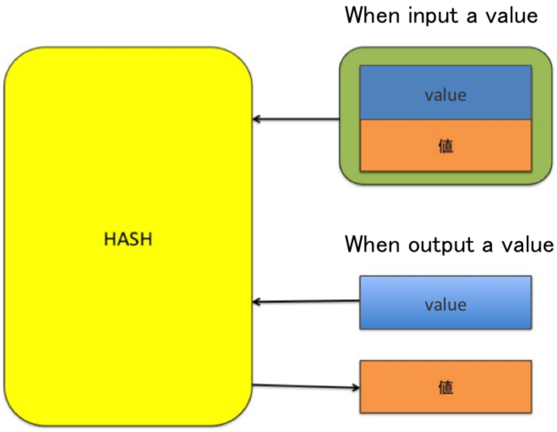


Figure 11: Summary of the hash

5 DEVELOPMENT TO A HUMAN PROBE

We design current "MAP for childcare" as independent smartphone application. On the other hand, we include an independent type dispersion of the human probe in the study of Yusuke Sugimoto and others [11] and we are unified with a system called this AHLE and am going to push it forward in future.

Feelings, local specialization type Autonomous System (Autonomous Human probes system with Local and Emotion functions) collects word of mouth and gathers it and analyzes it and we analyze it and realize it and am comprised of four elements shown below.

(1) The human log that stores away the word of mouth
(2) The analytical engine which parses data
(3) The interface that performs a human log and the mediation between clients
(4) The acquisition of the word of mouth and the client whom we display

We show a figure of summary of AHLE in figure 12. Notation called the human log is done not word of mouth during a figure here, but this is because AHLE assumes the collection of various human log and utilization as well as word of mouth.

We use data model showing in table 1 now in AHLE. These data model made it in reference to the article [12] of Nakamura and others. The Nakamura and others classify each data included in the human log from the viewpoint of 5W1H and can drop different human log into one data model.

We will implement this AHLE in future to MAP for childcare and, besides, analysis of the word of mouth classifies extraction of the feelings of word and performs the classifications of childcare support facilities using feelings word. We thereby inspect whether the user of plural people can provide the information that each user demands from a variety of childcare information.

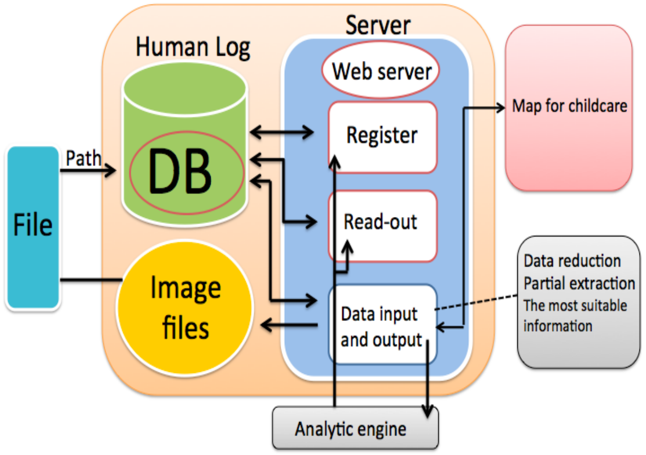


Figure 12: Summary of AHLE

Table 1: Comparison of the data model

Standpoint	Nakamura model	Proposed model	Note
WHEN	date	date	date
	time	time	time
WHO	user	user	user
	party	party	party
	object	——	object
	——	target	target
WHERE	location	——	location
	——	form Location	form Location
	——	to Location	to Location
HOW	application	application	application
	device	device	device
WHAT	content	——	content
	ref_schema	——	ref_schema
	——	category	category
	——	description	description
	——	picture	picture
	——	evaluation	evaluation
	——	emotion	emotion
WHY	——	——	(unused)

6 CONCLUSION

In this study, we designed smartphone application for childcare support "MAP for childcare". We plan visualization of childcare support carried out in the city by this smartphone application. We thereby try improvements such as the limited recognition of problems regarding childcare when going out and the childcare support of the city by the Toyota city citizen. The citizen who has a child by being a chance to know the policy of Toyota city in the process when the "MAP for childcare" that we use it and do childcare which we produced in this study collects reporting becomes easy to use for childcare support that the city performs positively and assumes it when a town may be full of pulling in customers effects by the word of mouth.

As a future problem, we perform the proof through the implementation "MAP for childcare" and improve the smartphone application based on problems that are identified with the usage. In addition, we let human log data accumulate by implementing AHLE and perform the classifications of childcare support facilities using feelings word of the word of mouth. In this way inspect whether there are many users, people can provide the information that each user demands from a variety of childcare information. As it stands, the smartphone application allows for only one to one to incorporate information of the one in one spot, but in the future the goal is to have many users input word of mouth of one spot. As the users of the smartphone application will have children with them, we hope to further develop the application for ease of use both in terms of style and practicality.

ACKNOWLEDGEMENTS

We would like to thank All of Toyota city government office child Department and Toyota citizens for cooperate.

REFERENCES

[1] Statistics Bureau of the Management and Coordination Agency, 2010 state of the nation survey, Japanese Statistics Bureau (2010).

[2] Attitude survey about Cabinet Office, a 2013 family and the childcare in the area, Cabinet Office government unification official (2014).

[3] Ayaka Nishiwaki, Takaaki Hishida, Tadanori Mizuno: Basic Study For The Construction Of A System For Parent-Child Cooperation Oriented Child Care Support Service, Workshop on Informatics 2013(WiNF2013), pp.180- 184(2013.12).

[4] Nobuaki Ohmori, Ayako Taniguchi, Rikutaro Manabe, Yoshihiko Terauchi: What Barriers Are Mothers with Small Children Facing When Participating in Out-of-Home Activities? http://library.jsce.or.jp/jsce/open/00039/200906_no39/pdf/263.pdf (2014/6).

[5] Toyota city, Aichi general affairs department general affairs department, Toyota city statistics book 2011 version.

[6] Toyama childcare support http://www.city.toyama.toyama.jp/special/child.html (2014/1).

[7] Aichi childcare home courtesy card business, http://www.pref.aichi.jp/kosodate/card/(2014/1).

[8] The Toyota city child Ordinance http://www.city.toyota.Aichi.jp/division/ak00/ak01/1194139_7170.html (2014/1).

[9] Toyota City child comprehensive plan (new Toyota child smile plan) http://www.city.toyota.Aichi.jp/division/ak00/ak01/1209206_7170.html (2014/1).

[10] Basic operation of the SQLite database MONOist http://monoist.atmarkit.co.jp/mn/articles/1209/13/news001(2014/1).

[11] Yusuke Sugimoto, Tadanori Mizuno, Hishida Takaaki: Classification of tourist sites using the emotional words in user reviews, dispersion, cooperation and mobile (DICOMO2014) symposium (2014).

[12] Masahide Nakamura, Akira Shimojo, and Hiroshi Igaki: Considering Common Data Model to Mash-up Databases, the Institute of Electronics, Information and Communication Engineers, technology research report, Vol. 109, No. 272, pp. 35-40 (2009).

A Proposal for CS Unplugged Utilizing Regional Materials

Hisao Fukuoka[†], Akane Kawakami[‡], and Yuji Wada[†]

[†]Tokyo Denki University, Japan
[‡]WACOM-IT Co.Ltd., Japan
fukuoka@mail.dendai.ac.jp

Abstract - This paper proposes CS Unplugged utilizing regional materials. CS Unplugged is a method of teaching computer related technology to children, who are in general lacking in mathematical or some other scientific background. The feature of CS Unplugged is that it does not use actual computers in its teaching process. Instead of using computers, it teaches computer related technology through some desktop play activities. We propose those activities utilizing regional materials. When we practice CS Unplugged activities in elementary or junior high schools, the target pupils are local children in general. Using regional materials in the activities can make them more friendly to the children. In this paper we discuss three directions for utilizing regional materials in CS Unplugged in detail. We also show the concrete example of activity that successfully utilizes the geographical feature of Matsue City where one of the author's former campus is located. In this activity we regard the geographical feature of Matsue as one of its regional materials.

Keywords: Computer Science Unplugged, Teaching Method, Regional Materials

1 INTRODUCTION

The lack of interest in computer science among children has increased in recent years and the aversion of young people to enroll in information system curricula at universities is a growing problem. To reverse these trends, some means of making computer science fun to learn and easy to understand for children is essential.

One such means is Computer Science Unplugged (CS Unplugged), an educational method comprised of desktop play activities, such as board games and so on, without the use of computers[1]. In other words, children can learn computer related technology indirectly. We believe that it is an appropriate teaching approach to children, because they are so lacking in mathematical and some other scientific background that it is difficult for them to learn computer related technologies directly.

CS Unplugged has originally been proposed by Timothy Bell of University of Canterbury, New Zealand. According to the homepage of CS Unplugged, around 20 activities are already defined. Twelve activities of them have been translated into Japanese by Japanese researchers and published[2]. In Japan, the subject ”Information” has been compulsory since 2003 in high schools and the subject ”Measurement and Control by Programming” has been compulsory since 2012 in junior high schools. These trends mean the importance of teaching computer related technology to relatively young generation. Some junior high school and senior high school teachers have tried to apply the CS Unplugged approach in their classes and have reported their experiences[3][4].

Although it is a promising teaching method, the current CS Unplugged activities are considered to be inadequate in terms of the coverage of technologies. Some new activities corresponding to technologies that have not been addressed in the manner of CS Unplugged have to be developed. These new activities must be friendly to children in order for them to be effective. The activities are practiced by the pupils in elementary schools or junior high schools, most of whom are in general local children. In this sense, utilizing regional or local materials as the content of these activities is considered to be effective, because the children must feel familiar with them.

In this context we have already considered three directions of utilizing regional materials in the development of CS Unplugged activities: Direction 1, Direction2 and Direction 3[5]. Direction 1 replaces the content of the existing activities with some regional materials. Direction 2 utilizes regional materials in a completely new activity corresponding to the technology not addressed so far. Direction 3 develops a new type of activity corresponding to the technology already addressed, in which we can utilize some regional materials.

As for Direction 2, we have developed the new activity in which children can learn the phenomenon of ”Bottlenecks” [6]. In this activity we have utilized the geographical feature of Matsue City in Japan. The activity is two-player board game and through tackling this game children can understand the phenomenon of bottlenecks and learn how to avoid the occurrence of them.

The rest of this paper is organized as follows. Section 2 describes the outline of CS unplugged. Section 3 discusses above three directions in detail with some concrete examples, especially for Direction 1 and Direction 3. From Sections 4 through 6, the details of the newly proposed activity are shown. Section 7 presents the feedback from children. Finally, conclusions are drawn in Section 8.

2 CS UNPLUGGED

CS Unplugged is a method of computer science learning originated by many people, especially Tim Bell of the University of Canterbury in New Zealand, and currently includes around 20 activities as shown in Table1. Each of these activities addresses a computer science concept that is taught in specialized courses at senior high school and college, but it is designed so that elementary and junior high school students can understand the concept while enjoying the performance of a game, magic trick, body movements, or other form of

Table 1: Activities in the Japanese-language teachers' guide

No.	Title	Concept
1	Count the dots	Binary numbers
2	Color by numbers	Image representation
3	You can say that again !	Text compression
4	Card flip magic	Error detection and correction
5	Twenty guesses	Information theory
6	Battleships	Searching algorithms
7	Lightest and heaviest	Sorting algorithms
8	Beat the clock	Sorting network
9	The muddy city	Minimal spanning trees
10	The orange game	Routing and deadlock in network
11	Treasure hunt	Finite-state automata
12	Marching orders	Programming languages
13	The poor cartographer	Graph coloring
14	Tourist town	Dominating sets
15	Ice roads	Steiner trees
16	Sharing secrets	Information hiding
17	The Peruvian coin flip	Cryptographic protocol
18	Kid Krypto	Public-key encryption
19	The chocolate factory	Human Interface design
20	Conversations with computers	Turing test
21	The intelligent piece of paper	Artificial intelligence

activity. The CS Unplugged activities are characterized by features such as the followings[7].

- Learning at play
- Learning by trial and error with concrete physical objects and bodily sensations
- Learning in groups
- Ease of execution
- Learning through interwoven story elements

A teacher's guide for CS Unplugged is publicly available and can be downloaded free-of-charge from the official CS Unplugged website[1].

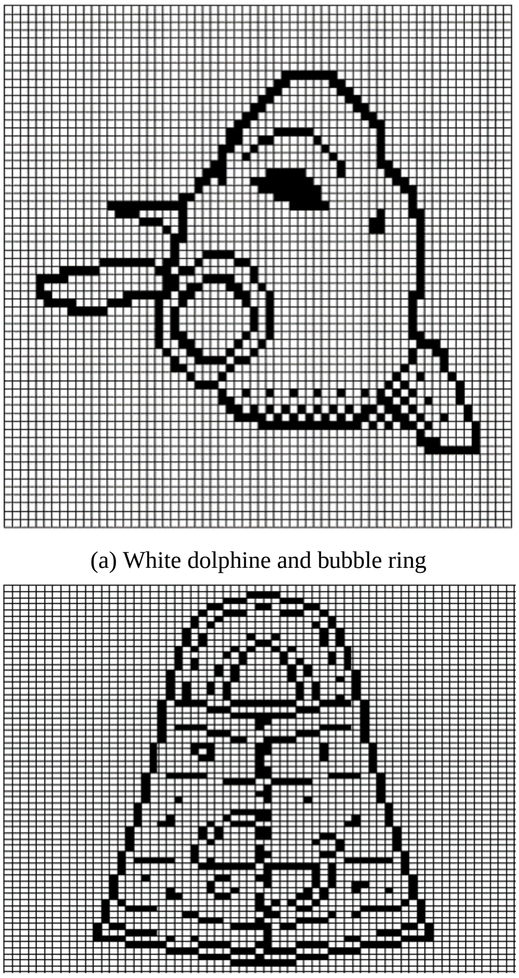
A Japanese translation containing the first twelve activities of the activities shown in Table 1 was published in Japan in 2007[2].

3 DIRECTIONS FOR UTILIZING REGIONAL MATERIALS

We have considered three directions in developing CS Unplugged activities utilizing regional materials: Direction 1, Direction 2 and Direction 3[5].

3.1 Direction 1

Direction 1 replaces the content of the existing activities with some regional materials. One example of Direction 1 is Activity 2, in which children learn a run-length representation of images. Children learn how to transform given images to their run-length representation and vice versa. As those given images, we can easily introduce the images that represent some local materials. Figure 1 shows some examples. These images represent the materials local to Shimane Prefecture.



(a) White dolphine and bubble ring

(b) Dotaku(Bell-shaped vessel)

Figure 1: Images local to Shimane

Another example of Direction 1 is Activity 3, in which children learn data compression technique similar to Lempel-Ziv compression algorithm. In this activity, children are given some short sentences such as poems that have many repeated phrases. We can easily replace these sentences with the ones locally well known, such as lullabies and folk songs, in some specific region.

The effectiveness of Direction 1 should be confirmed through several experiments, and we are now in planning phase of them.

3.2 Direction 2

For Direction 2, we utilize some regional materials in a completely new activity corresponding to the technology not addressed so far.

As an example of Direction 2, we have already developed the new activity in which children can learn the phenomenon of "Bottlenecks" frequently encountered in various computer systems[6]. As far as we have investigated, there is no activity that deals with bottlenecks. In this activity we have utilized the geographical feature of Matsue City in Japan. The geographical feature can be one of the regional materials. The activity is two-player board game and through tackling this game children can understand the phenomenon of bottlenecks and learn how to avoid the occurrence of them. In the following sections we will show the details of this bottleneck activity and also show the results of practicing this activity to the children in a junior high school and a college of technology, both in Matsue City.

We are also developing another new activity that deals with ”Scheduling”. In this activity we try to utilize ”Yamata-no-Orochi(Eight-headed dragon)” legend in Izumo Myths, a regional material in Shimane Prefecture.

3.3 Direction 3

Direction 3 develops a new type of activity corresponding to the technology already addressed, in which we can utilize some regional materials. Although we believe it must be a possible direction, we have not realized any concrete examples of Direction 3 and need some more studies.

4 TARGET CONCEPT OF NEW ACTIVITY

The concept addressed in the new activity is a computing bottleneck[8]. It is believed to be a new activity, as a preliminary survey of the current CS Unplugged activities did not show the same concept.

4.1 Computing bottlenecks

In computer science, a bottleneck acts as a constraint on communication via a medium that links two points and thus presents a barrier to increasing the processing speed of a computer or the communication speed of a network beyond that of the bottleneck itself. In von Neumann computers, von Neumann bottlenecks tend to occur because of the difference between the processing speeds of the CPU and the computer memory or because of insufficient width in the bus that links them. These problems limit or decrease the processing efficiency of computers.

4.2 Geographical bottleneck in Matsue City

Matsue City, where one of the author’s former campus is located, is in Shimane Prefecture, Japan. Its geographical features, and in particular its river-straddling configuration, are well suited for illustration and explanation of the bottleneck concept.

Matsue City lies in a region between two bodies of water called Lake Shinji and Nakaumi Lagoon, and its urban center is divided into northern and a southern district by the Ohashi-gawa River, which flows from Lake Shinji to Nakaumi Lagoon. The river is shown in Figure 2. The northern district is called Kyohoku and the southern district is called Kyonan. Until the early Showa period, they were connected only by the Matsue Ohashi bridge. Today they are connected by the Shinjiko Ohashi, Matsue Ohashi, Shin Ohashi, Kunibiki Ohashi, Enmusubi Ohashi and Nakaumi Ohashi bridges. The Enmusubi Ohashi bridge, the latest one, was constructed in 2013 to relieve the congestion occurring in the traffic attempting to cross the older bridges during the morning and evening rush hours.

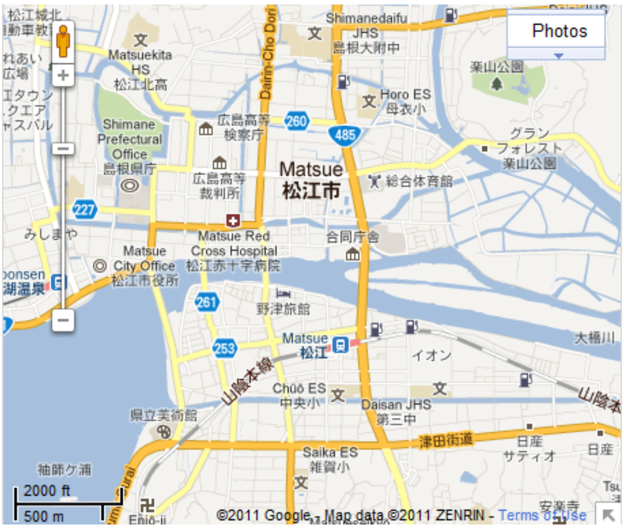


Figure 2: Ohashi-gawa River vicinity in Matsue City

The proposed activity is a board game modeled on these geographical features of Matsue City and designed to help students gain a clear experience-based understanding of the bottleneck concept.

5 DETAILS OF THE ACTIVITY

5.1 Overview

In this activity, two students compete against each other in a board game to see which one’s pieces can reach the goal more quickly. The learning materials are the board and the pieces. In the game, they experience bottlenecks and devise strategies to relieve them.

5.2 Learning materials and rules

Figure 3 shows the following learning materials used in this activity: the board, eight pieces, and eleven bridge-building cards.

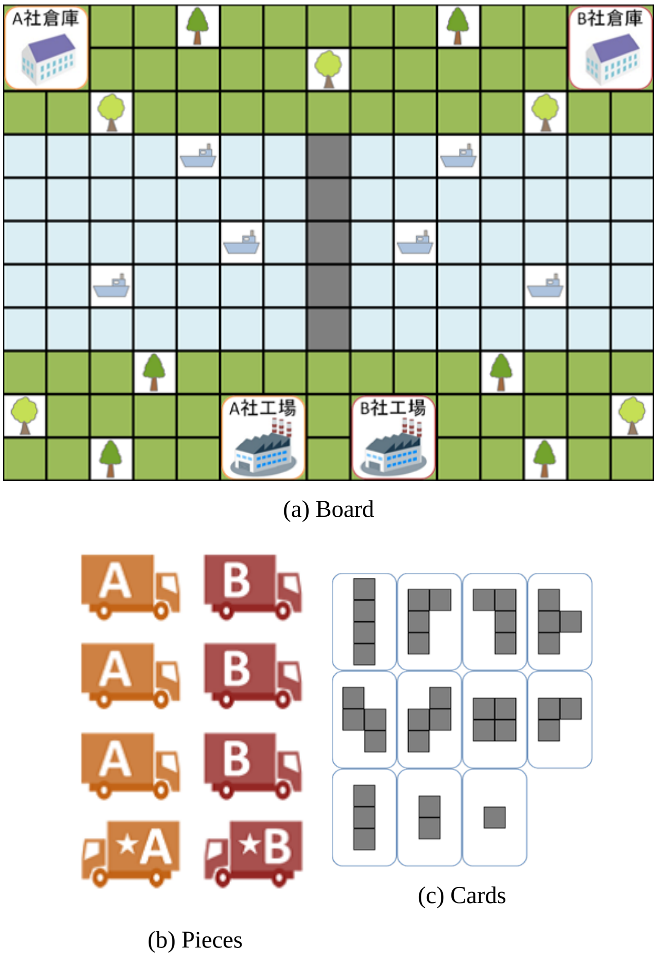
5.2.1 Board

The board represents a city divided into north and south sides separated by a river, much like Matsue City. As shown in Figure 4, the board comprises 165 squares (11 vertical and 15 horizontal) of six types, either singly or in combination, to represent the land, a river, one or more bridges, two warehouses, two factories, and twelve obstacles (trees and boats).

Each of the warehouses and factories occupies four squares. One warehouse and one factory belong to Company A, and the others belong to Company B. Before starting the game, the students choose between Company A and Company B. The student taking Company A makes the first move. For each student, the warehouse of his or her company is both the starting point and the final goal.

5.2.2 Pieces

Each student has four pieces, which represent trucks that transport production materials and products between the warehouse



(a) Board

(b) Pieces

(c) Cards

Figure 3: Learning materials

and the factory of his or her company. The students take turns, with each student moving all four of his or her pieces per turn throughout the game. Each piece can be moved just one space per turn. They start on the warehouse squares, with each piece representing a truck loaded with production materials. When a piece reaches the factory squares, it is turned over to represent a truck loaded with the products which, on the next turn, sets out for the company warehouse–the final goal.

As illustrated in Figure 5, a piece can be moved from the square it occupies to a land, warehouse, factory, or bridge square on any of its four sides, but cannot be moved diagonally or onto any square representing the river, an obstacle, or the other student's factory or warehouse, nor to any occupied square other than those of the student's own factory or warehouse.

During one turn, the student must move each of the four pieces one space unless if all adjacent squares are already occupied or represent the river, obstacles, or the other company's facilities.

5.2.3 Bridge-building cards

Each bridge-building card shows at most four squares arranged in a different block configuration, such as those shown in Figure 3(c). Before the game, each student chooses three of the cards. During the game, when he or she cannot move any of his or her pieces, the student in turn may use one of his or her own cards to mark river squares in the same block form as shown on the card. The pieces of each student can be moved to any bridge addition space marked by either student.

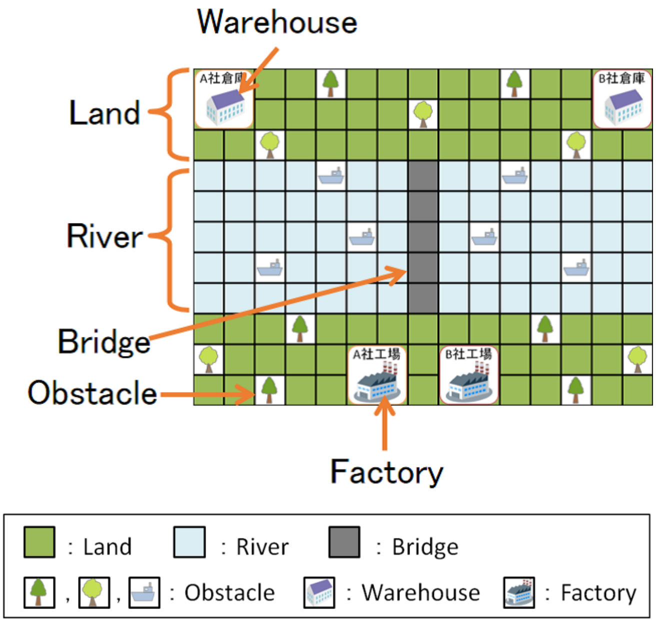


Figure 4: Board components

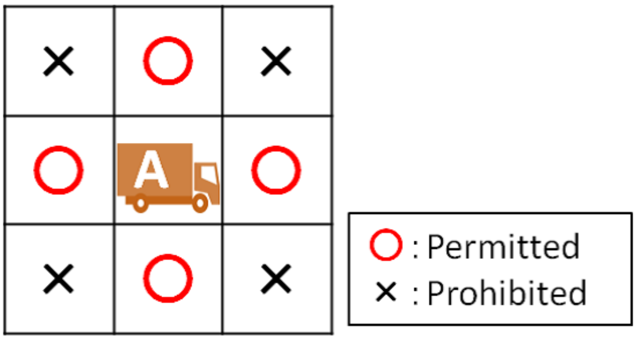


Figure 5: Permissible moves

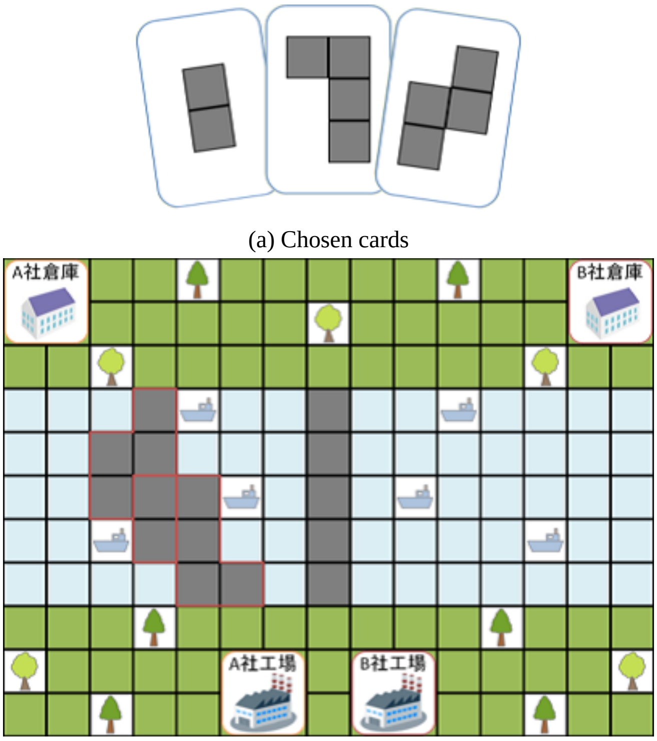
Assuming that a student chose three cards as shown in Figure 6(a), one of the possible bridge configurations during the game can be shown in Figure 6(b). In this case, the student has used a card three times.

6 TEACHING PROCESS

The overall process for the student experience and learning of the bottleneck concept begins with an introduction by the teacher, followed by playing a practice game and then the actual game by the students.

6.1 Introduction

The teacher first tells the students a story about the two adversarial companies, A and B, which are mutually competing to perform faster operations. For each company, the operations require the transport of production materials from its warehouse to its factory on trucks that must then transport the



(a) Chosen cards

(b) A possible bridge configuration

Figure 6: Bridge-building cards and their usage

factory products back to the warehouse because they cannot be stored at the factory. In both cases, the warehouse and the factory are separated by a river that is currently spanned by only one bridge, and their trucks therefore cause traffic congestion in that area. In this way, the story provides the setting for the game.

6.2 Practice

After hearing the story, the students first perform a practice game without the bridge-building cards. The board therefore remains in its initial state throughout the game with no increase in bridge squares, as illustrated by the interim stage of the practice game shown in Figure 7, in which the trucks of the two companies have caused a traffic jam as they attempt to move onto the bridge while driving toward their respective factories. In playing the practice game, the students gain their initial experience in the bottleneck problem in the form of traffic congestion at the bridge between the two land areas.

6.3 Game

After completing the practice game, the students next proceed to the actual game, with the use of the bridge-building cards to make bridge additions prior to the game itself. Figure 8 shows an interim stage of one such game. As the game proceeds, the students experience first-hand the effects of these bridge additions in relieving traffic congestion and enabling smooth progress toward the final goal.

In Figure 8, the students have increased the number of bridges from one to three, which is comparable to the widening of a bus connecting CPU and memory to relieve a von Neumann bottleneck. These three bridges are basically single-lane bridges, but the addition of a ”siding” such as that in the center bridge enables trucks moving in opposite directions to pass each other, which was not possible on the original bridge.

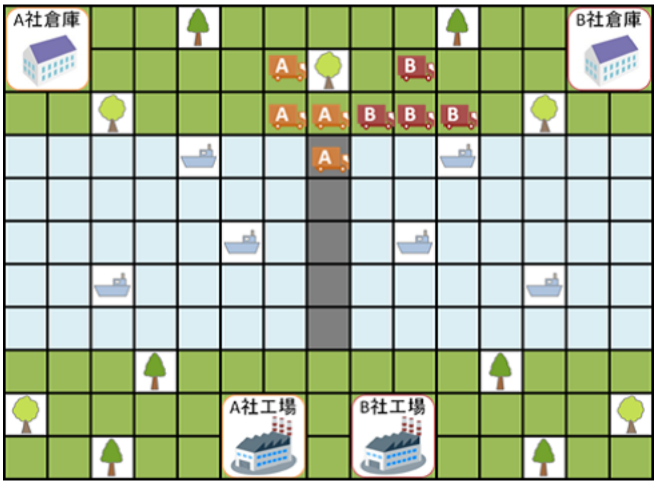


Figure 7: A traffic jam in the practice game

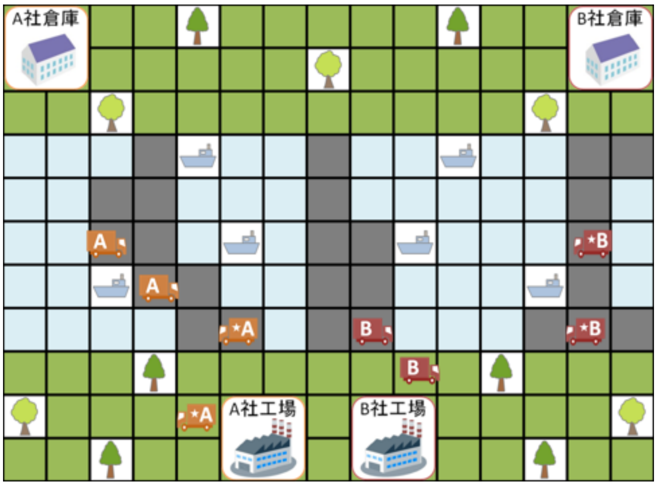


Figure 8: A game in progress with alleviated traffic congestion

The outcome of each game is governed largely by the width and shape of the bridge additions constructed by both students, which leads each of them to develop and try out various bottleneck mitigation strategies in successive games.

7 EXPERIMENTS

We conducted three experiments for the ”Bottleneck” activity.The specifications of these experiments are shown in Table 2. A snapshot of the third experiment is shown in Figure 9.

From the first experiment through the third, the activity had been gradually enhanced. The primary enhancement was to give users more choices of tactics to win the game, such as increasing the number of pieces and card patterns and so on.

There were thirty four participants in total, and thirty two of them had not had any knowledge of ”Bottlenecks” before the experiments.

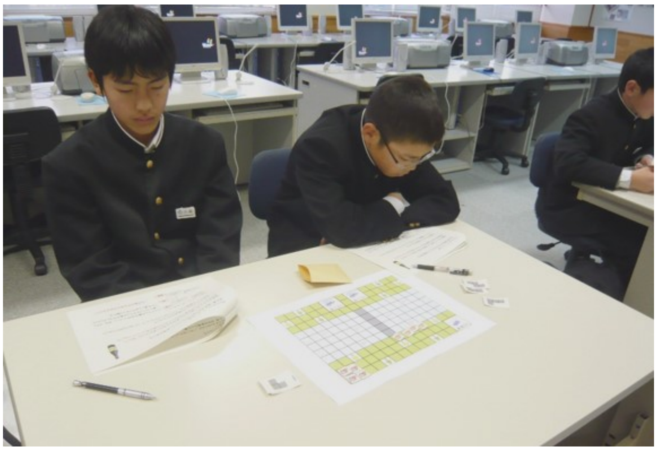


Figure 9: Snapshot of the third experiment

Table 2: Specifications of three experiments

Grade	Number of children
First graders of college of technology (15-16 years old)	7
First graders of college of technology (15-16 years old)	19
First graders of junior high school (12-13 years old)	8

After each experiment, a questionnaire was give to the participants. The aim of the questionnaire was to investigate the interest of the children and to determine if they understood what the bottlenecks is. The most important question was ”Did you understand what the bottleneck is?” The children were given four choices for this questions: yes, rather yes, rather no and no.

Table 3 shows the result for this question. We did not have any negative answers. In addition, we observed that the children really enjoyed the activity and tackled it positively.

Although the scale of our experiments was small and we only obtained the subjective evaluations, we believe that the activity could be effective in learning the concept of bottlenecks.

8 CONCLUSION

In the study described here, we have proposed CS Unplugged utilizing regional materials. We have set three directions for utilizing regional materials in CS Unplugged activities.

Following the Direction 2, we have constructed a new activity designed to increase children’s understanding of computer science through an enjoyable learning experience, with the ultimate objective of contributing to the increase of CS Unplugged adoption and use.

The target concept of the proposed activity is the computer bottleneck. For the learning experience, we designed a board game and materials in which the board is modeled after the geographical features of Matsue City, where one of the author’s former campus is located, and, in particular, on its bottleneck-inducing configuration.

In order to evaluate the activity, we conducted three experiments at a junior high school and a college of technology. The participants of the experiments showed their interest in the activity and gave us rather positive response in terms of understanding level of bottlenecks.

The cumulative results of the evaluations and trial-and-error modifications will be applied to the intended achievement of the activity in which children can learn the bottleneck concept more easily and enjoyably.

Table 3: Understanding level of participants

Did you understand what the bottleneck is?		
Answer	Number	Percentage
Yes	24	71
Rather yes	8	24
Rather no	0	0
No	0	0
No answer	2	6

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 25350275.

REFERENCES

[1] T. Bell, Computer Science Unplugged Activities, http://csunplugged.org/activities (2014.6.10).

[2] T. Bell, I.H. Witten and M. Fellows (Translation supervised by S.Kanemune), Computer Science Unplugged (Japanese Version), e Text Lab. Inc.(2007).

[3] S. Kanemune, R. Shoda, S. Kurebayashi, T. Kamada, Y. Idosaka, Y. Hofuku, and Y. Kuno, An Introduction of Computer Science Unplugged - Translation and Experimental Lessons -, Proceedings of Summer Symposium in Suzuka 2007, pp. 5-10 (2007) (in Japanese).

[4] T. Nishida, Y. Idosaka, Y. Hofuku, S .Kanemune, and Y. Kuno, New Methodology of Information Education with ”Computer Science Unplugged”, Proceedings of the 3rd International Conference on Informatics in Secondary Schools–Evolution and Perspectives:Inormatics Education–Supporting Computational Thinking(ISSEP’08), pp. 241-252(2008).

[5] H. Fukuoka, A Study on CS Unplugged Utilizing Regional Materials, The 76th National Convention of IPSJ, 4F-2, Vol.4. pp. 371-372(2014)(in Japanese).

[6] A. Kawakami, and H. Fukuoka, A New CS Unplugged Activity for a Learning EXperience in Bottlenecks, Proceeding of the 1st International Symposium on Technology Sustainability (ISTS2011), CIT-015, pp. 45-248(2102).

[7] T. Nishida, and S. Kanemune, Learning Computer Science with Fun, IPSJ Magazine, Vol.50, No.10, pp. 980-985(2009) (in Japanese).

[8] http://en.wikipedia.org/wiki/Bottleneck (2014.6.10).

Session 4: Intelligent Systems and Applications

(Chair : Yoshitaka Nakamura)

Estimating Babies' Awakening-time at Night Improving Childcare Support

Rina KANAZAWA*, Ikuma SATO**, and Yuichi FUJINO*

*Graduate School of System Information Science, Future University HAKODATE, Japan
**Future University HAKODATE, Japan
{g2113009, ikuma-is, fujino}@fun.ac.jp

Abstract - Recently, a childcare environment has been changing due to increase in the number of nuclear family. Therefore, parents have to take care their child by only themselves. We focused on a babies crying at night. We believe it is possible to estimate a baby's crying and awakening-time at night by changing his/her sleep rhythm. We propose a method of reducing a mother's stress by informing her about her baby's sleeping rhythm and awakening-time at night. We tried to estimate babies' awakening-times and sleep rhythms at night by using sensors. We conducted three experiments to detect babies' sleep rhythms through the cooperation of young parents. We confirmed the presence of body movements, awakening and crying at night by using video and the sensor-mat data. We analyzed the data and detected breathing rates, heart rates and body movements by fast Fourier transform and digital filtering. We also classified active and quiet sleep based on body movements. This study showed that a baby's sleeping condition at night could be evaluated by detecting the breathing, heart rate and body movements. We were able to detect babies' breathing rates with or without body movements; however, their breathing rates did not change before or after the awakening. We estimated the babies' awakening-times after a few sleep cycles, which shortened sleep just before awakening.

Keywords: childcare support, sleep rhythm of baby's, night crying, awakening, sleep sensor

1 INTRODUCTION

1.1 Stress on Childcare

Many young mothers become nervous due to the by stress of childcare. Parents are under stressful in their first childcare; especially stress is large to a mother. We'd like to think about a stress and it's influence only for a mother in her childcare. A few decades years ago, parents of a married couple and neighbors sometimes help young mothers with their childcare. However, a young mother has to take care of her child alone today. Therefore, a young mother may become nervous. At worst, they may mistreat their babies [1].

There is a large amount of stress in the childcare [2]. This is especially evident in a mother's sleeping problems. We previously investigated mothers living in HAKODATE, Japan who were experiencing the stresses of childcare. Most said they experienced sleep problems at night. Specifically, the stress of most mothers was due to their babies crying. Therefore, it was necessary for mothers to have a good amount of sleep to reduce their uneasiness and stress.

Once a baby wakes up at night, it is difficult for the baby to go back to sleep. As a result, a mother lacks sleep and become stressed. A study found that large amount of an awakening hormone is secreted in adults when they are suddenly awakened [3]. Therefore, we believe this secretion causes stress in mothers.

The reason of a baby's crying and awakening at night is that his/her circadian rhythm has been not established [4]. The circadian rhythm for adults synchronizes to the earth's time cycle. Adults can distinguish morning from evening by establishing a circadian rhythm. However, babies cannot distinguish morning from evening because they have not yet established it. Therefore, a baby repeats sleep and awakening throughout the day. A baby's sleep is not defined in terms of light sleep (REM sleep) or a deep sleep (Non-REM sleep) through brain waves but in terms of active and quiet sleep through the differences in brain waves [5]. Active and quiet sleeps have certain characteristics such as change in breathing rate and body movements, in addition to the brain waves. These cycle are related to sleeping and awakening rhythms. We believe that babies awaken after a few active sleeps cycles.

1.2 Purpose

We believe that knowing a baby's awakening-time in advance will be reduce a mother's stress. The first childcare is the most difficult for all mothers because they cannot estimate their babies' awakening-times. However, they will be able to estimate them in their second or third babies because of their experiences. Therefore, we think that knowing a baby's awakening-time will be reduce a young mother's stress on first childcare.

However, there has not been researches involving recording a baby's sleep without attaching sensors or estimating a baby's awakening-time at night. In the traditional research, some recording method of baby's sleep is using a wrist type sensor or a mother's-write record [6][7][8]. It is stressful for mothers to attach sensors to their babies. Attaching a sensor to a baby may also disrupt his/her sleep. Therefore, measuring a baby's sleep with unconstraint sensors and estimating his/her awakening-time are effective for supporting young mothers.

We propose a method for measuring a baby's sleep rhythm without restriction sensors, explain the results of active and quiet sleep cycles from babies' breathing rates and body movements using our proposed method, and investigate their awakening-times at night.

2 METHOD

2.1 Baby's Sleep Rhythm

A baby's sleep rhythm is classified as active or quiet sleep [5]. They are similar to the REM or Non-REM sleep cycles of adults. Active and quiet sleeps have similarities in the change in breathing rate and body movements.

We estimated a baby's awakening-time by observing a baby's sleep rhythm, which includes observing the intervals change in breathing rate, heart rate and body movement interval. The breathing rate of adults changes when they are in a deep sleep. We think this is also the same for babies. We defined a baby's sleep rhythm through his/her body movements at night because we believe a baby's awakening-time is connected to body movements and sleep rhythms. Finally, we tried to estimate a baby's awakening-time by observing the breathing rate, heart rate and body movements analyzed from a baby's sleep data.

Our method does not restrict babies or burden mothers, which is the important for young mothers. In conventional studies, some methods were reported in which mothers have to record or write down her babies' sleep by themselves or they have to attach some sensors to their babies. Therefore, a sensor without restricting a baby's movements is necessary.

2.2 Sensor

We used a sensor mat (the sleepscan made by TANITA Co., Ltd) filled up with purified water and a microphone to detect sleeping vibrations; breathing rates, heart rates and body movements. The sensor mat was made specifically for adults, but we think it is appropriate detecting a baby's sleep rhythm. It can be used without attaching any materials to a baby. A sensor mat is put under the bedding. We also used a video camera to check the babies sleeping conditions at night.

2.3 Method

We processed the raw data obtained by the sensor mat using digital filtering and Fast Fourier Transform (FFT) and were able to detect babies breathing rates, heart rates and body movements. A baby's breathing rate is 25 ~ 45 times per minute and heart rate is 80 ~ 160 times per minute [9]. We used a high-pass filter and FFT to reduce the effects of body movements. We designed the high pass filter of 10th butter function (Eq. (1)). We set the cut-off frequency to 0.3 Hz because a baby's breathing rate is usually over 25 times per minute, which is 0.42 Hz. We also set the cut-off frequency to 1.3 Hz because a baby's heart rate is usually over 80 times per minute, which is mean 1.33 Hz.

$$H(z)=\frac{b(1)+b(2)z^{-1}+\ldots+b(n+1)z^{-n}}{1+a(2)z^{-1}+\ldots+a(n+1)z^{-n}} \qquad (1)$$

We observed babies' body movements through FFT and compared their sleeping video data with the sensor mat data obtained by FFT. We found the "power" obtained by FFT with body movements were over 70. The other part of the "power" less than 70 was estimated with no body movements.

We estimated babies' awakening-times from their changing of breathing and heart rates. A baby's breathing rates decrease when he/she is sleeping and increase when he/she is awakening. But, a baby's heart rate is not confirmed. We tried to confirm it when he/she is sleeping. Therefore, we estimated a babies' awakening-time using these information. We also examined the relationship between babies' awakening-times and sleep rhythms.

One sleep cycle of a baby is defined from the beginning of a quiet sleep to the end of an active sleep. It is said that an adults sleep cycle is from a beginning of Non-REM sleep to the end of REM sleep about 90 minutes per cycle, and repeats 4 ~ 5 times in one night. We believe a baby's sleep cycle is the same as an adult's. However, the time of one cycle is different. It is 40 ~ 50 minutes on from a newborn to a babyhood stage, and 50 ~ 60 minutes on over three months [10].

We also estimated a babies' awakening-times by changes of breathing and heart rates. We examined the relationship between awakening-times and sleep rhythms classified by the presence of body movements.

3 EXPERIMENTS

We measured three babies' sleep rhythms using the sensor mat, which is not disturbance on babies, and estimated their awakening-times at night. We detected their breathing, heart rates and body movements using the analyzed the data from the sensor's data. We classified active and quiet sleep using these information. We also estimated the babies' awakening-times using their breathing and heart rates and the classified sleep rhythms.

3.1 Experimental Environment

We measured babies' sleep rhythm at night using the sensor mat and video camera. Figure 1 shows a photo of the experimental environment. The sensor mat was placed under a futon. The video camera was set near the baby's bedside to check whether he/she was sleeping or not.

Table 1 lists the experiment conditions which shows one baby of the three times experiments. We were not able to gather correct data in other two babies because they did not cry at night or the sensor data was not suitable.

3.2 Analysis

We analyzed the babies' breathing and heart rates using digital filtering and FFT. The babies breathing rates were detected using a high-pass filter with a cut-off frequency of 0.3 Hz and FFT for 32 seconds. The babies' heart rates were also detected using the high-pass filter with a cut-off frequency of 1.3 Hz and FFT under the same conditions.

We classified active and quiet sleep rhythm from babies' body movements analyzed by FFT and estimated their awakening-times by changes in breathing and heart rates and the relationship between awakening and sleep rhythms from their body movements.

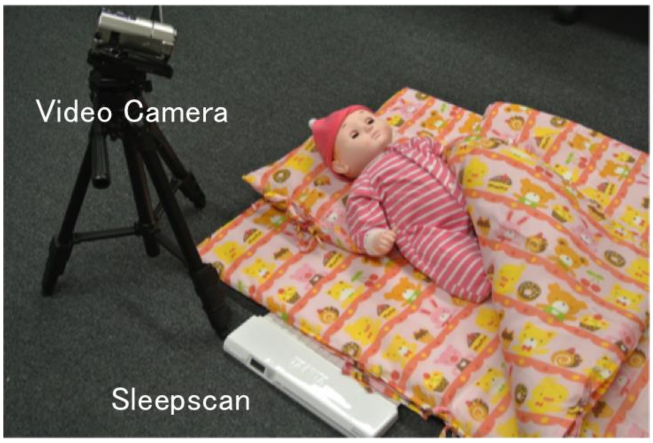


Figure. 1: Photo of experimental environment.

Table. 1: Experimental conditions.

	Baby 1
age	Three - six month
gender	boy
weight	5.9Kg-7.6Kg
data	①September, 27th, 2013 22:30～06:46 ②November, 6th, 2013 23:28～08:56 ③January, 28th, 2014 23:11～06:50

4 RESULT

We detected a peak in the frequency band of 0.33 ~ 0.75 Hz in the raw data of the sensor mat, which were processed the high-pass filter (a cut-off frequency of 0.3Hz) and FFT. Figure 2 shows baby 1's raw data, which were obtained by the sensor mat and classified babies' conditions at night using the video data. The baby woke up two times at night in the first experiment. In Fig. 2, two periods with sensing data are shown since the mother held her baby because he was crying at night. Figure 3 and 4 show the FFT results. Point "A" and "B" are 32 seconds. Point "A" in Fig. 2 shows the characteristic frequency with no body movements (Fig. 3) and point "B" shows that with body movements (Fig.4). The horizontal axis is frequency and the vertical is power in Fig.3 and Fig.4. A peak frequency of 0.438 Hz, 26.3 times per minute, was detected, as shown in Fig. 3. A peak frequency of 0.375 Hz, 22.5 times per minute, was detected as shown in Fig. 4.

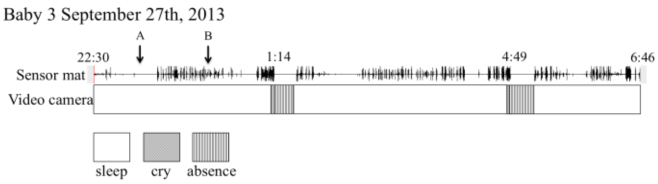


Figure 2: Baby 1's raw data and classified babies' conditions at night using the video data.

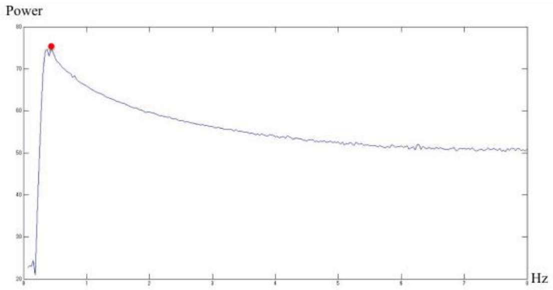


Figure 3: FFT results with body movements.

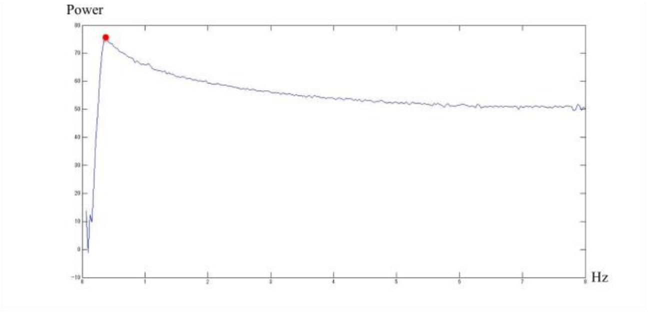


Figure 4: FFT results with no body movements.

Figure 5 shows the peaks through one night for baby 1. The raw data through high-pass filter at 0.3 Hz, and were divided every 32 seconds, and FFT in order to analyze his breathing rate. The horizontal is time and the vertical is frequency in Fig.5. We detected the characteristic frequency, which was near the babies' breathing rates, from 20 to 40 times per minute as shown by the rectangle in Fig. 5-1, Fig. 5-2 and 5-3 show the results of the second and third experiments for baby 1, respectively. Therefore, we detected a breathing rate from 20 from to 40 times per minute.

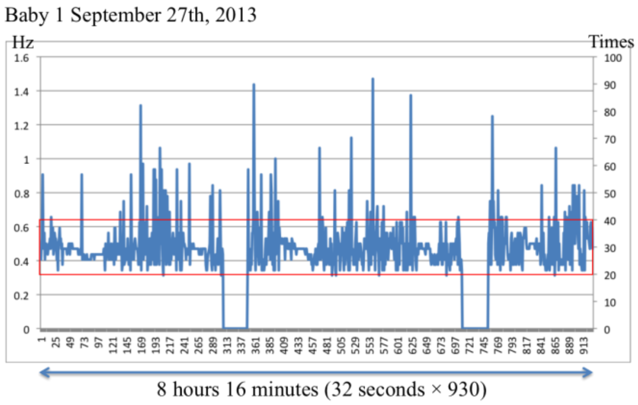


Figure 5-1: First experimental results of peaks through the high-pass filter at 0.3 Hz and FFT throughout night.

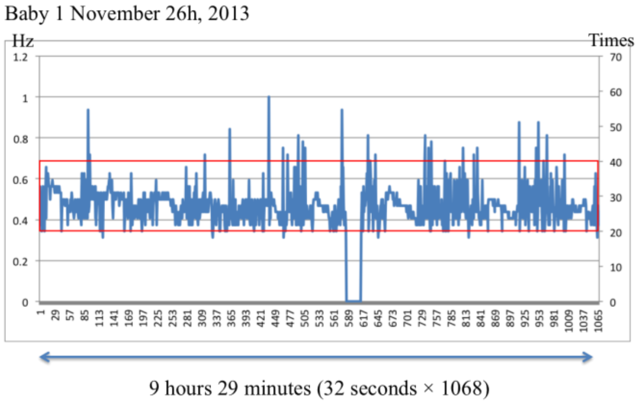


Figure 5-2: Second experimental results of peaks through the high-pass filter at 0.3 Hz and FFT throughout night.

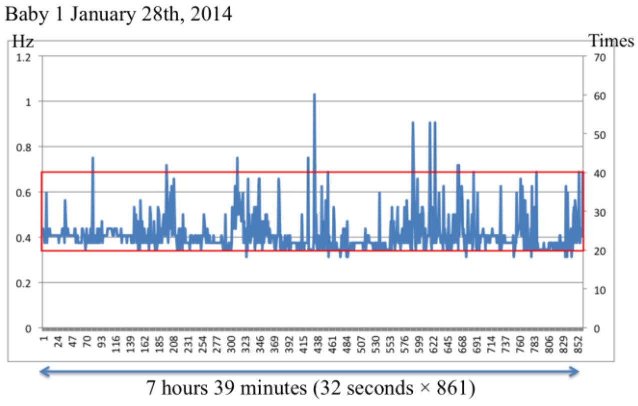


Figure 5-3: Third experimental results of peaks through the high-pass filter at 0.3 Hz and FFT throughout night.

We detected a peak in the frequency band from 1.33 to 2.67 Hz through the high-pass filter (cut-off frequency of 1.3 Hz), the raw data divided every 32 seconds and processed FFT in only the third experiment for baby 1. Figure 6 shows the results. The horizontal is time and the vertical is frequency in Fig.6. All peaks were plotted the baby's heart rates. We also observed another characteristic frequency near the baby’s heart rates. This result had been estimated to heart rate from 80 to 120 times per minute as shown in the rectangle in Fig. 6.

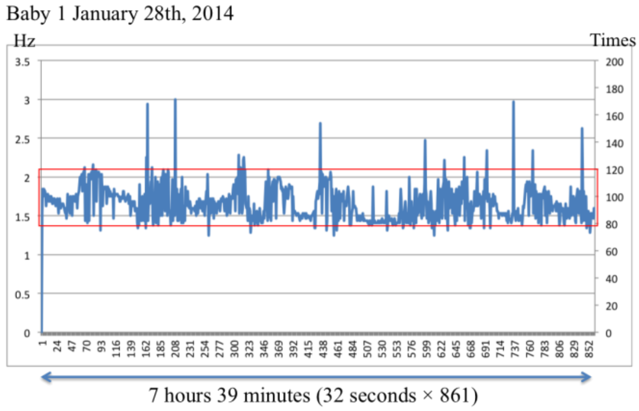


Figure 6: First experiment's result of peaks through the high-pass filter 1.3Hz and FFT throughout night.

We classified active and quiet sleep rhythms using the power got by FFT result. We confirmed body movements in the first experiment for baby 1. Active sleep was confirmed from the body movements, and a quiet sleep was confirmed from no body movements. We classified presence of body movement from FFT result and video data. We weighted by baby's weights. Therefore body movement when power is 70, on the other hand no body movement. As mentioned above, a baby’s sleep rhythm consists of an active and quiet sleeps. One rhythm is from the quiet sleep to the end of active sleep. Figure 7-1, 7-2, and 7-3 show the sleep rhythms from the first, second and third experiments, respectively. The black sections in these figures denote active sleep, and the white sections quiet sleep. We had found a characteristic pattern showing that an awakening would occur after one long and one short sleep cycle, as shown in Fig. 7-1. This occurred two times in one night. We had also found another characteristic pattern, which is shown in Fig. 7-2. There is no evidence of awakening in Fig. 7-3.

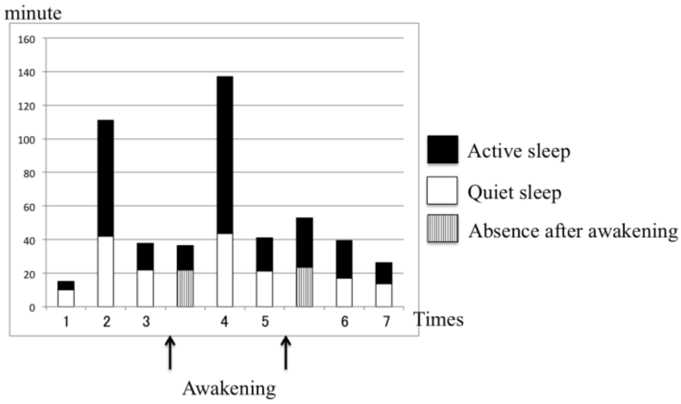


Figure 7-1: Sleep rhythm from first experiment.

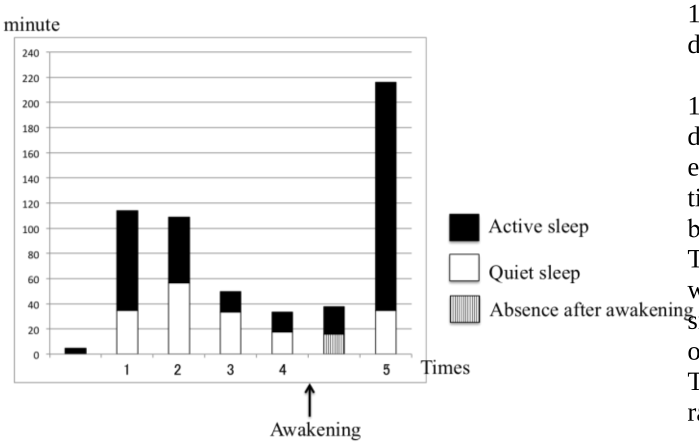


Figure 7-2: Sleep rhythm from second experiment.

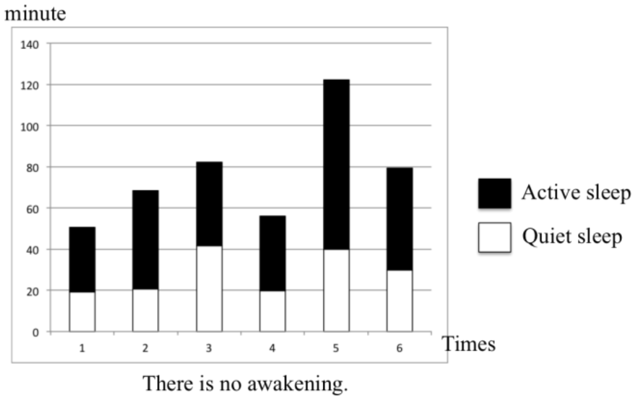


Figure 7-3: Sleep rhythm from third experiment.

5 DISCUSSION

We first discuss the babies' breathing rates analyzed with a 0.3 Hz high-pass filter and FFT. We detected periodical data, from 20 to 40 times per minute, as shown in Fig. 5. As mentioned above, a baby's breathing rate is about 25 to 45 times per minute. Our detected periodical data was started at 20 times per minute. This is because a baby's breathing rate decreases when sleeping, which our experiments confirm. We also detected a stable breathing rate on baby's quiet sleep condition and a changing it on the baby's active sleep condition. We believe the change in the babies' breathing rates was due to their body movements. We detected the babies' breathing rates with or without body movements; however, their breathing rates did not change before and after their awakening. We were also able to detect stable frequencies under all conditions with grown up of the baby 1 on the third experiment. We believe the stable result was due to the babies gaining weight.

Next, we discuss the babies' heart rates, analyzed using the 1.3 Hz high-pass filter and FFT. Since there were no stable data in the experiments 1 and 2, we only discuss the third experiment. We detected periodical data, from 80 to 120 times per minute, as shown in Fig. 6. As stated above, a baby's heart rate is about from 80 to 160 times per minute. The reason we could not detect them in experiments 1 and 2 was that the captured vibration power of heart rate was smaller than any other data, so raw data are hidden in any other data due to the lack of sensitivity of the sensor mat. Therefore, we have not been able to detect the babies' heart rates in the experiments 1 and 2.

We now discuss estimating the babies' sleep rhythms. Figure 7-1, 7-2, and 7-3 show periodical active and quiet sleeps from using the sensor mat and video data, which clearly show the babies' sleep rhythms.

Finally, we discuss the babies' awakening-time. We could not detect any clear change in breathing rate before awakening. We could not confirm any change in the babies' heart rates in experiment 1 because the baby did not wake up and cry throughout the night. If we used a more sensitive sensor mat, we might have captured the change in heart rate before awakening. Regarding the babies sleeping rhythms, awakening occurred after one long and one short sleep cycle, as shown in Fig. 7-1. However, Fig. 7-2 shows that awakening occurred after one short sleep cycle. In both cases, the babies awakened after a few sleep cycles, which were shorter just before awakening. We think we are able to estimate awakening-time by counting sleep cycles. When there was no crying at night, the babies' sleep cycles became longer just before awakening in the morning.

6 CONCLUSION

We detected babies' breathing and heart rates to estimate their awakening-times at night by using the sensor mat. We also detected sleeping rhythms, which were classified based on the existence of the babies' body movements. The rhythms were classified as active or quiet sleep.

We detected the babies' breathing rates with or without body movements; however, they did not change before and after awakening. Therefore, we could not estimate their awakening-times using the changes in their breathing rates. We also detected heart rates in the third experiment, but could not examine the relationship between the baby‘s awakening-times and heart rates, since he did not cry at night. In the future, we will try to collect other data with a few months older babies and use a more sensitive sensor mat.

We found that the baby woke up after a few sleep cycle, which were shorter just before awakening. We believe we are able to estimate awakening-times by counting sleep rhythm. When a baby does not cry or awaken at night, his/her sleep cycle will be longer awakening in the morning. Since we think these data are influenced by personal difference, we would like to carry out additional experiments and collect more data.

ACKNOWLEDGNENTS

This work was supported in a part by JSPS KAKENHI Grant Number 26560418. We also appreciate young mothers for their cooperation to the experiments.

REFERENCES

[1] The Yomiuri Shimbun medical information bureau, Medical care of child is dangerous, Chuokoron-Shinsha, 2002.

[2] Hiroko Hashimoto, Nobuko Miyata, Katsuko Shimoi and Sayoko Yamada, Mother's Anxiety and Helping Need around Child-rearing, Gifu medical care science university bulletin, Vol.2, pp.33-38 (2008).

[3] Jan Born Kirsten Hansen, Lisa Marshall Matthias Molle, and Horst L. Fehm, Timing the end of nocturnal sleep, Nature, vol.397, pp.29-30 (1999).

[4] Etsuko Shimizu, The good sleep guide to kind for a baby and mother, Kanki bublication Co., 2011.

[5] Japanese Soc. of Sleep Res., The hypnology, Asakura bookstore, 2009.

[6] Mieko Shimada, Masaya Segawa, Makoto Higurashi, Rumiko Kimura, Kikuko Oku, Sadao Yamanami, Hiroshi Akamatsu, A Recant Change of Sleep Times and Development of Sleep-wake Rhythm in Infants, The journal of child health Vol. 58, No.5, pp.592-598. 1999

[7] Mayumi Hiramatsu, Izumi Takahashi, Takahide Omori, Taeko Teramoto, Taiko Hirose, Hisami Mikuni, Sleep Rhythm of Infants and parenting Stress, The journal of child health Vol. 65, No.3., pp.415-423. 2006

[8] Avi Sadeh, Christine Acebo, R. Seifer, “Activity-based assessment of sleep-wake patterns during the 1st year of life”, Infant Behavior and Development, Volume 18, Issue 3, 1995, pp329-337

[9] Kazuo Shiraki, Tetsu Takada, Pediatrics for nurse and co-medical, Japanese infant medical affairs publishing company, 2010.

[10] Anders TF, Sadeh A, Appareddy V, Normal sleep in neonates and children, WB, Sanders, Philadelphia, pp.7-18 (1995).

Verification of The Feedback Effect of a Learning System Using Simple Electroencephalographs

Koji Yoshida[*], Fumiyasu Hirai [**] and Isao Miyaji [***]

[*] Department of Information Science, Shonan Institute of Technology, Japan
[**] Graduate School, Shonan Institute of Technology, Japan
[***] Faculty of Informatics, Okayama University of Science, Japan
yoshidak@info.shonan-it.ac.jp, 13T2002@sit.shonan-it.ac.jp, miyaji@mis.ous.ac.jp

Abstract – Simple electroencephalographs (EEG devices), which have recently been used commercially to an increasing extent, are portable and wearable to a degree that they do not restrict a wearer's actions. This convenience of use allows the ordinary use of electroencephalography both inexpensively and widely. This study examines the construction of a simple EEG system that can feed back obtained EEG information for instruction assistance in distance learning. This paper specifically examines two frequency components, gamma waves (low γ) and θ waves, which exhibit reactions during memorization work. We assume the ratio of (θ + α) / (10 × low γ) as an index that presumes memorization capability inspired by the frequent synchronous behavior of (θ + α) / 10 and low γ during memorization work. Accordingly, a feedback system was constructed based on the properties. Then comparative analysis was conducted of correlation in EEG data during a student's memorization work in an EEG measurement experiment. The enhanced percentage of correct answers in memorization work tasks attributable to the support of the feedback system demonstrated the effectiveness of (θ + α) / (10 × low γ) as an index for measuring the degree of memorization. Furthermore, the usefulness of a feedback support system was confirmed.

Keywords: Feedback, Simple electroencephalograph , Memorization , Distance learning

1 INTRODUCTION

Electroencephalography (EEG), which provides biological information, is widely used as a performance index of information processing that takes place in the brain. Frequency response among EEG characteristics is known to be related closely to cognitive processes such as learning, language, and perception. By virtue of continued development of brain science and technology, electroencephalographs (EEG devices), which used to be expensive and bulky, have been miniaturized for portability. Simple EEG devices that are wearable and sufficiently compact to permit a wearer to have unrestricted movement have become commercially available recently. We specifically assess the merits of a simple EEG device and explain the construction of a system that feeds back EEG information for use with distance learning [1].

Distance learning systems are beneficial because the progress and results of learning can be fed back and checked immediately. However, they do have shortcomings: student learning cannot be observed directly. In addition, insufficient information such as the learning state and progress information can offer only limited support. It is therefore indispensable to observe the cognition and mental condition of user students with biological information obtained from EEG devices to enable the support of students in light of their actual conditions. Such a system for observation can be expected to improve distance learning shortcomings and to encourage instruction assistance and student learning. This study specifically examines the fact that θ waves reacted on working memory and that low_γ reacted on memorization work in an experiment conducted last year. Moreover, this study investigates the relation between these two wavelengths. A support system using the properties thereof was constructed. Then an experimental comparative analysis was conducted of correlation of electroencephalogram data of students' memorizing work in actual electroencephalogram measurements.

2 ELECTROENCEPHALOGRAPHY

Electroencephalography (EEG), which can measure an index of human information-processing steps, is widely employed in medicine for integrative functional evaluation of the brain, and for the investigation of brain disorders through epileptology and angiopathy. An electrical signal is created at times of neural ignition or synaptic transmission in the brain. This biological signal can be recorded using an electrode placed on the scalp. It is shown as brain potential change. A trace of those signals is an electroencephalogram [2]. The EEG data are classifiable into five types according to the frequency range. Listed below are the designation, frequency range, and typical mental state of the appearance of each wave.

- δ waves, 1–4 Hz, in sleep
- $\underline{\theta}$ waves, 4–8 Hz, in sleep/attention
- α waves, 8–12 Hz, relaxed/eyes closed
- β waves, 15–20 Hz, concentrated/moving
- γ waves, 30 Hz -, processing memory and vision

Fourier analysis of original collected EEG data yields the power spectrum of each frequency.

However, EEGs exhibit many individual differences. The relation between EEG and the cognition state varies with circumstances, even for the same person. The emergence of α waves does not always imply that a subject is relaxed. Accordingly, it is necessary to perform repeated measurements and to compare data with EEG obtained in various circumstances.

3 SIMPLE ELECTROENCEPHALOGRAPH

An EEG device measures and records electroencephalographic data. Conventional experimental studies of brain physiology use a large-scale apparatus. However, such an EEG device is unsuitable for ordinary use. Medical-grade instruments with many electrodes bother subjects because of their inconvenient requirements for wearing and restriction of movement, which burden subjects with stress. This situation might inhibit their learning, which is the objective of this study. When medical level precision is necessary for the acquisition of EEG data, EEG equipment should be used. However, small portable EEG devices that are readily available are desirable in cases where EEG information is used for applications, assuming a simple EEG input interface and ordinary use.

For these reasons, it is easy and effective to use a simple EEG device rather than the medical type of EEG device for introducing EEG devices to educational use, as in this study. Therefore, this study conducts EEG experiments using MindSet^{TM} produced by NeuroSky, Inc. [3], which is inexpensive and wearable. MindSet transmits digital EEG data to a PC. The potential between a sensor on the forehead and an electrode on the ear is measured, collected EEG data are analyzed with an on-board chip built in an ear pad, and data are transmitted to a PC using Bluetooth, a wireless communication system. Figure 1 shows communications with MindSet and a PC.

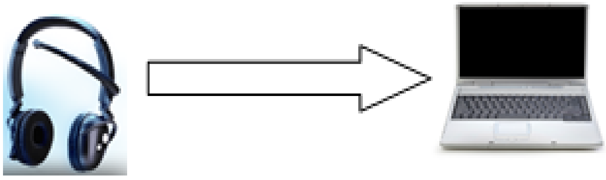


Figure 1: MindSet^{TM} and communication.

Features of MindSet are listed below.

- Measurement point: at the frontal lobe with one sensor (international 10 / 20 system (Fp1)).
- Reference is set on an earlobe.
- Dry sensor type EEG module.
- A chip in the ear pad performs from sensing to analysis.
- Operable with most processors and DPS.
- Data transfer to PC employs Bluetooth communication.
- Sampling at 512 Hz.
- Each frequency component is extracted using fast Fourier transform (FFT) for every second.

The sampling frequency of 512 Hz assures 512 original EEG data obtained every second. Frequency components are extracted by application of FFT to these data, which are then digitized and transmitted to a PC. Other signals are transmitted and received as data aside from these, including poor-sig-lev (noise intensity) and e-sense meter (an original index of NeuroSky) such as the attention level and meditation level. Table 1 presents the range of each frequency component at FFT.

Table 1: Frequency component table of brain wave

Type	Measurable data (Hz)	State of mind
δ waves	0.5–2.75	Deep sleep without dreaming, non-REM sleep, unconsciousness
θ waves	3.5–6.75	Intuition, creativity, remembrance, imagination, illusion, dreams
Low α waves	7.5–9.25	Relaxed but not lazy, peaceful, consciousness
High α waves	10–11.75	Formerly designated as sensorimotor rhythm (SMR), relaxed but concentrating, integrative
Low β waves	13–16.75	Contemplation, recognizing self and environment
High β waves	18–29.75	Alert, wakefulness
Low γ waves	31–39.75	Memorization, high-order cognitive activity
Mid γ waves	41–49.75	Visual information processing

Libraries and applications accompanying the simple NeuroSky EEG device can facilitate users' research and development. The system environment of this study collects EEG data using an application provided by NeuroSky.

4 STATE OF EEG AND LEARNING STATE

Previous research findings in psychology and brain science empirically show that EEG waveforms are useful as an index of a mental condition if observed with a related event. The measurement of the following is regarded as effective to observe human mental conditions: The power spectrum of α and β waves obtained by discrete Fourier transform of obtained EEG, the fraction of α or β waves to the whole EEG, and the ratio of α waves to β waves [4,5]. α waves are generally observed during rest and wakefulness. The α wave amplitude is generally enlarged in a relaxed state, but it shrinks with tension and the appearance of β waves.

Particularly, β waves (13–30 Hz) are regarded as closely related to cognition states. Some reports describe studies that address the relation between intellectual tasks and EEG. Giannitrapani et al. [6] measured the EEG of healthy persons during an intelligence test. They discovered that the low-frequency component of β waves became predominant during reading tests, mathematics tests, and a figure alignment tests, but they are less dominant during other tests, which demonstrates that β waves are effective to some degree as an index for inferring a cognition state.

γ waves (low γ waves) react to memorization work [7], whereas θ waves react to temporary memorization operations such as working memory [8]. This study specifically examined the graphical presentation of low γ and θ waves. Synchronous movement in the EEG appearance patterns of θ + α and 10 × low γ is often observed at the time of memorization work. Figure 2 depicts plots of θ + α and 10 × low γ at the time of music appreciation and memorization work. Accordingly, the ratio of (θ + α) / (10 × low γ) is assumed as an index for presuming the capability for memorization work. Contribution of this index to the feedback effect was verified by experimentation.

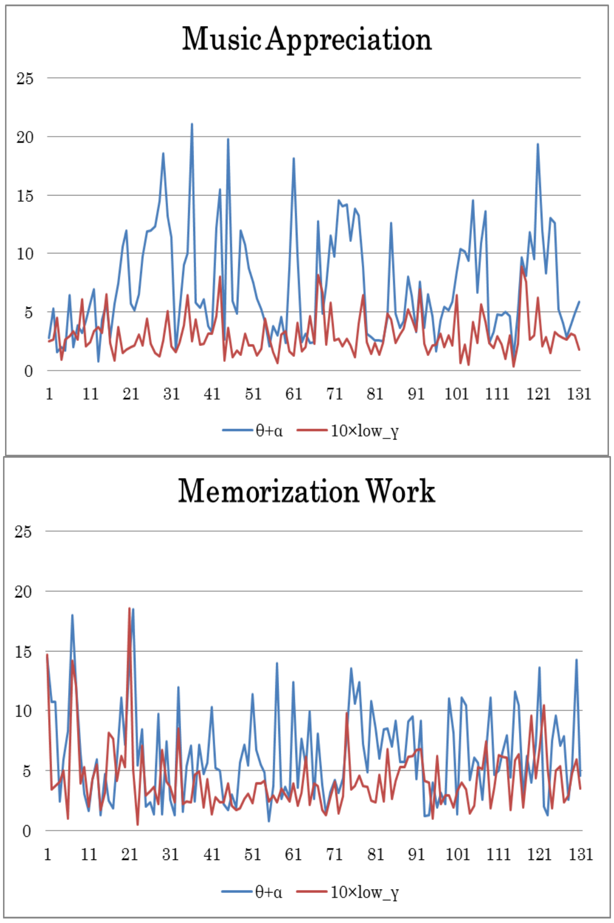


Figure 2: θ +α and 10 × low_γ

4.1 Working Memory

Working memory, located in the prefrontal area, serves a crucial role in cognition. It is the domain for conducting operations while maintaining information temporarily. It affects mental arithmetic, conversation, thinking capability, etc. For example, working memory is always in action when composing a formula in the mind when performing mental arithmetic or when uttering words when talking with people in a conversation.

4.2 N-Back task

An N-Back task is used for experimentation, investigation, and measurement of the "capability of a brain for using temporary memory", designated as working memory. Simple tests of memorization and forgetting are repeated, starting from a state where *n* easily memorable items, such as a number, a character, and a position, are memorized. The capabilities of working memory are measured using test results.

5 EEG DATA FEEDBACK LEARNING SYSTEM

Because the API of MindSet cannot record EEG data, it is necessary to install a Java program that records EEG data. In this study, MindSet is connected through a server program (ThinkGearConnector) provided by NeuroSky. The Java program prepared in this study performs socket communication to the server program and receives data. Data reception is conducted every other second using packet communications.

The present feedback learning system was constructed based on this procedure with graphical presentation, analysis, and evaluation functions of EEG data appended. This system compares the average of the designated index data at a definite interval. The system analyzes and evaluates brain activity by its fluctuation. It conducts feedback to a user after evaluation by displaying an instruction and a graph of index data. In addition, we judged as the noise which we made by setting the threshold of each frequency component about the feedback system. Figure 3 presents a system overview.

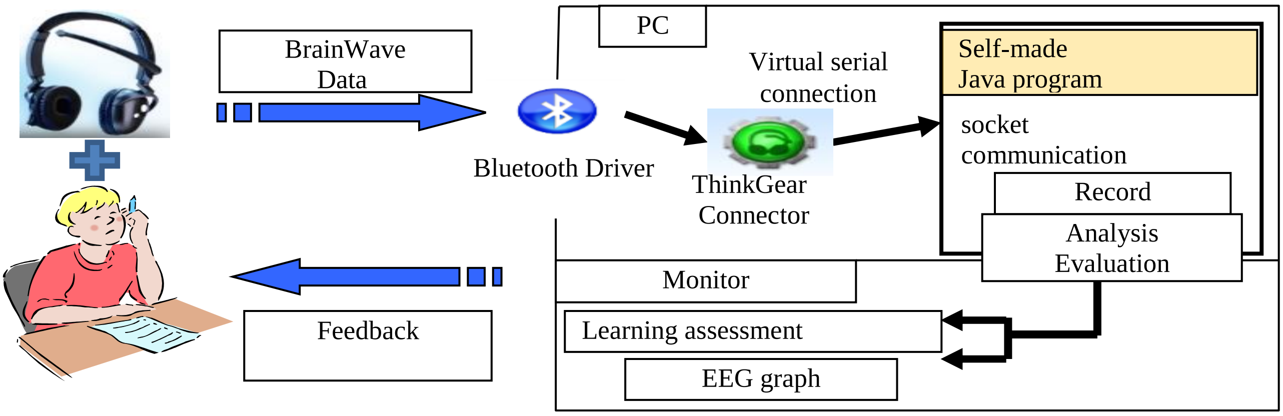


Figure 3: System overview

6 EEG MEASUREMENT EXPERIMENT

This experiment is aimed at analyzing the relation between an EEG and cognition state in a learning state under cognitive work with a simple wearable EEG device. Accordingly, an EEG under cognitive work is measured. Then the relations between the measured EEG, cognition, and frequency response are observed. In addition, the effects of the assumed index and the prototype feedback system are verified.

EEG measurements of several student participants were conducted in our laboratory. Measurements were conducted in a seminar room. The attention of the participants was maintained by providing sufficient intermissions.

Power spectra at respective frequency ranges and sensor sensitivities were recorded during experiments. As a precondition of analysis, measurements were conducted only when sensor sensitivity was at its best. Unusual numerical values occurred, although rarely, among continuous and stable data, even given the best sensor sensitivity. Such cases were excluded as noise.

6.1 Experiment Outline

• Participants: Five men in their 20s (university students studying natural sciences)
• Measuring time: until work completion
• Cognitive theme: visual N-Back task

Conditions for the visual N-Back task are presented below.

• Performed at 3-Back.
• Within a 3 × 3 matrix.
• A participant shall memorize the location of squares displayed successively, and answer whether the location three steps ago and the present one are the same.
• A square is displayed at an interval of 2.5 s.
• A theme finishes as 50 squares are displayed.

6.2 Experimental Procedures

A participant carries out a cognitive process wearing a simple EEG device. The feedback system successively records, analyzes, evaluates, and graphically presents an EEG. Experiments are conducted with/without the system. The results are compared.

Support by the system is evaluated by comparison of the average of index $(\theta + \alpha) / (10 \times \text{low } \gamma)$ taken every 5 s. When the average drops successively, capability for memorization work is judged to have declined, and a restart instruction is displayed on the screen with a monotone. The participant suspends the cognitive task immediately, refreshes the brain, and restarts.

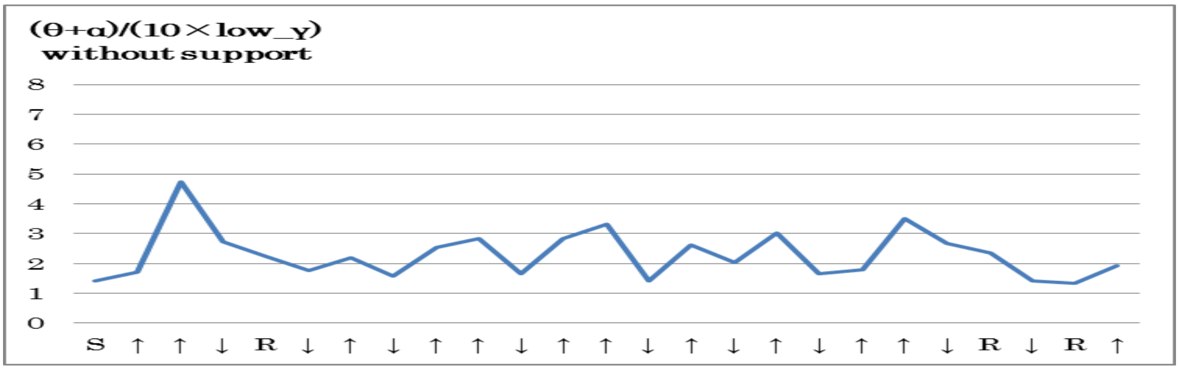
In addition, conditions when the restart is instructed are investigated using a questionnaire administered to each participant after the experiment.

6.3 Experimental results

Table 2-6 presents experimentally obtained results, which include the percentage of correct answers and the average of index $(\theta + \alpha) / (10 \times \text{low } \gamma)$ of each participant with/without support. Figure 4–8 present graphs of the average of $(\theta + \alpha) / (10 \times \text{low } \gamma)$ taken for each participant every 5 s with and without support. Dotted lines on the graph with support show that a restart instruction was given by the feedback system.

Table 2: Percentage of correct answers and $(\theta + \alpha)/(10 \times \text{low } \gamma)$ for Participant A

	Support	**Percentage of correct answers**	$\boldsymbol{(\theta + \alpha)/(10 \times \text{low\_}\gamma)}$
Participant A	without support	30%	2.2875
	with support	53%	3.3263



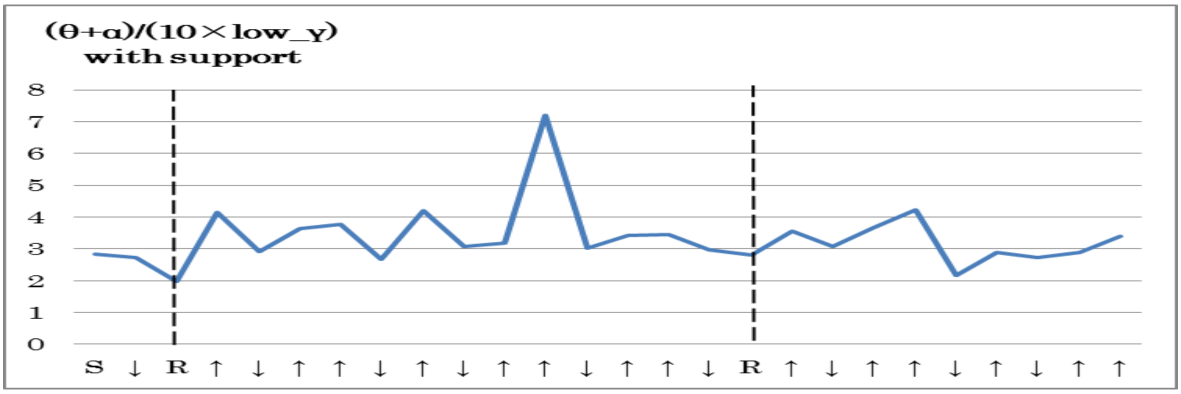


Figure 4: (θ + α)/(10 × low γ) for Participant A

Table 3: Percentage of correct answers and (θ + α)/(10 × low γ) for Participant B

	Support	Percentage of correct answers	*(θ + α)*/(10 × low_*γ*)
Participant B	without support	46%	0.7903
	with support	73%	0.9407

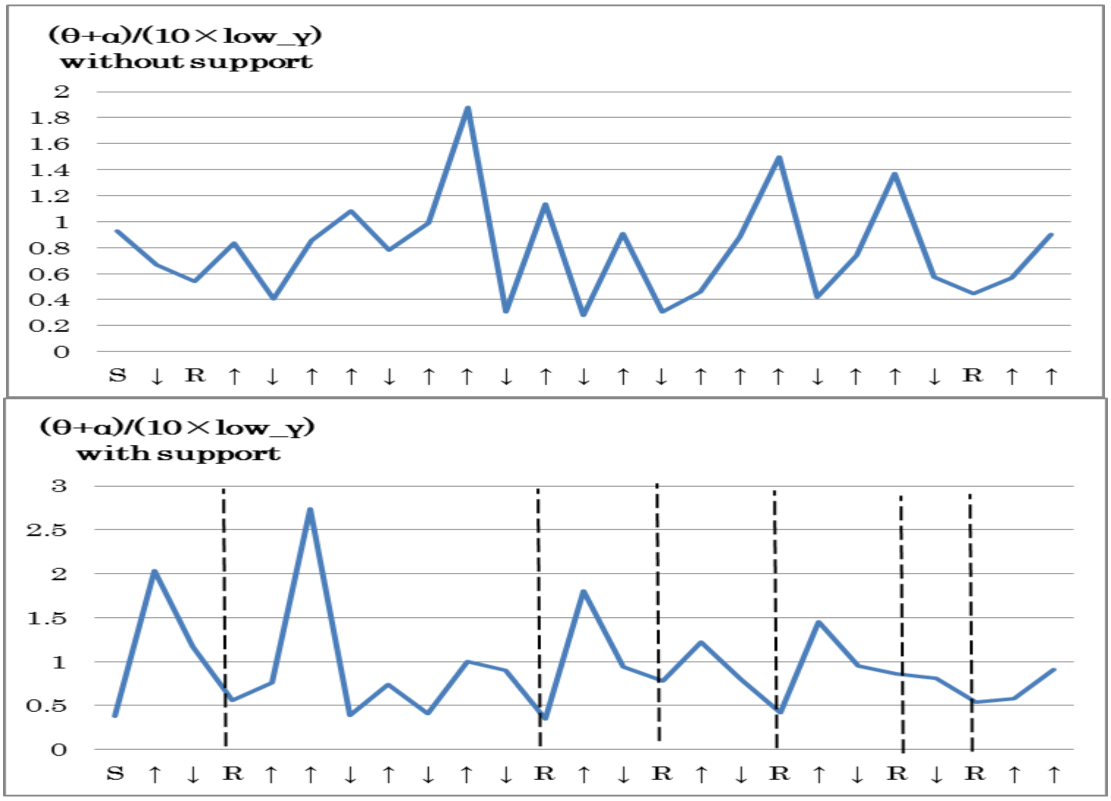


Figure 5: (θ + α)/(10 × low γ) for Participant B

Table 4: Percentage of correct answers and (θ + α)/(10 × low γ) for Participant C

	Support	Percentage of correct answers	*(θ + α)*/(10 × low_*γ*)
Participant C	without support	21%	1.1377
	with support	38%	1.2260

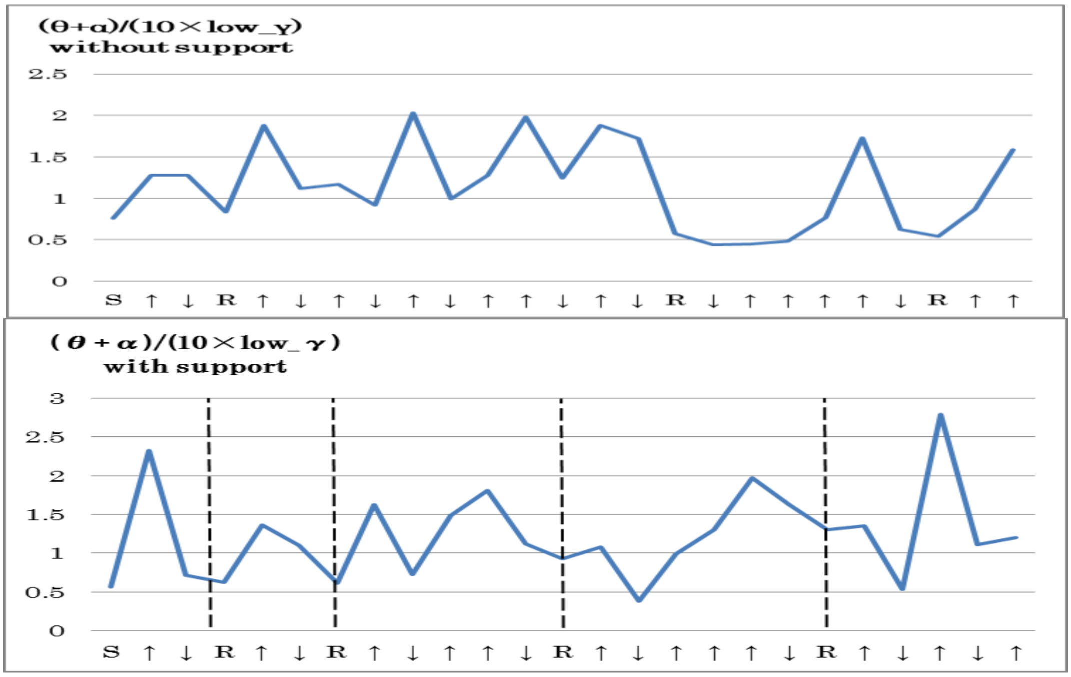


Figure 6: (θ + α)/(10 × low γ) for Participant C

Table 5: Percentage of correct answers and (θ + α)/(10 × low γ) for Participant D

	Support	**Percentage of correct answers**	**$(\theta + \alpha)/(10 \times \text{low\_}\gamma)$**
Participant D	without support	100%	0.4552
	with support	80%	0.3957

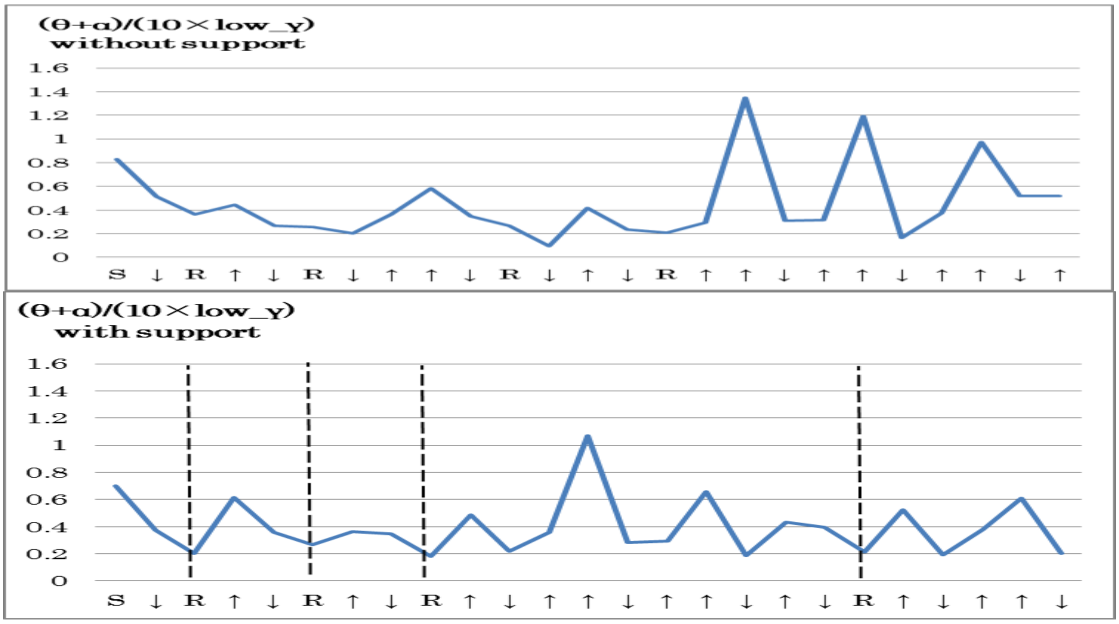


Figure 7: (θ + α)/(10 × low γ) for Participant D

Table 6: Percentage of correct answers and (θ + α)/(10 × low γ) for Participant E

	Support	**Percentage of correct answers**	**$(\theta + \alpha)/(10 \times \text{low\_}\gamma)$**
Participant E	without support	71%	1.3890
	with support	67%	0.9808

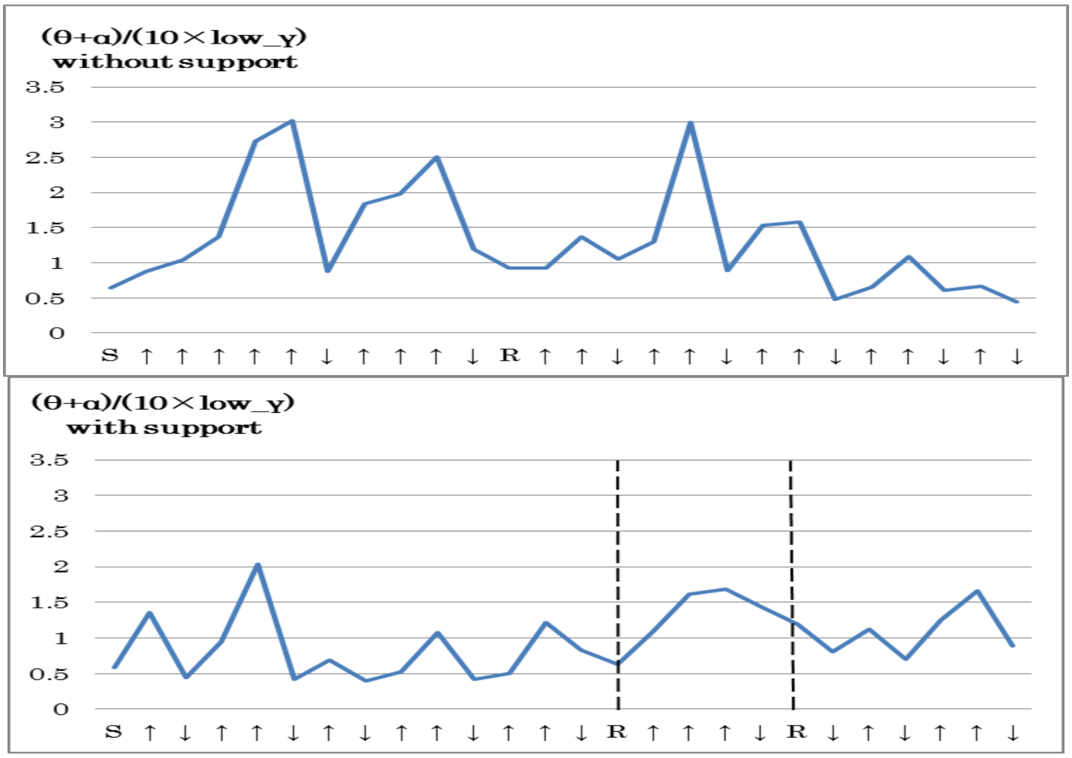


Figure 8: (θ + α)/(10 × low γ) for Participant E

6.4 Evaluation of results

Results obtained with and without a support system revealed a higher percentage of correct answers with support. The overall average of (θ + α) / (10 × low γ) with support was also higher. The graphs suggest that the percentage increased after restart instructions.

A questionnaire survey of participants after the experiment provided answers showing that their memorization work had not caught up immediately before and after a restart instruction. Some participants nevertheless complained that they were directed to restart even though their memorization work had launched satisfactorily. In addition, some participants judged independently that no more memorization work was possible and restarted before an instruction. Some participants felt that the support instruction was too late. Therefore, their results without support were better than those with support.

7 DISCUSSION

The condition of participants as assessed by a questionnaire and the increased values in the graph suggests that a high percentage of correct answers with support are attributable to the fact that a restart direction at a proper timing encourages a participant to address a task more efficiently than ordinarily. A restart tended to raise the average of (θ + α) / (10 × low γ). Presumably, this is true because the memory domain was initialized and capacity was secured.

Malfunction of the restart direction at an early stage is considered because a brain in a comfortable state was suddenly burdened and became unstable. Because a malfunction rarely takes place in a late stage of cognitive tasks, our future subject is to assess responses to the change immediately after a start.

For participants with better results without support, it is presumed that the difficulty of cognitive tasks does not match their level.

Consequently, the discussion presented above verifies the effectiveness of $(\theta + \alpha)/(10 \times \text{low } \gamma)$ as an index for presuming the degree of memory and the usefulness of a feedback support system.

8 CONCLUSION

This study specifically examines two frequencies of low γ and θ waves. The former reacts to memorization work, whereas the latter reacts at a short-term memory domain called working memory. We have assumed (θ + α) / (10 × low γ) as an index of memorization work, produced a prototype feedback system using this index, and conducted experiments thereon. The experimentally obtained results suggest that (θ + α) / (10 × low γ) is effective as an index during memorization work. The present system is expected to be effective as a feedback support system.

Our future subjects include response to a sudden change at an early stage and improvement in the reliability of the index for proper feedback to each participant. We expect to obtain more analytical data with more participants and to improve the analytical precision. Furthermore, we expect to verify differences by sex, and to conduct a cumulative examination of correlation between the appearance pattern of an EEG and learning to consider individual differences.

This study was supported by a grant-in-aid for scientific research No. 24501219.

REFERENCES

[1] D.Szafir,B.Mutlu:”ARTFul:Adaptive Review Technology for Flipped Learning,”CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.1001-1010 (2013)

[2] K. Kitajo and Y. Yamaguchi, “Study on visual perception by EEG phase synchrony analysis,” Vision, vol. 19, no. 4, pp. 193-200 (2007). (in Japanese)

[3] Neuro SKY Sensor http://www.neurosky.com/

[4] Uwano, Ishida, Matsuda, Fukushima, Nakamichi, Ohira, Matsumoto and Okada: “Evaluation of Software Usability Using Electroencephalogram – Comparison of Frequency Component between Different Software Versions,” Human Interface Society, Vol.10,No.2,pp. 233–242 (2008). (in Japanese)

[5] K. Yoshida, F. Hirai, Y. Sakamoto and I. Miyaji: “Evaluation of the change of work using simple electroencephalography,” KES'2013, Proceedings, Knowledge-Based Intelligent Information and Engineering Systems, pp.1817-1826 (2013)

[6] D. Giannitrapani: “The role of 13–Hz activity in mentation,”The EEG of Mental Activities, pp.149–152 (1988).

[7] M. Kawasaki, Y. Yamaguchi:”Neural Substrate for the Effects of Subjective Preference on Visual Working Memory Capacity,” IEICE Transactions on Information and Systems, J94-D(9), 1570-1578, (2011) .(in Japanese)

[8] F. Hirai, K. Yoshida, and I. Miyaji:” Comparison analysis of the thought and the memory at the learning time by the simple electroencephalograph,” Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO) 2013, Symposium Proceedings, pp. 1441-1446 (2013).(in Japanese)

Measurement of Olfaction in Children with Autism by Olfactory Display Using Pulse Ejection

Eri Matsuura[*], Risa Suzuki[*], Shutaro Homma[*], and Ken-ichi Okada[**]

[*] Graduate School of Science of Technology, Keio University, Japan
[**] Faculty of Science and Technology, Keio University, Japan
{matsuura, risa, honma, okada}@mos.ics.keio.ac.jp

Abstract - Autism Spectrum Disorders (ASD) is considered as one of the developmental difficulties caused by dysfunctions of the brain. There are a variety of cases with ASD, but these can be significantly improved by appropriate treatment and education. Therefore, it is important to find the patients with ASD while young. Recently, research on the olfactory senses of patients with ASD is being done. It is reported that there are differences of odor detection and identification abilities between people with ASD and controls. Our aim was to develop the screening examination by olfaction, so we developed the application to assess odor detection and identification abilities in children by olfactory display using pulse ejection. We also investigated the olfactory abilities of the children with ASD using the application. During the experiments, we found some problems with application, so we made improvements. After the improvements, we investigated the olfactory abilities of the typically developing children. As a result, we saw similar tendency as in the related works on olfactory ability in children with ASD by the olfactory display.

Keywords: olfactory display, pulse ejection, autism spectrum disorders, children, interface application

1 INTRODUCTION

Autism Spectrum Disorders (ASD) is considered as one of the developmental difficulties by dysfunctions of the brain [1]. People with ASD have difficulties in three main areas referred to as the “Triad of Impairments” coined by Lorna Wing: impairment of social communication, social imagination, and social relation. Some examples are sudden start of conversations with strangers, difficulty in meeting their eyes, and parroting the question back. There are a variety of other cases with ASD, and it is not a uniform state. However, it is known that such difficulties can be significantly improved by appropriate treatment and education [2]. Therefore, it is important to detect a person with ASD while young. On the other hand, the examinations of ASD have many problems, and it is not easy to carry out an examination. Multiple staff spend time to make preparations for an examination, and it takes a long time to do an examination. For example, one examination takes a maximum of four hours in Japan, but it may take longer in other countries. Furthermore, the staff need to have a discussion to select the appropriate method. Because the examinations require preparation, more than one year of waiting is often needed, and cost too much money to take it lightly. In fact, no matter how much you want, it is not easy to have the examinations of ASD. There are also screening examinations of ASD, but they have some problems as well. For example, the Autism-Spectrum Quotient is unsuitable for detecting ASD while young, because it is developed for adults [3]. There is also the Modified Checklist for Autism in Toddlers (M-CHAT) as a screening examination for children [4], but the Japanese version of M-CHAT [5] has some problems likewise. It is reported that Japanese version of M-CHAT cannot detect 15% of children with ASD.

Recently, research on the olfactory senses of patients with ASD is being done. Because one of the cases with ASD is characteristic behavior towards odor, for example, they have a fetish for certain odors. It is reported that there are differences of odor detection and identification abilities between people with ASD and controls, so checking olfactory abilities may make it possible to screen ASD. In this study, we assess olfactory abilities in children with ASD for developing a screening examination by olfaction.

2 OLFACTION IN PEOPLE WITH ASD

Many children with ASD have sensory difficulty [6]. For example, they don’t want to get on the buses because of its odor, and they keep on sniffing certain odors. For these abnormal responses to sensory stimuli, E.Gal et al. claim that there is a change in cases during puberty, and in responses to sensory stimuli as well [7]. The studies on changes in cases with age have not been done much, but there have been reports that the cases were improved or severer as people with ASD get older. Thus, it is thought that people with ASD don’t show a similar trend, but the cases may vary according to age.

There are a variety of the cases of sensory difficulty, and it is reported that abnormal responses for olfaction and gustation are stronger than for vision and audition. Moreover, it is reported abnormal responses for olfaction, gustation, and touch are stronger for girls than boys, and the difficulties in olfaction of girls with ASD are more serious. Thus, it is believed that many of the patients with ASD respond characteristically to olfactory stimuli, and research on the olfactory senses is being done. Suzuki et al. assessed olfaction of the adults with Asperger Syndrome (AS) in 2003 [8]. 24 participants took part: 12 men with AS and 12 control participants. They assessed odor detection and identification abilities using University of Pennsylvania Smell Identification Test (UPSIT) [9]. Odor identification ability of participants with AS were impaired, but there was no difference in odor detection ability. Bennetto et al. focused on odor identification ability in 2007 [10]. 48 participants aged 10 to 18 took part, 21 patients with ASD and 27 control participants. They assessed odor identification ability using Sniffin' Sticks [11], and reported impaired ability of patients with ASD. There is also research focusing on pleasantness of odors. Hydlicka et al. assessed pleasantness of odors for 70 children aged around 10 years old (35 children with AS or high functioning autism (HFA), 35 control participants) using Sniffin' Sticks [12]. They reported that children with AS or HFA, compared to controls, perceived the odor of cinnamon and pineapple as less pleasant, the same was true of cloves. Dudova et al. assessed odor detection and identification abilities using Sniffin' Sticks, and relation between preference for odors and identification ability [13]. 70 children aged around 10 years old took part, where 35 were children with AS or HFA, and 35 were control participants. The paper reported impaired detection ability of children with AS or HFA, and compared to controls, they were better in correctly identifying the odor of orange and worse at correctly identifying the odor of cloves. As it was found that the odor of orange was favorite odor and the odor of cloves was least favorite odor for children with AS or HFA by Hydlicka et al. , they argued that odor identification ability of children with ASD was related to pleasantness of odors.

The patients with ASD have some characteristics regarding olfaction. However, the characteristics differed among age groups. In other words, it is believed that olfaction may be a physiological indicator of ASD by the age limit. Thus, by assessing olfaction in children, screening patients with ASD may be possible.

3 MESUREMENT OF OLFACTION BY OLFACTORY DISPLAY

It is important to find patients with ASD at while young. But the examinations of ASD have a variety of problems, and it is difficult to do an examination. The screening examinations of ASD also have some problems. Therefore, it is necessary to develop an easy method to screen ASD. Because there are several characteristics of olfaction in people with ASD, we think measurement of olfaction in children will lead to screening examination of ASD. However, the current method of olfaction test in Japan requires great care. There are many tasks and takes too much time. There is also the problem of scent scattering in the air and filling the room. In addition, the current method may be tiring for children because it consists of repetition of simple tasks for a long time, and it has a possibility of measurement failure. Moreover, the odors used in the identification test are not always familiar to children. Thus, it is important to make a simpler method of olfaction test targeted at children. We therefore use olfactory display in the examination to solve these problems, and develop a screening application for children with ASD.

The screening examination is composed of odor detection and identification test. We use the olfactory display using pulse ejection, which means scent is presented for a short time, and the PC for measurements. The person conducting the examination uses the PC to control the olfactory display, and scents are presented simply by operating the PC. The usage of PC reduces the time and effort involved to present a scent. Since part of the operation is automatic, even if the person conducting the examination doesn't have enough knowledge about the olfaction tests, the person can still conduct the examination. There are two screens in our system, one for doctors to check the status and the other for children. Odor detection and identification test are composed of games manipulated by children so as not to bore children. The operations are simple, and children can manipulate intuitively using a touch panel monitor. Since the operations that doctors must do are made minimum, doctors can observe patients. We propose a screening application for children with ASD based on these concepts.

We hypothesized that odor detection ability in children with ASD is impaired and odor identification ability is not different from it in typically developing children, and assessed the olfaction in children.

4 OLFACTORY MESUREMENT METHOD

4.1 Olfactory Display for Medical Purposes

We developed an olfactory display as shown in Figure 1. We call this display "Fragrance Jet for Medical Checkup (FJMC)." FJMC uses the technique used in ink-jet printers in order to produce a jet, which is broken into droplets from the small hole in the ink tank. This device can create pulse ejection for scent presentation: thus the problem of scent scattering in the air can be minimized. The 2D diagram of FJMC is as shown in Figure 2. This display has one large tank and three small tanks, and one fragrance is stored in each tank. There are 255 minute holes in the large tank, and 127 minute holes in small tanks. It is possible to emit scent at the same time through all these holes. We refer to the average ejection quantity per minute holes as "the unit average ejection quantity (UAEQ)", and the number of minute holes that emit scent at the same time as "the number of simultaneous ejection (NSE)." The device can change the ejection time at 667 μs intervals so the measurement can be controlled precisely. This is defined as "the unit time" in this device. UAEQ from one minute hole on the large tank is 7.3 pL, and on the small tanks is 4.7 pL. As this display can emit a fragrance from multiple holes at the same time, the

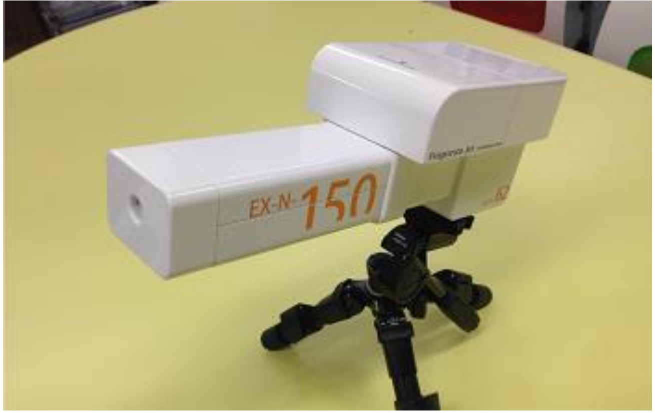


Figure 1: Olfactory Display

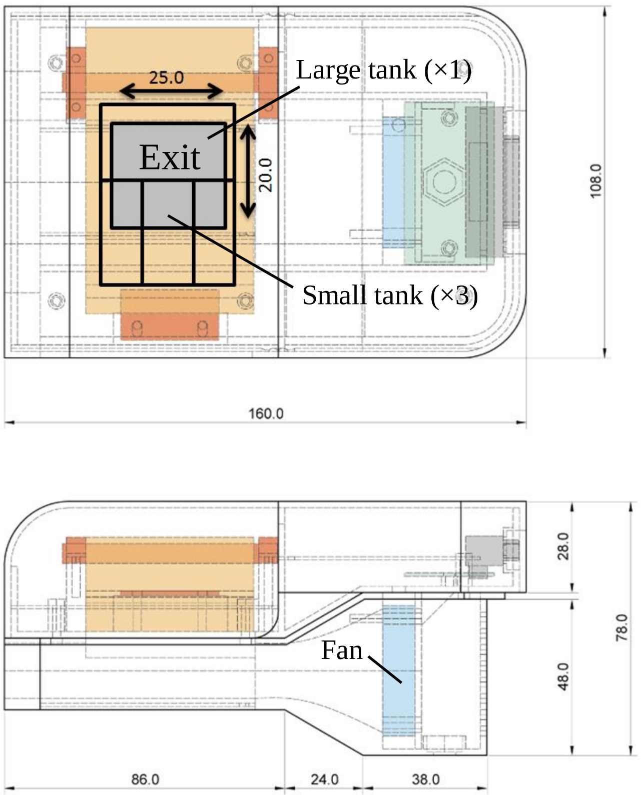


Figure 2: Plane and Side View of the Olfactory Display

range of UAEQ is 0 to 127 for small tanks, and 0 to 255 for the large tank. These values determine the ejection quantity per unit time. Thus, it is possible to control the intensity of scent by the ejection quantity per unit time and the ejection time (ET), and the actual ejection quantity (EQ) can be calculated as follows.

$$EQ(pL) = UAEQ \times NSE \times (ET \div 667\,\mu s) \qquad (1)$$

4.2 Application for Measurement of Olfaction

We developed the screening application for children with ASD. The purpose of developing the application was to reduce time and effort for the olfactory test and not to bore the children. We used FJMC, the PC, and the touch panel monitor in this system. Figure 3 shows a schema of the system. As FJMC uses pulse ejection, scent is emitted for just short periods of time, and presented odor can be changed quickly by operating the PC. The reason why we used touch panel monitor was to make children move their own hands. In addition, we incorporated gaming element in the olfactory test, and we also made the examination simple. The examination is composed of the odor detection and identification test. In the next section, we describe the method of use.

The odor detection test was designed as a treasure hunting game where patients try to find a smelling box out of three boxes. We used the odors of banana and pineapple, and there were four levels of intensity of the odors: 10, 20, 40, and 80. The intensity was expression by NSE. Because the odor of banana is familiar to people, we used it. The odor of pineapple is also familiar to children, we think, but it was reported that children with AS or HFA perceived the odor of cinnamon and pineapple as less pleasant in the research of Hydlicka et al. Thus, we thought children with ASD had some characteristics regarding detection the odor of pineapple, and we chose it. Moreover, the ejection time was 200 msec, and we used only the small tanks. Detection thresholds were determined by using the raising method (the first intensity was 10) and three alternative forced choice procedure. There were three boxes on the screen for patients, one of them with a smell and the others odorless, and the patient select the smelling box from the three boxes. Detection threshold was determined by the intensity which the patient answered correctly twice in a row. If the patient selected the wrong answer, the intensity level was raised one level. In addition, the test was carried out firstly with the odor of banana and then the odor of pineapple. Figure 4 shows the screen for patients. The patient pushed the arrow buttons to move the dog in front of the box the patient wanted to sniff. By pushing the button of the dog, the scent was presented and a sound rang at the same time. When the patient found the smelling box, he or she moved the dog in front of the box and pushed the “this” arrow button.

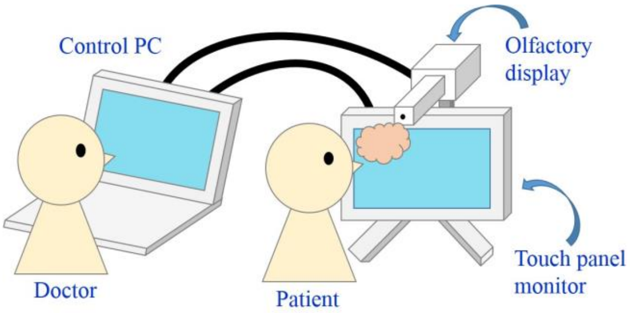
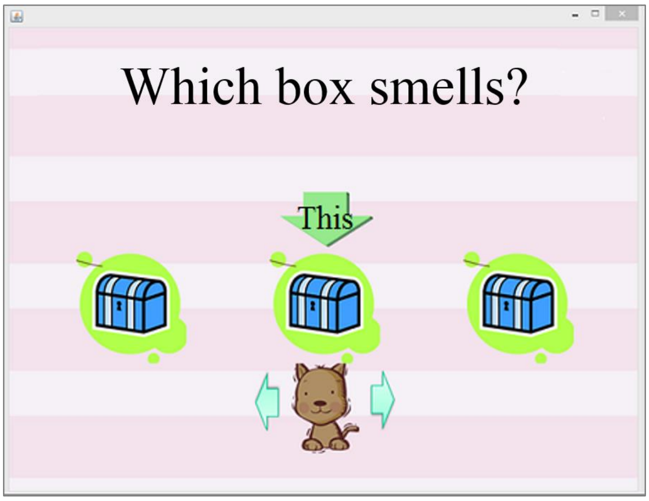


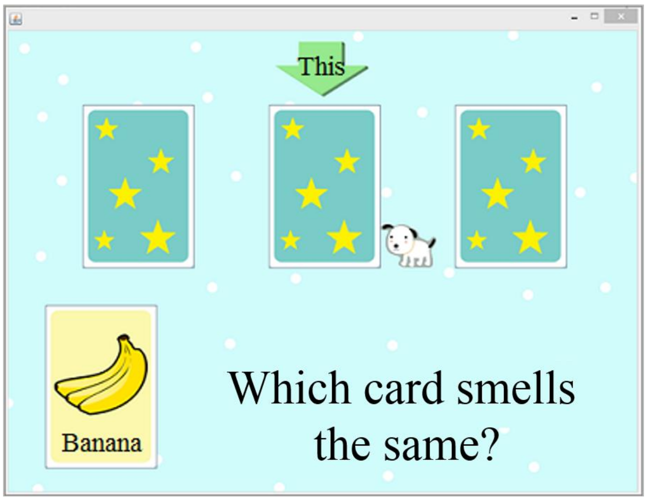
Figure 3: Conceptual Diagram

The odor identification test was designed as a card game where patients try to find the same card by using odor as a clue. In this test, we used odors of banana, rose, and lavender. The intensity was decided by preparatory experiment, and was strong enough to sniff. Two tests were conducted in the odor identification test. Firstly, the patient selected a card which the odor matched the illustration (Trial1). Secondly, the patient sniffed the odor of the target card, and selected a card with the same odor (Trial2). The reason why we carried out two tests was that we wanted to discuss whether the causes of impaired odor identification was the brain (the ability of associating odors and the image of odors) or simply olfaction. Both tests were carried out first with the banana and rose. In this examination, the number of correct answers was evaluated. Before the actual test, the patient checked the odor of banana and rose. The first two of the four tests were Trial1. Figure 5 shows the screen of patients for Trial1. The odor of banana, rose, and lavender was assigned in a random order on the three cards above. The odor was emitted by pressing the card, and the patients sniffed the three cards to find the card with the odor matching the illustration on the bottom left. Since the goal of this test was the identification of odor and illustration, the odor was not emitted when touching the bottom left card. When they found the card with the odor that matched the illustration, they touched the card again and the “this” arrow



Which box smells?

Figure 4: Odor Detection Test



Which card smells the same?

Figure 5: Odor Identification Test

button to answer. After the two tests, Trial2 was conducted. Compared to the screen for Trial1, only the bottom left card was different. Since there was no illustration on the card, first the patient had to touch the bottom left card to sniff the target odor. After sniffing the each odor for the three cards, the patient selected the card with the target odor.

5 EXPERIMENTS AND EVALUATION

5.1 Experiments for Children with ASD

The subjects were 15 patients with ASD (11 boys, 4 girls). Their ages were from 10 to 16 (mean age of 14.3 years, SD 1.74 years). We used the PC, the olfactory display, and the touch panel monitor in the experiment. The interface for doctors was displayed on the PC, and that for patients was displayed on the touch panel monitor. The touch panel monitor was placed in front of the olfactory display so that sniffing the odor can be done while watching the monitor. The experiment was conducted from 10:00 to 17:00 in a quiet room, and the subjects sat on the chair during the experiment. Figure 6 shows the experimental environment.

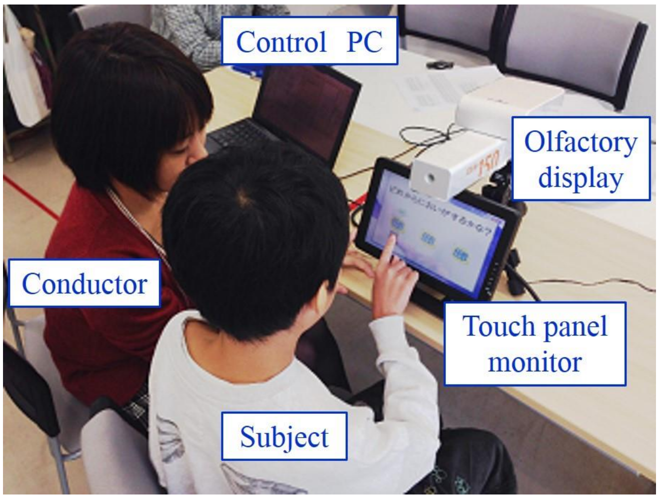


Figure 6: The State of the Experiments

The odor detection threshold for banana was 65.7 ± 55.0 and the threshold for pineapple was 46.9 ± 44.3. However, two subjects had to do the test twice and other two subjects could not finish the test because of operation errors in the odor detection test, so the average was calculated by excluding the two immeasurable subjects. Furthermore, the detection thresholds of the patients who could not detect in 80 was 160. The accuracy rate in odor identification test was about 60 to 70%. We examined the olfactory properties of the patients with ASD. The calculation was carried out by excluding the immeasurable subjects. There were no differences between the detection threshold of banana and pineapple by the t-test ($p = 0.27 > 0.05$). Then, we examined the differences between the results of identification Trial1 (odor and illustration) and Trial2 (odor and odor) by the z-test, and we saw no differences (banana: $p = 1.00 > 0.05$, rose: $p = 0.40 > 0.05$).

During the experiment, the subjects took a variety of characteristic behaviors. There were many children who had shown a large interest in the touch panel monitor and the olfactory display. Some of them looked into the ejection hole of the olfactory display and stuck their nose in. In addition, they asked about the equipment very much but not so much about the examination, and the operation error was sometimes occurred in the odor detection test. Thus, we thought that it is necessary to improve the application to allow more intuitive manipulation. We also thought that it is essential that the screen and operation were made simpler in order to do the experiment easily for the children.

We made mainly three improvements: 1) automation of screen transition, 2) display method of the correct answer and the operation time of the screen of doctors, and 3) adding a practice mode. Firstly, the screen control in this application was not automatic, and the doctor went on to the next screen by clicking on the “next” button. However, during the examination, the doctor was talking with the patients and helping them, so the screen transition was troublesome. Therefore, the screen control was made almost automatic without the cases that the doctor should ask the patient whether the next test can be carried out. Secondly, as only the results were displayed on the screen of doctors, they could not know which answer the patient had tried to select. Moreover, some of the patients took a long time to answer. Therefore, the correct answer and the operation time were displayed on the screen of doctors. Finally, when

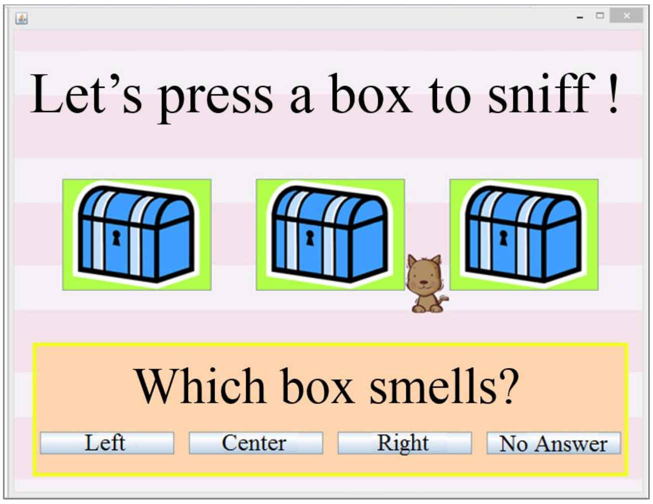


Figure 7: Odor Detection Test after the Improvements

describing the operation method verbally, there was no problem for the adults, but children could not understand without seeing the actual operation. In order to have a good understanding the usage, a practice mode was added in the odor detection and identification tests.

In the odor detection test, the patients often pressed the box or the “this” arrow button, and not the dog when they tried to sniff. So that children could make a more intuitive operation, we made the size of box bigger and the odors were emitted by touching the box. Moreover, time and effort was needed to move the dog. In addition, there was no choice available when the patient couldn’t detect the odor, but had to choose one box. In order to solve these problems, the “this” arrow button was eliminated, and four choices (left, center, right, no answer) were placed at the bottom of the screen. Figure 7 shows the screen for odor detection test after the improvements were made. There was also an improvement of the algorithm. Since the test started from the weakest odor intensity, children who could not detect the odor often appeared to have lost interest in the examination. It was most important to conduct the whole examination, so the examination method was changed to the descending method (the first intensity was 80) in order to enhance the motivation of the children. The timing of changing the level of odor intensity was also improved. If the answer was correct twice in a row, the intensity was lowered one level. Therefore, if the answer was wrong once, the examination was finished and the last intensity which they detected twice became the detection threshold.

During the odor identification test, operation errors were not occurred. However, there was no choice available for when the patient couldn’t find an answer as well. Thus, the “this” arrow button was also eliminated, and four choices were placed at the bottom of the screen. Before the test, the patients could check the odor of banana and rose but not lavender, so we modified the application so that the odor of lavender can be checked in the same section. Figure 8 shows the screen for odor identification test after the improvement.

5.2 Experiments for Controls

The experiments were also conducted for the typically developing children using the application after the improvements. The subjects were nine children (five boys, four girls). Their ages were from 9 to 12 (mean age of 10.8 years, SD 0.916 years). The experiment was conducted from 17:00 to 21:00 in a quiet room. The instruments we used were the same as the previous experiments.

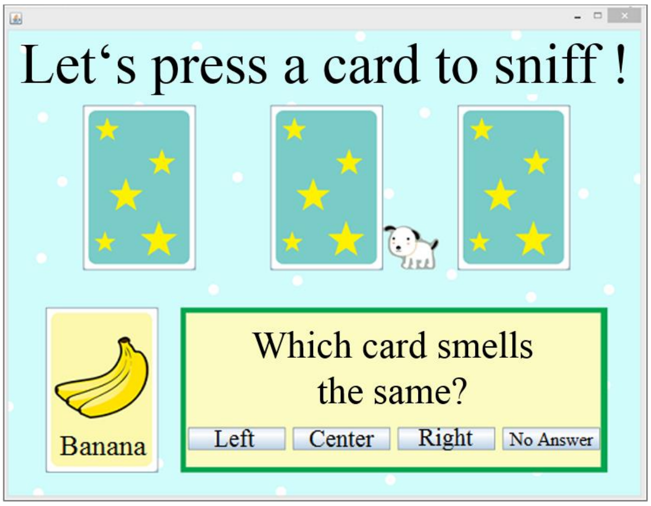


Figure 8: Odor Identification Test after the Improvements

The odor detection threshold for banana was 15.0 ± 10.0 and the threshold for pineapple was 12.2 ± 4.16. In this experiment, operation error was occurred once in the odor detection test, but there were no subjects who could not be assessed. The accuracy rate in the odor identification test was from 70% to 100%. We examined the olfactory properties of controls as well. There were no differences between the detection thresholds of banana and pineapple by the t-test ($p = 0.51 > 0.05$). Then, we examined the difference between the results of identification Trial1 (odor and illustration) and Trial2 (odor and odor) by the z-test, and there were no differences (banana: $p = 0.29 > 0.05$, rose: $p = 1.00 > 0.05$). As with the children with ASD, significant differences could not be found in the odor detection ability between banana and pineapple, and in the odor identification ability between Trial1 and Trial2.

Table 1: The results of detection threshold (M ± SD)

Group	Banana	Pineapple
ASD	65.7 ± 55.0	46.9 ± 44.3
Control	15.0 ± 10.0	12.2 ± 4.16

Table 2: The accuracy rate in odor identification test [%]

Group	Trial1		Trial2	
	Banana	Rose	Banana	Rose
ASD	73.3	73.3	73.3	60.0
Control	100	66.7	88.9	66.7

We compared the result of the children with ASD and the typically developing children (Table1, Table2). The detection thresholds of banana differed significantly between the ASD group and the control group ($p = 0.006 < 0.01$), and pineapple differed significantly as well ($p = 0.02 < 0.05$) by the t-test. Thus, it can be said that the odor detection ability of children with ASD is impaired. There were no differences between the identification ability of odor and illustration of the ASD group and the control group (banana: $p = 0.09 > 0.05$, rose: $p = 0.73 > 0.05$) by the z-test, as well as the identification ability of odor and odor (banana: $p = 0.36 > 0.05$, rose: $p = 0.74 > 0.05$). In other words, it can be said that there are no differences in the odor identification ability between ASD group and control group. In summary, we obtained the results that the odor detection ability of children with ASD was worse compared to typically developing children, but there were no differences in the odor identification ability. This was a result in line with our hypothesis, but the experimental conditions were different. To compare the olfactory abilities accurately, we should reassess children with ASD using the improved application. However, we found on previous study that there was a tendency of the odor detection threshold to be lower using the raising method than the descending method. Therefore, we expect there is the same tendency when we conduct the experiments for the children with ASD again.

6 CONCLUSION

In this study, we developed the screening application for children with ASD. Since it is known that the patients with ASD have characteristics related to olfaction, we introduced the odor detection and identification test for screening ASD. The application targeted children because it is important to find the patients with ASD while young. Therefore, we incorporated gaming element in the olfactory test so as not to bore the children. Moreover, we used the olfactory display in the examination in order to minimize the problems in the old way of assessing olfaction in Japan. We conducted the experiments to assess the olfactory abilities of children with ASD using the application so as to develop the way of screening ASD. During the experiments, we found some problems with the application, and we improved the application. After the improvements, we conducted the experiments for typically developing children. Compared to their results, the children with ASD had less ability of odor detection. However, there were no differences in the odor identification ability between the ASD group and the control group. In short, we were able to see similar results as in the related works on olfactory abilities in children with ASD by the olfactory display using pulse ejection, and the result was in line with our hypothesis. In the future, we will conduct the experiments for children with ASD and typically developing children under the same experimental conditions, and hope to contribute to the construction of the screening tests for ASD by increasing sample size.

REFERENCES

[1] Rich Stoner et al., Patches of Disorganization in the Neocortex of Children with Autism, The New England Journal of Medicine, Vol.370, No.13, pp.1209-1219 (2014).

[2] Masahiko Nishiwaki, Early Intervention for an Improvement of ASD Chidren's Development, Journal of Aichi University of Education Center for Clinical Practice in Education, Vol.3, pp.47-54 (2013) (in Japanese).

[3] Simon Baron-Cohen, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley, The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians, Journal of Autism and Developmental Disorder, Vol.31, No.1, pp.5-17 (2001).

[4] Robins DL1, Fein D, Barton ML, Green JA, The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders, Journal of Autism and Developmental Disorders, 31(2), pp.131-44 (2001).

[5] Naoko Inada, Tomonori Koyama, Eiko Inokuchi, Miho Kuroda, Yoko Kamio, Reliability and validity of the Japanese version of the Modified Checklist for autism in toddlers (M-CHAT), Reseach in Autism Spectrum Disorders, 5 (1), pp.330-336 (2011).

[6] Scott D. Tomchek, Winnie Dunn, Sensory Processing Disorders in Children with Autism: Nature, Assessment, and Intervention, Growing Up with Autism: Working with School-Age Children and Adolescents, The Guilford Publishers, pp.95-123 (2007).

[7] E.Gal, SA.Cermak, A.Ben-Sasson, Sensory Processing Disorders in Children with Autism: Nature, Assessment, and Intervention, Growing Up with Autism: Working with School-Age Children and Adolescents, The Guilford Publishers, pp.95-123 (2007).

[8] Y. Suzuki, H. Critchley, A. Rowe, P. Howlin, D.G.M. Murphy, Impaired Olfactory Identification in Asperger's Syndrome, Journal of Neuropsychiatry and Clinical Neuroscience, Vol.15, pp.105-107 (2003).

[9] Richard L. Doty, Richard E. Frye, Udayan Agrawal, Internal consistency reliability of the fractionated and whole University of Pennsylvania Smell Identification Test, Perception & Psychophysics, 45(4), pp.381-384 (1989).

[10] L. Bennetto, E.S. Kuschner, S.L. Hyman, Olfaction and Taste Processing in Autism, Biological Psychiatry, Vol62. No.9, pp.1015-1021 (2007).

[11] T. Hummel, B. Sekinger, S.R. Wolf, E. Paul, G. Kobal, 'Sniffin' Sticks': Olfactory Performance Assessed by the Combined Testing of Odor Identification, Odor Discrimination and Olfactory Threshold, Chemical Senses, 22 (1), pp.39-52 (1997).

[12] M. Hrdlicka et al., Brief Report: Significant Differences in Perceived Odor Pleasantness Found in Children with ASD, Journal of Autism and Developmental Disorders, Vol.41, No.4, pp.524-527 (2011).

[13] I. Dudova et al. Odor detection thresholds, but not odor identification, is impaired in children with autism, European Child & Adolescent Psychiatry, Vol.20, No.7, pp.333-340 (2011).

Finding An Area with Close Phenotype Values to Predict Proteins That Control Phenotypes

Takatoshi Fujiki[†], Empei Gaku[†], and Takuya Yoshihiro[‡]

[†]Graduate School of Systems Engineering, Wakayama University, Japan
[‡]Faculty of Systems Engineering, Wakayama University, Japan
{s101044, s121010, tac}@sys.wakayama-u.ac.jp

Abstract - Because phenotypes of living creatures are expressed reflecting on interactions among genes and proteins, relations among phenotypes and proteins (or genes) have been regarded as a key issue to be clarified to understand the system of creatures. In this paper, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples G, whose expression levels of A and B are both close to one another, and they always have close values of P. In this paper, we propose a method to extract a pair of proteins that effect on a target phenotype, from a dataset that consists of protein expression profiles and phenotype values.

Keywords: Proteomic Analysis, Two-Dimensional Electrophoresis, Phenotype, Expression Profile, Data Mining

1 Introduction

After the entire human DNA sequence was made public, many post-genome researches started to investigate the systems of living creatures. Proteome analysis is a field of such a post-genome research. The proteome analysis is a research to analyze comprehensively the entire protein sets. Namely, the overview of the functions and the interactions of proteins in *vivo* is clarified by the proteome analysis.

As a method in proteome analyses, there is a technique called 2D electrophoresis [1]. The 2D electrophoresis enables us to measure expression levels of thousands of proteins in a biological tissue simultaneously. From the protein expression profiles obtained by the technique, we can clarify the functions and the interactions of proteins.

In many researches, major goal of researchers is to identify proteins that effect on a certain phenotype. For this purpose, a method for discovering the relationship between one protein and one phenotype is often used. One of the most basic methods is to calculate the correlation coefficient between expression levels of a protein and phenotype values. Relationship between two items can be revealed by a relatively simple statistical method. However, the correlation coefficient evaluates only the liner relationship between two items. In contrast, Qu, et al. proposed a method to discover the nonlinear relationship between a gene and a phenotype using orthogonal polynomials.

On the other hand, there are a few researches that try to discover relationships in which more than one proteins effect on one phenotype. Zhang, et al. studied the interaction among a triplet of genes by comparing the correlation coefficients of genes A and B between two cases where another gene C expresses and does not express [3]. As another method, Inoue, et al. developed an algorithm to predict interactions among three proteins A, B and C based on correlation coefficient [4], and Fujiki, et al. developed an algorithm to predict interactions among three proteins A, B and C based on conditional probability [5]. If we regard C as a phenotype, those methods can be used to investigate the relationship between proteins and phenotypes.

In this paper, we propose a new method to detect interactions from different approaches. Specifically, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples G, whose expression levels of A and B are close to one another, and they always have close values of P. We evaluate the proposed method by applying the proposed method to the real data set.

Note that, to the best of our knowledge, this study is the first study that tries to find a set of two proteins that effect on a phenotype based on a group of samples whose expression levels of A and B are close to one another, and they always have close values of P.

The remainder of this paper is organized as follows. In Section 2, we describe the relation among two proteins and a phenotype assumed in this paper. In Section 3, we describe the proposed algorithm in detail. In Section 4, we evaluate our method by applying it to a real protein expression profile and a data set of phenotype. Finally, in Section 5, we conclude our study.

2 The Relationship between Proteins and Phenotype We Suppose

2.1 Phenotype of Creature

Phenotype is a character that a creature has. It is said that a species of creatures is determined by genes. However, there are cases that a size of body, a color, a pattern, etc. are different from each other among individuals of the same species. Moreover, it is thought that phenotypes vary according to heredity and growth environment. The influence that genes give in phenotype has been studied widely. Particularly, in the field of agriculture and livestock, inbreeding researches are performed to produce individual and breed that are economically valuable.

2.2 Protein Expression Profile

Protein expression levels are the amount of each proteins included in a biological sample. The protein expression levels are typically measured by the 2D electrophoresis method

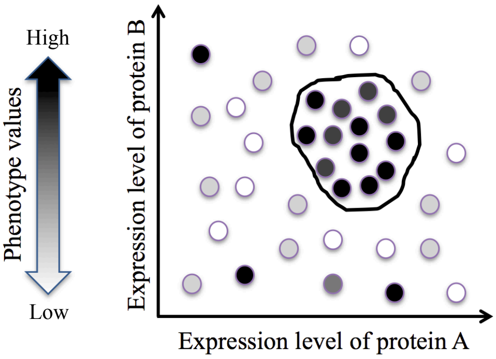


Figure 1: Example of Area That We Want to Demand

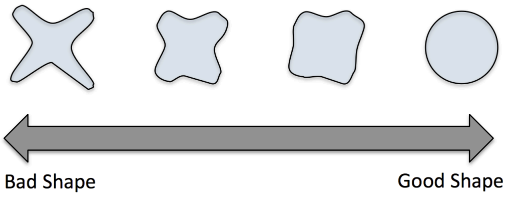


Figure 2: Good or Bad Shape of Area

[1]. This method is used for protein expression measurement widely.

The 2D electrophoresis is a method to separate proteins with 2-dimensions through two steps of electrophoresis. Generally, proteins are separated with isoelectric point in the first dimension, then they further are separated by molecular weight in the second dimension. Typically, the number of proteins included in a profile ranges from several hundred to thousands.

The expression profiles are obtained by 2 steps. First, we obtain a 2D electrophoresis image through the 2D electrophoresis experiment. Second, we measure the areas of the islands revealed by the first step using image processing techniques.

2.3 Relationship of Two Proteins and A Phenotype We Suppose

We suppose two proteins that effect on a phenotype. In this paper, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples whose expression levels of A and B have close values a and b with each other, and they always have close value p of P.

Figure 1 shows an example of this relationship. We consider a 2-dimensional plane that has two axes of expression levels of proteins A and B. Each sample is plotted in this plane, and the deepness of the color of the samples represents phenotype values. Here, if there are no relationship among those two proteins and the phenotype, the distribution of the color of the samples would uniform. That is, the samples up to a deep color from a light color are plotted at a uniform density. In contrast, if some relationships exist, it is thought that the distribution would ununiform. In this paper, as shown in Figure 1, we extract the area in which all the samples have high phenotype values. By extracting such areas, it is possible to estimate the combination of two proteins and those expression levels that effect on a phenotype.

Table 1: Data Format of Protein Expression Profiles

Sample ID	Protein ID			
	A	B	C	···
1	0.000582	0.000107	0.000541	···
2	0.000563	0.000111	0.000458	···
3	0.000495	0.000126	0.000333	···
···	···	···	···	···

Table 2: Data Format of Phenotype Data Set

Sample ID	Phenotype			
	Beef Marbling Standard	Carcass Weight	Rib-eye Area	···
1	4	422.7	44	···
2	9	470.7	53	···
3	7	433.5	50	···
···	···	···	···	···

3 Extraction Method of Area with Close Phenotype Values

3.1 Format of Input Data

We use two sets of input data in the proposed method. One is a protein expression profile and the other is a set of phenotype data. We assume that the protein expression profile is obtained from the 2D electrophoresis experiment. The expression profile consists of the expression levels of each protein contained in each biological sample. We let $i(1 \leq i \leq I)$ be a sample, and let $j(1 \leq j \leq J)$ be a protein. Then, the expression level e_{ij} of a protein j included in a sample i is a real value. We show an example of the expression profile in Table 1.

We assume that a phenotype data set is represented by a table. Then the phenotype data set consists of the real values that represent the degree of phenotype (hereafter, we call them the $phenotypevalues$). We let $p(1 \leq p \leq P)$ be a phenotype, and the phenotype value p_i of a phenotype p included in a sample i is a real value. We show an example of the phenotype data set in Table 2. This example shows a case of brand cattle, in which we have BMS (Beef Marbling Standard), carcass weight, rib-eye area, etc. as phenotypes.

3.2 Areas That We Wish to Extract

In this paper, we extract a pair of proteins A and B that effect on a target phenotype p, by finding an area in which there is a set of samples whose expression levels of A and B are close to one another, and they always have close values of p on the 2-dimensional plane. In this section, we describe the criteria that the area should satisfy.

We consider two criteria with which we evaluate areas. Two criteria are on the phenotype values, and on the shape

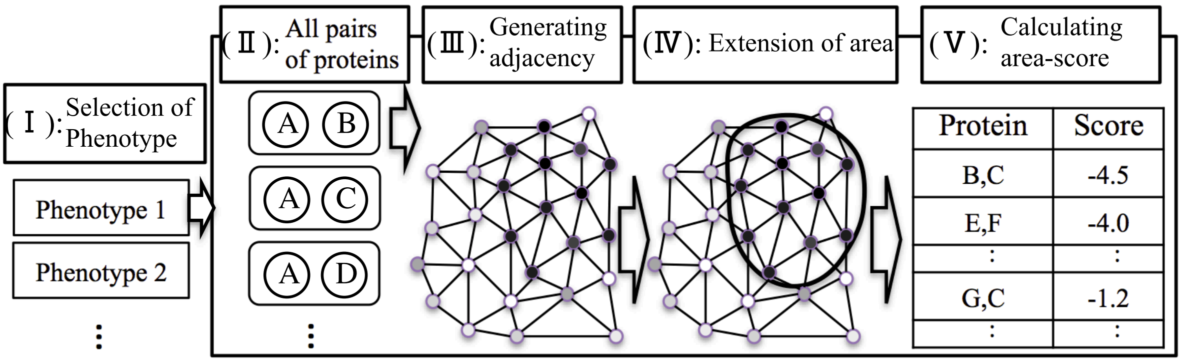

(Ⅰ): Selection of Phenotype
Phenotype 1
Phenotype 2
(Ⅱ): All pairs of proteins
A B
A C
A D
(Ⅲ): Generating adjacency
(Ⅳ): Extension of area
(Ⅴ): Calculating area-score
Protein | Score
B,C | -4.5
E,F | -4.0
G,C | -1.2

Figure 3: Overview of Our Proposed Method

of the area, respectively. First, we describe the criterion on the phenotype values included in the area. It is required that the variance of the phenotype values included in the area is smaller than the variance of phenotype values of all samples. Namely, it means that the samples that take a narrow range of phenotype values are included in it.

Next, we describe the criterion on the shape of the area. The criterion on shape is that the shape does not have big unevenness on a boundary line (i.e. the shape is “not warped”). Namely, in this paper, we regard the circle as the best shape, whereas we regard the “warped” shape as bad shape. (See Figure 2). Without this criterion, if an area is allowed to be any shape, we can make the area of any shape by choosing the samples with close phenotype value freely. By limiting shape of areas, we can evaluate the area properly according to the sample distribution.

3.3 Overview of Proposed Method

We designed an algorithm to find the area that keeps the criterion we described in Section 3.2. We describe the overview of the proposed method as follows. (See Figure 3 in parallel.)

(a) We select one phenotype to analyze (Figure 3(I)).

(b) We compose all the possible pairs of proteins for the phenotype selected in step (a) (Figure 3(Ⅱ)).

(c) We generate an adjacency graph from the samples on a 2-dimensional plane whose two axes are the expression levels of proteins A and B (Figure 3(Ⅲ)). We generate the adjacency graph as the Delaunay graph. We will give a short explanation of the Delaunay graph in the following Sections 3.4.

(d) We repeat extending the area using the graph that is generated in step (c) (Figure 3(Ⅳ)). We start with the area that includes one arbitrarily sample (we call this sample *starting sample*). Then, we repeat extending the area with the most suitable samples until it comes to contain all samples. We perform this process from every starting sample. We describe this extending process in the following Sections 3.5.

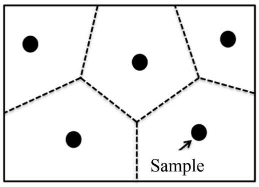


Figure 4: Volonoi Diagram

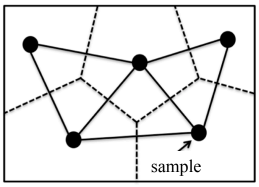


Figure 5: Delaunay Diagram

(e) We calculate the variance of the phenotype values for all the areas throughout the extending process i.e., we calculate the variance every time after extending the area with one sample. Then, for each individual area, we calculate its z-value (we call it the *area-score*) that indicates the statistical probability that the value of the variance occurs (Figure 3(V)). We extract the areas whose area-score is greater than the threshold. We describe about the calculating area-score in the following Sections 3.6.

3.4 Step(c): Generating the Adjacency Graph from Samples

In this Section, we explain the algorithm to generate the adjacency graph from the samples on the 2-dimensional plane.

First, we generate a Voronoi diagram [6] on the 2-dimensional plane. A Voronoi diagram (Figure 4) is a diagram obtained by dividing space into a number of areas. The boundary lines (dotted lines) between samples are composed of perpendicular bisectors between two samples. The plane is divided into areas (called Voronoi area) corresponding to each sample by the boundary line.

By connecting two samples corresponding to two adjacent Voronoi areas, the Delaunay diagram (Figure 5) that represents the adjacency among samples is generated. Then, we let $N(i)$ be the sample set adjacent to sample i.

3.5 Step(d): Extension of Areas

3.5.1 Overview of Extension Algorithm

We describe an algorithm to extend the area we wish to extract. We show the overview of the process as follows.

(1) An initial area consisting of one sample is determined by selecting a starting sample arbitrarily.

(2) We select a set of *extension candidate samples* from the samples that are adjacent to the current area so as not to make the shape of the extended area donut-shape.

(3) We select an *extension sample* from the set of *extension candidate samples* and extend the area by adding this extension sample.

(4) If the area does not include all samples, we return to (2).

As for (1), There are various choices in selection of the starting sample. For example, when we want to extract the area in which high phenotype values are included, it is desirable to select a sample with high phenotype value as a starting sample. Steps (2) and (3) are described in the following sections 3.5.2 and 3.5.3, respectively.

3.5.2 Method to Select Extension Candidate Samples

In this Section, we explain the method to select the set of *extension candidate samples* mentioned in Section 3.5.1 (2). We let C be a set of extension candidate samples, and let D be the current area. C is a set of samples that satisfy conditions (i), (ii) and (iii) among the samples that is adjacent to D. We prevent extensions from becoming donut-shape by setting these conditions as follows.

(i) Candidate sample must be adjacent to more than one samples that are included in D.

(ii) Samples on the boundary of D adjacent to the candidate sample must be continuous on the boundary.

(iii) There is no other samples between the candidate sample and the sample on the boundary of D.

We explain that the area does not become donut-shape using an example. In the area shown in Figure 6, the samples that are surrounded by a black square are the samples that satisfies condition (i). Among them, the X-mark sample does not satisfy the condition (ii) because the three samples on the boundary line of D adjacent to this X-mark sample are not continuous on the boundary line. Moreover, this X-mark sample does not satisfy the condition (iii), neither, because other samples exist between the X-mark sample and D. If we add this X-mark sample to D, the area that is extended becomes the donut-shape.

3.5.3 Method to Select Extension Sample

We explain the algorithm to select a *extension sample* mentioned in Section 3.5.1. An extension sample is the sample that is the most desirable to be added to D from a set of extension candidate sample. As described in Section 3.2, the sample that is the most desirable is the sample that satisfies two criteria on the phenotype values and on the shape of the area, respectively.

We describe the steps to select the extension sample from the extension candidate sample set C. First, we calculate the *shape-cost* $T(x)$ for every *extension candidate sample* $x \in C$. $T(x)$ evaluates the shape of the area that is created by adding the sample x to the current area D. Now, we let D_x be the area that is created by extending D with x. The less $T(x)$ is, the more the shape of D_x is warped.

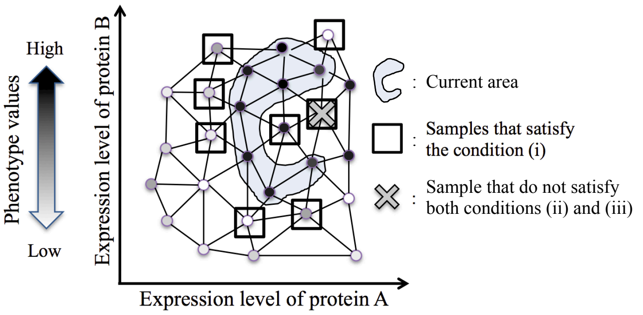

High
Phenotype values
Low
Expression level of protein B
Expression level of protein A
: Current area
: Samples that satisfy the condition (i)
: Sample that do not satisfy both conditions (ii) and (iii)

Figure 6: Conditions of Extension Candidate Sample

Next, we calculate the *phenotype-cost* $Z(x)$ to evaluate the phenotype value of the samples included in D_x. If $Z(x)$ is small, the phenotype value of x take a value close to the samples in D, and x would not increase the variance of the phenotype values of samples included in D. That is, we select the extension sample x such that $Z(x)$ is the smallest in C, and satisfy $T(x) \geq T_{thresh}$, where T_{thresh} is the threshold to the $T(x)$, and we consider the shape of the area that satisfies $T(x) \geq T_{thresh}$ to be "not warped." If there is no x to satisfy $T(x) \geq T_{thresh}$ we regard x such that $Z(x)$ is the smallest in C as the extension sample.

We explain the method to calculate the shape-cost $T(x)$. We calculate $T(x)$ based on the ratio between the boundary length of D_x and the area of D_x. In general, if the area of land is the same, the boundary length is shorter when the shape is close to circle. We calculate $T(x)$ using this property. First, we let L_x and S_x be the length of the boundary line and the area of D_x, respectively. Here, the radius r_x of the circle whose circumference is just L_x is written as $r_x = \dfrac{L_x}{2\pi}$. Similarly, the radius r'_x of the circle whose area is just S_x is written as $r'_x = \sqrt{\dfrac{S_x}{\pi}}$. Finally, We define $T(x)$ as the ratio of r_x and r'_x as follows:

$$T(x) = \frac{r'_x}{r_x} = \frac{2\sqrt{\pi S_x}}{L_x}.$$

Note that, $T(x)$ takes a value between 0 and 1, and the more $T(x)$ increases, the more the shape close to a circle.

The phenotype-cost $Z(x)$ evaluates the variance of phenotype values included in D_x, which is created by extending D with x. We calculate $Z(x)$ as the z-value, which is known as a kind of statistic. $Z(x)$ is defined as $Z(x) = \left|\dfrac{p_x - \mu_x}{\sigma_x}\right|$, where p_x is the phenotype value of x, and μ_x and σ_x are the average and the standard deviation of the phenotype values of samples in D_x. Note that, $Z(x)$ represents the amount of difference between the phenotype value p_x and the average μ_x of the distribution of phenotype values in D_x, which is measured as the number of the unit value σ_x. If the absolute value of $Z(x)$ is small, it means that p_x is close to the phenotype values of samples in D_x. Namely, if we add such p_x to D, the variance of D would not be increased.

3.6 Step(e) Calculating Area Score

In this Section, we explain the retrieval of the areas we wish to extract.

In this paper, we wish to extract the area with small variance of the phenotype values of the samples in the area. However, in general, if the number of samples in the current area is low, the variance is small. Therefore, in this paper, in order to retrieve the good area under variation of the number of samples, we calculate the variance of every area throughout our process to extend areas, and aggregate the variance of all the areas. Then, we judge whether the area is the one we wish to extract, by calculating the *area-score* based on the result of aggregation. Now, we let $(i_1, i_2)\ \ (1 \leq i_1 < i_2 \leq J)$ be the pair of proteins, and let $n\ \ (1 \leq n \leq I)$ be the number of samples in the area. We suppose the extending process of the area with a starting sample m on the plane whose axes are two proteins (i_1, i_2). Here, we define the area where number of samples in the area is n as $D^{(n)}_{m,(i_1,i_2)}$. Note that, $D^{(n)}_{m,(i_1,i_2)}$ is determined uniquely by n, m and (i_1, i_2). Now, we let p_i be the phenotype value of sample i $(i \in D^{(n)}_{m,(i_1,i_2)})$, and $E[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ be the average and the variance of the phenotype values of samples in $D^{(n)}_{m,(i_1,i_2)}$.

We explain the method to calculate the *area-score*. First, we calculate $E[D^{(n)}_{m,(i_1,i_2)}]$ correspond to combination of n, m and (i_1, i_2) as follows:

$$E[D^{(n)}_{m,(i_1,i_2)}] = \frac{1}{n} \sum_{i \in D^{(n)}_{m,(i_1,i_2)}} p_i.$$

Similarly, we calculate $V[D^{(n)}_{m,(i_1,i_2)}]$ as follows:

$$V[D^{(n)}_{m,(i_1,i_2)}] = \frac{1}{n-1} \sum_{i \in D^{(n)}_{m,(i_1,i_2)}} (p_i - E[D^{(n)}_{m,(i_1,i_2)}])^2.$$

Next, we calculate the average μ_n and the standard deviation σ_n of $V[D^{(n)}_{m,(i_1,i_2)}]$ with all areas whose number of samples in the area is n as follows:

$$\begin{aligned} \mu_n &= \frac{1}{|M| \times J(J-1)/2} \sum_{m \in M} \sum_{1 \leq u < v \leq J} V[D^{(n)}_{m,(i_1,i_2)}], \\ \sigma_n &= \sqrt{\frac{\sum_{m \in M} \sum_{1 \leq u < v \leq J} (V[D^{(n)}_{m,(i_1,i_2)}] - \mu_n)^2}{|M| \times \frac{J(J-1)}{2} - 1}}. \end{aligned}$$

Finally, we calculate the z-value for the variance $V[D^{(n)}_{m,(i_1,i_2)}]$ of each area using μ_n and σ_n as the area-score $R^{(n)}_{m,(i_1,i_2)}$. The area-score $R^{(n)}_{m,(i_1,i_2)}$ is defined as follows:

$$R^{(n)}_{m,(i_1,i_2)} = \frac{V[D^{(n)}_{m,(i_1,i_2)}] - \mu_n}{\sigma_n}.$$

If $R^{(n)}_{m,(i_1,i_2)}$ is small, it means that the area rarely appears statistically. Therefore, we expect the area $D^{(n)}_{m,(i_1,i_2)}$ whose are-score $R^{(n)}_{m,(i_1,i_2)}$ is small enough for the output of the proposed method. For such areas D, we suppose there would be an interaction among two proteins i_1, i_2, and the phenotype.

4 Evaluation and Discussion

4.1 Evaluation Method

We evaluate the proposed method by applying it to real protein expression profiles and a phenotype data set obtained by the author's collaborative work in Wakayama [7]. The protein expression profiles that we use in our evaluation are obtained by a 2D electrophoresis-based experiment [8].

A measurement error occurs in the measurement of the protein expression levels. Therefore, we performed 2D electrophoresis twice for each sample to confirm the accuracy of each electrophoresis experiment. From the result of the duplicated measurement, we removed the values considered to be low reliability from expression profiles. Specifically, we measured two expression values for each pair of a protein and a sample. If the larger expression level is larger than 1.3 times the value of the smaller expression level, we consider the expression level for the protein and the sample to be a null value as they are not reliable. Otherwise, the average of the two expression levels is used for each sample-protein pair. As a result, the expression profiles used for our evaluation consist of 90 samples and 47 proteins. In addition, the expression profiles are standardized in advance so that the average and the standard deviation of the expression levels with each sample are 0 and 1, respectively.

We performed an evaluation using "Carcass weight" as an important phenotype among many items included in the phenotype data set of beef cattle. As a pre-processing, we also standardized the phenotype data.

In order to evaluate the performance of the proposed method, we implemented a *simple method* to extend areas to be compared with the proposed method. The simple method is the method that replaces the extension algorithm explained in Sections 3.5 and 3.4. The simple method adds a sample that is close in the Euclidean distance to the start sample m to the current area $D^{(n)}_{m,(i_1,i_2)}$. Consequently, the shape of the area $D^{(n)}_{m,(i_1,i_2)}$ that is obtained by the simple method is nearly a circle centered on the start sample m. Thus, the simple algorithm is equivalent to the algorithm that retrieves the best circular areas in the plane.

We evaluate the performance of the proposed method by comparing it with the simple method by calculating the variance $V[D^{(n)}_{m,(i_1,i_2)}]$ and the average $E[D^{(n)}_{m,(i_1,i_2)}]$.

Here, we describe the parameters in the evaluation experiment. We set the threshold of the shape-cost $T(x)$ as $T_{thresh} = 0.7$, and we set the number of samples in the area between 20 and 40 in order to ensure the reliability of the variance of the phenotype values in $D^{(n)}_{m,(i_1,i_2)}$. In addition, as the starting sample m, we use the sample whose phenotype value is within the bottom 10% among all samples. As actual require-

Table 3: Ranking of Areas with Proposed Method

Ranking	Protein A	Protein B	Number of samples	Area score	Variance in area	Average in area	Shape score
1	3899	4491	39	-2.7545	0.2510	-0.5539	0.7003
2	5639	5735	31	-2.5615	0.1862	-0.6034	0.7012
3	3648	4491	38	-2.4033	0.3012	-0.4405	0.7057
4	828	5733	36	-2.3852	0.2832	-0.3981	0.7002
5	3648	5727	40	-2.3596	0.3283	-0.3444	0.7010
6	3899	3598	30	-2.3408	0.2153	-0.5549	0.7175
7	4491	5727	29	-2.3014	0.2090	-0.7281	0.7058
8	5636	5654	38	-2.3002	0.3193	-0.4944	0.7001
9	3648	5726	38	-2.2910	0.3209	-0.4495	0.7276
10	4491	5730	40	-2.2879	0.3406	-0.3662	0.7060

Table 4: Ranking of Areas with Simple Method

Ranking	Protein A	Protein B	Number of samples	Area score	Variance in area	Average in area
1	3648	4491	31	-2.9546	0.3939	-0.3227
2	4491	5657	40	-2.8999	0.5544	-0.2364
3	4491	5688	40	-2.8186	0.5688	-0.2780
4	4491	5686	39	-2.8077	0.5571	-0.2671
5	4491	5721	26	-2.8066	0.3203	-0.4939
6	828	5660	38	-2.8003	0.5436	-0.1725
7	4491	5724	39	-2.6507	0.5856	-0.3966
8	4491	5734	36	-2.6493	0.5437	-0.2875
9	828	4991	40	-2.6394	0.6005	-0.2203
10	5637	5644	25	-2.6194	0.3477	-0.4936

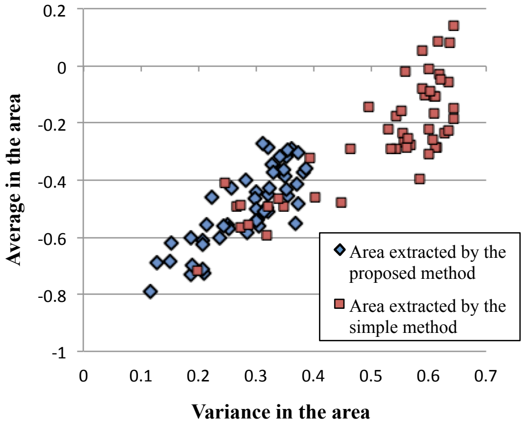


Figure 7: Distribution of Areas Extracted by Proposed Method and Simple Extension Method

ments, because it is expected to extract the areas whose samples have low phenotype values, we confirm that the proposed method extracts the area whose phenotype value is low.

4.2 Result and Discussion

Tables 3 and 4 show the results of the ranking of top 10 combinations of proteins with respect to the area-scores. Table 3 is the result of the case where we applied the proposed method to the expression profiles and the phenotype data. On the other hand, Table 4 is the result of the simple method. These tables include the columns of protein ID of proteins A and B, the number of samples in the area, the area-score, $V[D^{(n)}_{m,(i_1,i_2)}]$ and $E[D^{(n)}_{m,(i_1,i_2)}]$. Note that, in Table 3 and Table 4, we leave only the best area out of the same protein pairs.

These results show that both $V[D^{(n)}_{m,(i_1,i_2)}]$ and $E[D^{(n)}_{m,(i_1,i_2)}]$ in the proposed method are smaller than those in the simple method. It was found from the result that the proposed method could extract areas better than the simple method. In order to confirm it in detail, Figure 7 shows the scatter plots of the ranking of the top 50 areas extracted by the proposed method and the simple method. The vertical axis represents $E[D^{(n)}_{m,(i_1,i_2)}]$ and the horizontal axis represents $V[D^{(n)}_{m,(i_1,i_2)}]$. As is apparent from Figure 7, both $E[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ extracted by the proposed method is found to be lower values than those of the simple method. From these results, we confirmed that the phenotype values of the areas extracted by the proposed method are lower than those extracted by the simple method, and the samples included in the area have close phenotype value each other. In other words, it can be said that the proposed method can extract “good area,” compared with the simple method.

Next, we confirm whether the shape of the area extracted by the proposed method is “good shape” or not. As a typical example of the extracted areas, we show the shape of the rank-1 area in Figure 8.

Figure 8 shows the scatter diagram of the rank-1 area in Table 3. The horizontal axis and the vertical axis represent the standardized expression levels of protein A and B, respectively. The shape-cost of the area is 0.7003, which is the value close to threshold $T_{thresh} = 0.7$. We found that this area is close to a circular shape to same extent and is allowable as an area. That is, the shape of this area extracted by the proposed method is “good shape.”

Then, we see whether this area is a “good area” or not by examining the phenotype value of the samples in the rank-1 area in Table 3. Figure 9 shows the histogram of the phenotype values included in the area, and a histogram of the phenotype values of all samples. The vertical axis represents the number of samples and the horizontal axis represents the carcass weight. Since the carcass weight has been standardized, the average of the carcass weight of all samples is 0, and the variance is 1. The phenotype values in the extended area are distributed in a relatively narrow range between -1.5 and 0.5, and the distribution is unimodal. We find that $V[D^{(n)}_{m,(i_1,i_2)}] = 0.2510$ is considerably lower than the whole variance 1. Moreover, the $E[D^{(n)}_{m,(i_1,i_2)}] = -0.5539$ is sufficiently smaller than the whole average 0.

From the above reasons, we found that the area extracted by the proposed method is the area that we want to find because both $V[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ are small enough.

5 Conclusion

In this paper, we proposed a method to find areas with close phenotype values to predict proteins that control phenotypes. By extracting areas including samples with close phenotype values, which rarely occur statistically, it is possible to esti-

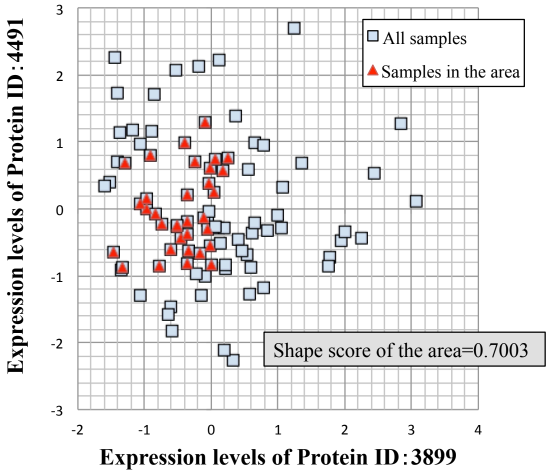


Figure 8: Distribution of Areas of Rank 1 in Table 3

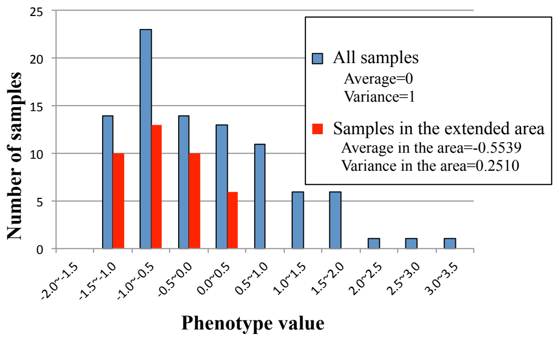


Figure 9: Histogram of Areas of Rank 1 in Table 3

mate the relationship among two proteins and a phenotype.

We performed the evaluation experiment using real data set obtained by the author's collaborative work in Wakayama [7]. In order to evaluate the performance of the proposed method, we implemented a simple method to be compared with the proposed method. As a result, we found that the proposed method extracted the area better than the simple method. That is, the proposed method is able to extract the area that the variance of the phenotype values in the area is small.

Acknowledgement

This work was partly supported by "the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry" of NARO (National Agriculture and Food Research Organization), and "the Program for Promotion of Stockbreeding" of JRA (Japan Racing Association).

REFERENCES

[1] A. Malcolm Campbell, Laurie J. Heyer, "Discovering Genomics, Proteomics and Bioinformatics," Benjamin Cummings, 2006.

[2] Y Qu, S Xu, "Quantitative trait associated microarray gene expression data analysis," Molecular Biology and Evolution, vol.23, no.8, pp.1558-1573, August, 2006.

[3] J. Zhang, Y. Ji, L. Zhang, "Extracting Three-way Gene Interactions from Microarray Data", Bioinformatics, vol.23, no.21, pp2903–2909, 2007.

[4] E. Inoue, S. Murakami, T. Fujiki, T. Yoshihiro, A. Takemoto, H. Ikegami, K. Matsumoto, and M. Nakagawa, "Predicting Three-way Interactions of Proteins from Expression Profiles Based on Correlation Coefficient," IPSJ Transactions on Bioinformatics, vol. 5, pp34–43, 2012.

[5] T. Fujiki, E. Inoue, T. Yoshihiro, M. Nakagawa, "Prediction of Combinatorial Protein-Protein Interaction from Expression Data Based on Conditional Probability", In: Protein-Protein Interactions - Computational and Experimental Tools, InTech Web Press, pp131–146, 2012.

[6] M. de Berg, O. Cheong, M. van Kreveld, M. Overmars, "Computational Geometry: Algorithms and Applications 3rd ed," Springer, 2008.

[7] Collaboration of Regional Entities for the Advancement of Technological Excellence in Wakayama, http://www.yarukiouendan.jp/techno/kessyu/

[8] K. Nagai, T. Yoshihiro, E. Inoue, H. Ikegami, Y. Sono, H. Kawaji, N. Kobayashi, T. Matsuhashi, T. Otani, K Morimoto, M. Nakagawa, A. Iritani and K. Matsumoto, "Developing an Integrated Database System for the Large-scale Proteomic Analysis of Japanese Black Cattle," Animal Science Journal, vol.79, no.4, 2008. (in Japanese)

A cooperative GPS/GNSS positioning method with neighboring receivers

Tomoya Kitani*, Hiroyuki Hatano, Masahiro Fujii, Atsushi Ito and Yu Watanabe
* Graduate School of Informatics, Shizuoka University, Japan
Graduate School of Engineering, Utsunomiya University, Japan
t-kitani@kitanilab.org

Abstract - In this paper, we propose an accurate single point positioning method for GPS (Global Positioning System) and a cooperative positioning method with neighboring users' receivers via communication networks, called the mobile differential GPS (mDGPS). To improve the accuracy and precision of a single point positioning, we utilize the residual sum of squares to eliminate pseudoranges that include gross error. In our methods, a receiver that estimates its position accurately with a single point positioning is used as a base station; it calculates differential correction of each pseudorange according to its measured position; it distributes the differential corrections to neighboring receivers via communication networks. Such a neighboring receiver can improve its positioning using the received differential corrections instead of the traditional navigation messages from satellites. The proposed methods need no other information than the pseudoranges and the navigation messages that conventional receivers have used for positioning. Through an experiment under the situation that a few pseudoranges have been added 20 meters as the intentional measurement error, the our methods have achieved the horizontal positioning error within 2 meters with 6 available satellites whereas that of the conventional one is more than 10 meters with 8 available satellites.

Keywords: GPS/GNSS, cooperative positioning, ad hoc network

1 Introduction

Location information has been utilized in various systems and devices. Services utilizing users' location information are called Location-based services (LBS). In the field of intelligent transport systems (ITS), location information is expected to be used for traffic safety assistance.

To obtain self-location outsides, GNSS (Global Navigation Satellite System) is common and GPS (Global Positioning System) provided by U.S. is the most popular[1]. It is preferable that such a positioning system is accurate and precious anywhere.

In GNSS, a receiver measures the distance between itself and a satellite with the received signal from the satellite. The distance called the *pseudorange* is derived from the delay of the signal from the satellite and the received message called the *navigation message* that contains information about the satellite position. Each pseudorange contains the measurement error by several factors and they degrade the accuracy of position determination. The most significant factor of them is the clock bias in a receiver, and it is treated as the forth unknown value to be estimated, following the three-dimensional coordinate position of the receiver. The most significant factors among the rest error factors are the delay in passing through the ionosphere and troposphere, hereafter referred to as the *atmospheric delay*, and the *multipath delay* due to structural objects on the ground. Atmospheric delays are geographically correlated and they are partially corrected by the information in the navigation message but they can be just halved. Multipath delays often occur where there are lots structures such as in an urban area. The accuracy of position determination is seriously degraded when the measurement error of any of the measured pseudoranges is large. To improve the accuracy, such pseudoranges should be excluded or corrected before position determination.

Since atmospheric delays are geographically correlated, a base station that knows its true position measures the atmospheric delay of each satellite at the position correctly and distributes the measured information called the *differential corrections* so that a neighboring receiver corrects its delays. It is called as the differential GPS (DGPS) and it has been used in the world. However, the real-time broadcast of differential corrections for civil receivers has been stopped in 2008 in Japan. Additionally, it is difficult for DGPS to correct the multipath delay, one of the most significant factors of the positioning error, because the multipath delay is not correlated geographically as much as the atmosphere delay.

In this paper, we propose an accurate single point positioning method for GPS and a cooperative positioning method with neighboring receivers via communication networks. We call this system the *mobile differential GPS (mDGPS)*. A receiver that estimate its position accurately by itself acts as a base station of DGPS, and it generates the information to correct measurement errors of pseudoranges for neighboring receivers.

The position determination of GPS is generally conducted by the use of the least-squares method. The accuracy of the positioning is seriously degraded when any of the pseudoranges used as the inputs for the method contains gross error. To improve the accuracy, such incorrect pseudoranges should be excluded or corrected before position determination.

In the proposed method, a GPS receiver acting as a base station detects pseudoranges that make the positioning error large, changing the combination of satellites to be used for position determination. Then, it can conduct the single point positioning with only accurate pseudoranges. Such a positioning result is generally accurate and precious.

After deriving the position with the above method, a receiver as a base station generates the differential correction for each pseudorange. A neighboring receiver as a user station receives the differential correction, corrects its measured pseudoranges, and conducts the position determination. The proposed method, mDGPS, is effective even when the number of visible satellites is small and some of pseudoranges contain

Table 1: Characteristics of Error Factors in Pseudorange

factor	size	obliquity	temporal	spatial
	[m]	factor	correlation	correlation
satellite clock bias	a few	–	15 min.	∞
satellite position	a few	–	15 min.	1000 km
ionosphere delay	0 to 20	1 to 3	15 min.	100 km
troposphere delay	2.4 at sea level	1 to 10	30 min. 30 min.	100 km 100 km
multipath delay	a few to more than dozen	lower angle, larger	a few min.	very narrow
receiver noise	less than 1	–	none	none

gross error. Through an experiment with the measured results with real receivers, we will evaluate the accuracy and precision of positioning with mDGPS.

We mention about the position determination of GNSS and related work in Sec. 2, the proposed accurate single point positioning method for a base station of mDGPS in Sec. 3, the proposed cooperative positioning method in Sec. 4, and the evaluation result in Sec. 5. Finally, we summarize this paper in Sec. 6.

2 Position Determination of GNSS and Related Work

2.1 Factors of Positioning Error of GNSS

As measuring a pseudorange in GNSS, the following factors mainly cause the measurement error except the clock bias in a receiver, treated as unknown value to be estimated: the clock bias in a satellite, the ionosphere delay, the troposphere delay and the random noise at a receiver. The position of a satellite is provided in a navigation message, and it also contains error slightly. Moreover, the multipath delay occurs in a pseudorange when the receiver cannot receive the signal directly and received the signal after reflected by a ground structure, and it will be significant.

Table 1 shows the quantitative characteristics of the error factors [1], [2]. The obliquity factor is used when a satellite is at low elevation angle. The ionosphere delay and the troposphere delay are approximated as a 8-dimensional model and the part of a navigation message called the *ephemeris* which contains its parameters. Each size of those delays is corrected by the model. The atmospheric delay (the ionosphere delay and the troposphere delay) and the multipath delay are most significant among the above factors.

2.2 Positioning Error Reduction

2.2.1 Multipath Avoidance

There are two prominent types of researches to avoid the effect of multipath to the position determination. These basically work on the problem before measuring a pseudorange. The first type is to separate the direct wave and the multipath waves of the signal from a satellite when a receiver receives both waves so that the pseudorange is measured by the direct wave. The following related works have been studied: designing the signal tracking module[3]–[6], improving the modulation method[7], smoothing the signal carrier[8], and designing the antenna[9]. The second type is to detect and eliminate a signal that are received via multipath when its direct wave cannot be received due to any obstacle. The positioning accuracy is degraded if the position determination is conducted with a biased pseudorange which contains a multipath delay. The direct signal cannot be received if there are any obstacle on the line of sight between a receiver and a satellite. It has been proposed to use a 3D-map of ground structures and a camera in order to detect such obstacles[10], [11].

Unlike the above related work, we focus on detecting multipath after measuring pseudoranges. Our proposed methods do not need any extra devices, information and hardware modification to detect a pseudorange that contains gross errors such as multipath delay.

2.2.2 DGPS: Differential GPS

Among all the error factors in a pseudorange, the clock bias and position error of each satellite can be canceled completely by using the differential between the measured result of two receivers. Since the atmosphere delay is temporary-and-spatially correlated, it can be almost canceled by using the measured result by a stationary receiver, called a *base station*, that knows its true position. A base station distributes its measured information called the *differential corrections* so that a neighboring receiver, called a *user station*, corrects its delays. It is called the differential GPS (DGPS) and has been used in the world.

DGPS is effective to improve the accuracy of GPS. However, the real-time broadcast of differential corrections for civil receivers has been stopped in 2008 in Japan because the intentional dilution of precision of GPS, called SA (Selected Availability), had been released in 2001 and the maintenance cost of the base stations is getting higher. Currently, the differential corrections are available via the web site of GEONET (GNSS Earth Observation Network System)[12] in Japan but they are not provided in real time, and they are just used for location surveys. There are 27 DGPS base stations and 1200 GEONET base stations in Japan. A DGPS base station is built every 200km on the coast and a GEONET base station is built every a few km squared patch of land basically. Since multipath delays occur locally, it cannot be eliminated by the conventional DGPS.

In our proposed methods, a receiver that estimates its position accurately acts as a base station. Such a receiver, we call it a *mobile base station*, calculates the differential corrections and distributes them to neighboring receivers via wireless communication networks. Because the distance between such base station and a neighboring receiver is close, it could find a correlation between the multipath delays of two nodes that are at the same distance from a ground structure that may reflect the signal from a satellite. To utilize the correlation between the multipath delays of neighboring two nodes, Tang et al. improve the accuracy of the relative positioning for the two nodes[13]. Furukawa et al. shows that 89.7% of the multipath delays is correlated between each receiver at two tandem vehicles on roads through experiments[14]. In our methods, we utilize this correlation to improve the accuracy of the absolute positioning.

As related work, the method called ”simple DGPS” has been proposed in [15]: A receiver measures the horizontal absolute positioning error, and it shares the amount of the error with a neighboring receiver. In the method, a base station needs to know its true position somehow, and the correction works well at a neighboring receiver only if the same combination of satellites to be used for position determination and the positioning algorithm in both of the two receivers. A base station can calculate and distribute the differential for all possible combinations from the visible satellites but it is costly.

In our methods, a receiver that acts as a base station does not need to know its true position. Moreover, the differential corrections are calculated for every satellite, not every combination of satellites, like the conventional DGPS. Therefore, both the calculation time and the size of the difference corrections are small. The target of our methods is to improve the accuracy of the absolute positioning.

2.2.3 Satellite Selection

In GNSS, a receiver needs at least 4 visible satellites to solve 4 unknown values for position determination with the least-squares method. The alignment of visible satellites is quite important to obtain the position of a receiver accurately. The position determination works well when visible satellites are spread in any azimuth and elevation angle. Meanwhile, even if the number of visible satellites is large enough but most of them are around the same azimuth, it is hard to make the position converge and slight measurement error in a pseudorange could cause gross positioning error. The variability of visible satellites is expressed as the index DOP (dilution of precision). DOP is derived from the positions of a receiver and its visible satellites, and it is smaller when the variability is smaller. The positioning accuracy is associated with DOP rather than the number of visible satellites. As shown in Fig. 1 later, the accuracy and precision is good enough when the number of visible satellites are 6 or more.

When DOP is still small even if a satellite has been eliminated and the pseudorange of the satellite contains gross errors, it is possible to improve the positioning accuracy by excluding the satellite. On the other hand, when one of the visible satellites has been eliminated and DOP becomes much worse, it is better not to exclude the satellite for position determination. The former feature can be used to improve the positioning accuracy in a mobile base station of our methods. The latter feature indicates that the accuracy can be improved if a pseudorange contains gross errors but it is able to be corrected before positioning.

In the pseudorange of a satellite at a low elevation angle, the atmosphere delay and the multipath delay may become large as shown in Tab. 1, To improve the positioning accuracy, a general positioning algorithm for GPS devises the following tactics: eliminating a pseudorange from positioning if the elevation angle of its satellite is lower than a threshold or if the residual of the pseudorange becomes larger than a threshold during positioning calculation, and weighting each pseudorange according to the elevation angle of its satellite [1], [2]. However, the order of the amount of errors in each pseudorange is poorly correlated with that of the residual of them after calculation with the least-squares method. Therefore, it is hard to detect pseudoranges that contain gross errors from the amount of the positioning error.

A method to detect pseudoranges that contains gross errors has been proposed [16]. It classes the possible combinations of satellites for positioning according to the derived position of each combination, and it estimates which class contains the true position. However, it does not work well even if two pseudoranges contain gross errors because the size of the cluster that contains the true position will be very small and it is hard to detect the class. Also, to calculate the position for all possible combinations is costly.

Our methods can detect all pseudoranges that contain gross error when at least n pseudoranges are measured well and these n satellites are located with a small DOP. The value n should be 5 or more, but it is usually 5 or 6 to let the methods work well practically.

2.3 Extra Information to improve Accuracy

WiFi Positioning systems have been in practical use. Such a system connects a WiFi access point and its location, and it lets a client node estimate its position using the connection information and the signal strength from thr access point. Such systems have been available by Google Inc. and Skyhook Wireless Inc.[17] in US, and Koozyt Inc. has provided the system named Place Engine[18] in Japan. Those systems with GPS have provided accurate positioning services, and they are called Hybrid Positioning System.

Our methods do not need such extra information, focusing on improving the positioning accuracy using only GPS receivers as devices.

3 Accurate Single Point Positioning Method utilizing Pseudorange Measurement Residuals

First, we propose a single point positioning method to improve the positioning accuracy with the residual of each pseudorange after positioning calculation.

As mentioned before, in GNSS, the position is derived from pseudoranges and their own satellite's position with the least-squares method. In the least-squares method, the position is estimated so that the residual sum of squares of the pseudoranges is minimized. The residual sum should be 0 if any pseudorange does not contain errors. Given at least one pseudorange that contains gross errors, the residual of each pseudorange is biased.

In this section, we propose a method to detect pseudoranges that make the positioning error large by changing the combination of satellites to be used for position determination.

Hereafter in this paper, we call a pseudorange that contains gross error an *inaccurate-pseudorange*, and we also call a pseudorange that contains little error an *accurate-pseudorange*.

3.1 Positioning Experiment

We have conducted an experiment at the 4th graded triangulation station named "Hosono," located at N 34°45'46.1622", E 135°47'35.6252"[19], on December 27, 2012. We used a

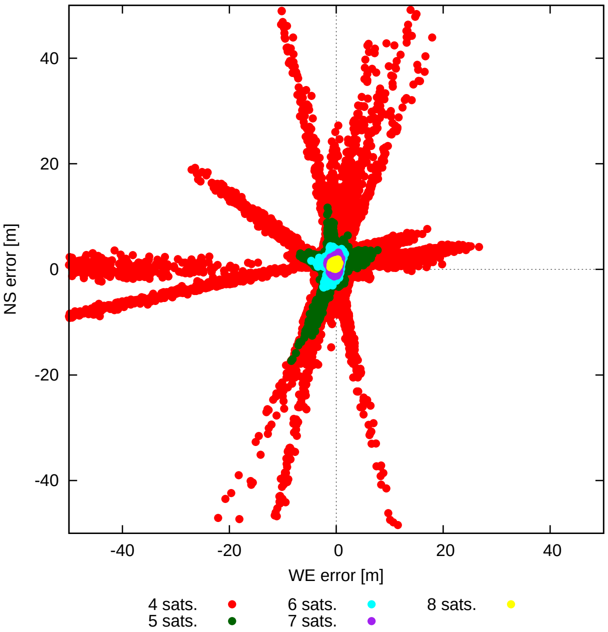


Figure 1: Positioning Results of Single Point Positioning by General Algorithm

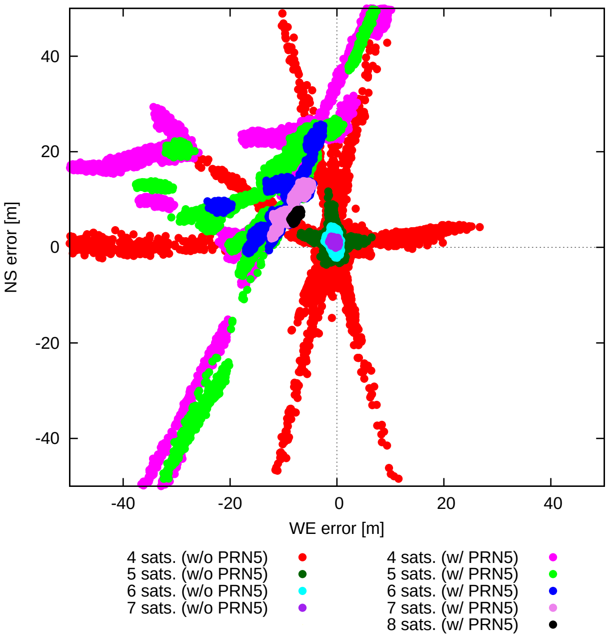


Figure 2: Positioning Results with 20 meters Intentional Error for PRN5

receiver DELTA by JAVAD GNSS Inc.[20] and have obtained the measurement and positioning result every second. This receiver outputs a result in the format RINEX[21] and the result contains the pseudorange of any visible satellites and the navigation messages from the satellites. We have obtained results after waiting time enough to receive the navigation messages completely, and we use the results of ten minutes between 17:28 to 17:38 on that day in this paper. There are no high structures to avoid the line of sight between the receiver and the 8 visible satellites, and all signals from the satellites had been received in good condition.

3.1.1 Positioning by General Algorithm

We use the positioning algorithm in [2]. The algorithm is usual and based on the least-squares method and it weights each pseudorange according to the elevation angle of its satellite. According to the experimental results in [1], [2], given a pseudorange and its satellite located at the elevation angle θ, the measurement error of the pseudorange is modeled as $\sigma_{EL}(\theta) = \frac{0.8}{\sin\theta}$. Therefore, the algorithm utilizes the inverse of $\sigma_{EL}(\theta)$ as the weight for the pseudorange so that the pseudorange of a high elevation satellite is prioritized.

Figure 1 shows the positioning results for 600 seconds and all possible combinations among the 8 satellites. The center (0,0) is the true position. The results are almost converged to the true position. Positioning results far from the true position are because the DOP of satellites used for position determination is high, and, in most case, the number of satellites is less than 6.

3.1.2 Positioning Trial with an Inaccurate Pseudorange

In this trial, we use the same measurement results as Sec. 3.1.1. We have conducted the positioning trial with the following condition: the position determination will be conducted after 20 meters as an intentional delay is added to the pseudorange of Satellite #5 (PRN5). Figure 2 shows the positioning results. Just one inaccurate-pseudorange makes the positioning errors biased. In other words, it makes the accuracy worse but it does not make the precision worse.

We have confirm from Fig. 2 that it is better to eliminate an inaccurate-pseudorange (PRN 5) to improve the positioning accuracy, as mentioned before. For instance, the results by 7 satellites excluding PRN 5 are much more accurate the results by 8 satellites including PRN 5.

3.1.3 Discussion about Convergence of Positioning by Least-Squares Method

We have conducted the following position determinations with the measured result of DELTA at the moment 17:28:00 on that day.

(1) Positioning with 8 pseudoranges measured by signals received in good condition

(2) Same as (1) except that 20 meter as an intentional delay is added to the pseudorange PRN 5

(3) Same as (1) but eliminated the pseudorange PRN 5 and positioning with the 7 pseudoranges

Table 2 shows the positioning parameters obtained from the navigation message and the residual of each pseudorange after the positioning process. The residuals of each pseudorange d_i in Tab. 2(b) are reduced the atmospheric corrections. Item (0) in Tab. 2(b) shows the residual when all pseudoranges did not contain any errors. The value “H error” denotes the positioning error. In other words, it is the horizontal distance between the true position and the positioning result.

The value $\sum(d_i/w_i)^2$ is the sum of square of the residuals of every pseudorange as the evaluation function in the least-squares method to be minimized.

Before the positioning process in GNSS, the atmospheric correction a_i for a satellite i is added to the pseudorange r_i, and $r'_i = r_i + a_i$ is used for positioning instead of r_i. The positioning result will converge when r'_i is equal to the true distance between the receiver and the satellite i for any i. Actually, as shown in Tab. 2(b)-(0), r'_i contains errors slightly.

The horizontal error is about 1 meter and the residuals are also small when the signals are received in good condition from all satellites that are used for the positioning process, as shown in Tab. 2(b)-(1) and (3). On the contrary, as shown in Tab. 2(b)-(2), the horizontal error is considerably large and it is about 10 meters. Although PRN 5 contains $+20$ meters-delay as the intentional delay, the residual d_5 does not become 20 meter larger than that of (1). The process of the least-squares method and the weight makes the residuals biased globally.

In (3), PRN 5 is eliminated before the positioning process but the residual of PRN 5, d_5, can be calculated back form the position derived the process and the position of Satellite #5 from the navigation messages. The residual of PRN 5 in (3) is derived as $d_5 = -21.08$. The difference of d_5 between (0) and (3) is obtained as $\Delta d_5 = -20.29$, and it is the correction that can cancel the added intentional delay 20 meters.

Due to the feature of the least-squares method, the accuracy is significantly degraded when at least one initial pseudorange r'_i for the positioning process has large difference from the true range between the receiver and the position of Satellite i. It is regardless of which factor causes the gross error. An inaccurate-pseudorange, not only due to multipath delay but also due to other factors, should be eliminated before the positioning process in order to improve the accuracy of position determination.

3.2 Proposed Method to detect Inaccurate-Pseudoranges

In this subsection, we propose a method to eliminate inaccurate-pseudoranges by using the residuals of the least-squares method in the positioning process. As shown in Tab. 2(b), the residual sum of square of the method after positioning process varies according to the combination of satellites. The residual sum of square of a combination of satellites gets considerably large if the combination contains inaccurate-pseudoranges.

The algorithm of the proposed method is following.

[Algorithm 1] Classifying Pseudoranges by Size of Errors

```
Inputs   $S$: the pseudoranges of all the visible satellites
         $th$: the threshold of classification
         $f(s)$: the function to obtain the residual sum of
         square after positioning with satellites $s$
Outputs  $C$: the set of accurate-pseudoranges
         $E$: the set of inaccurate-pseudoranges
vars.    $s$: selected satellite(s) from $S$
 1:  $C \leftarrow \emptyset, E \leftarrow \emptyset$
 2:  $s \leftarrow$ new combination of 5 satellites from $S$
 3:  if such $s$ does not exist then end with failure
 4:  if $f(s) > th$ then goto 2
 5:  $C \leftarrow s, S \leftarrow S \setminus s$
 6:  if $S = \emptyset$ then end with success
 7:  $s \leftarrow$ 1 satellite from $S$
 8:  if $f(C \cup s) < th$
 9:      then $C \leftarrow C \cup s$
10:      else $E \leftarrow E \cup s$
11:  $S \leftarrow S \setminus s$ goto 6
```

In the positioning process on GPS, since the number of the unknown variable is 4, it needs at least 4 pseudoranges to conduct the position determination. With the least-squares method which utilizes over-determination, it needs at least 5 pseudoranges to improve the position determination. Thus, at Line 2 in Algorithm 1, the number of selected satellites is 5. A set of the selected satellites should be found such that it consists of only accurate-pseudoranges, and the algorithm tries to find it at Lines 2 through 4. Given n visible satellites, this process needs to check at most ${}_nC_5$ combinations of the satellites. However, in practice, such combination will be found soon because the number of inaccurate-pseudoranges is less than 2 in most cases.

The proposed method determines the position of a receiver accurately by using only accurate-pseudoranges. Moreover, it can calculate back the differential correction of inaccurate-pseudoranges from the derived position of a receiver. In this paper, we use 10 as the threshold th according to the result shown in Tab. 2(b).

3.3 Our Contributions

The conventional methods conduct following techniques to derive the position of a receiver accurately: eliminating the pseudorange of a satellite if its angle of elevation is lower than a threshold, and eliminating pseudoranges if their residual exceeds a threshold during positioning process. However, an inaccurate-pseudorange does not make its residual large. As shown in Tab. 2(b)-(2), although PRN5 contains gross error, the residual of PRN28 is finally the largest.

The proposed method can detect any inaccurate-pseudoranges when there are at least 5 other accurate-pseudoranges. It does not any extra information and devices such as a map of ground structures which cause multipath delay and cameras which find such structures. It needs only the received information as same as that of the conventional receivers, and it can improve accuracy and precision of position determination.

Table 2: Position Determination at 17:28:00

(a) Parameters for Positioning

PRN number	PRN 5	PRN 9	PRN 15	PRN 18	PRN 21	PRN 24	PRN 26	PRN28
Atmospheric Correction a_i [m]	-8.269	-6.110	-6.168	-13.401	-8.952	-7.276	-8.157	-16.48
Elevation Angle θ_i [deg]	41.4	70.4	66.6	22.4	39.4	52.6	40.8	13.4
Weight $w_i (= 1/\sigma_{EL}(\theta_i))$	0.827	1.178	1.147	0.476	0.793	0.993	0.816	0.291

(b) Residuals after Positioning Process d_i ([m])

PRN number	H error	PRN 5	PRN 9	PRN 15	PRN 18	PRN 21	PRN 24	PRN 26	PRN28	$\sum(d_i/w_i)^2$
(0) Without any error	0	0.794	-2.017	0.293	1.824	1.037	0.382	0.728	1.078	8.226 [m^2]
(1) 8 prns.	0.764	0.592	-1.038	0.611	-0.662	-0.048	0.799	-0.142	-1.647	3.203 [m^2]
(2) 8 prns. added error	9.973	-8.515	0.795	-1.478	-1.578	-2.936	4.158	5.815	12.66	112.6 [m^2]
(3) 7 prns.	1.261	–	-0.924	0.482	-0.719	-0.227	1.008	0.227	-0.759	2.727 [m^2]

4 Cooperative Positioning Method with Neighboring Receivers

Second, we propose a cooperative positioning method with neighboring receivers to improve the positioning accuracy.

There is at least one DGPS base station within a few kilometer-square in Japan. However, it is not possible to correct local errors such as multipath delay using the differential correction information from the base stations because such local errors are caused by ground structures like buildings which is much smaller than a kilometer.

A current GPS receiver is likely to be equipped with a communication device such as a smartphones and a mobile phone, and such a device can communicate with neighboring devices via cellar networks and WiFi ad-hoc wireless networks. Since the accuracy of a single point positioning is getting better, a receiver that received lots signals from satellites in good condition can determine its position almost exactly. Therefore, such a receiver acting as a pseudo base station of DGPS, it is possible to improve the accuracy and precision of position determination for neighboring receivers in real time even if a neighboring receiver does not determine its position well by itself.

4.1 Mobile DGPS, a Proposed Method

As shown later in Fig. 1, the positioning results converge on the true position if 6 and more satellites are visible and all the pseudoranges are accurate-pseudoranges.

Suppose that there is a neighboring receiver, and it does not receive signals in good condition, which means that the number of visible satellites is less and there are some inaccurate-pseudoranges. The positioning results have a biased positioning error and do not converge on its true position as shown later in Fig. 2 if at least one inaccurate-pseudorange is used for position determination and especially the number of accurate-pseudoranges to be used is less than 4. Given the accurate differential correction for inaccurate-pseudoranges, the accuracy of position determination can be improved.

In this paper, we propose a system called *mobile DGPS (mDGPS)*. In mDGPS, a receiver that determines its position almost exactly is acting as a base station of DGPS, and we call such a receiver a *mobile base station*. Neighboring receivers which can receive the differential correction from a mobile base station is called a *user station*.

A mobile base station generates the differential correction of mDGPS, and it consists of the determined position of the mobile base station, the time of the position determination, and the correction value for each satellite. The size of this differential correction is less than 300 bytes if each item is expressed by a 4-byte value and because the maximum number of the visible satellites is 32. It is small enough to be broadcast via cellar networks and WiFi ad-hoc networks.

The protocols of a mobile base station and a user station of mDGPS are shown as follows.

[Protocol 1] Mobile Base Station of mDGPS

Step 1: Confirm that the receiver obtains at least 6 accurate pseudoranges by using the method shown in Sec. 3. If not, it cannot be a mobile base station, and it will act as a user station.

Step 2: Estimate its position with a single point positioning without inaccurate-pseudoranges.

Step 3: Generate the differential correction of the pseudorange of each visible satellite. The differential correction information of each satellite also includes which the pseudorange of the satellite is classified into the set of accurate-pseudoranges C or that of inaccurate-pseudoranges E.

Step 4: Broadcast the differential correction via cellar networks and WiFi ad-hoc networks. Moreover, it can share it through cloud-computing networks via Internet.

[Protocol 2] User Station of mDGPS

Step 1: Confirm that the receiver does not obtain at least 6 accurate pseudoranges by using the method shown in Sec. 3.

Step 2: Receive a differential correction directly via wireless communication or download it via Internet with its position roughly estimated by itself.

Step 3: Estimate its position with the received differential correction instead of that from the navigation messages.

At a user station in mDGPS, inaccurate pseudoranges can be corrected to be accurate pseudoranges with the generated differential correction by a mobile base station. A user station could estimate its position accurately when the number of its visible satellites is 4 or more but the number of accurate-pseudoranges is less than 4 due to multipath delay and other factors.

On the other hand, a user station might not correct such inaccurate-pseudoranges well if factors of the measurement error of the pseudoranges are different between in the user station and in the mobile base station. Note that mDGPS can correct an inaccurate-pseudorange of a user station if the same pseudorange is also inaccurate in the mobile base station. However, in the case where a pseudorange in a mobile base station is inaccurate but the same pseudorange in a user station is not inaccurate, the method shown in Sec. 3 could work well in a user station to select pseudoranges that will be used for position determination.

Here, assume that the number of visible satellites of a user station is more than 4. For example, given the pseudoranges: PRN1, PRN2, $\cdots$, PRN8 of a mobile base station and the pseudoranges: PRN1', PRN2', $\cdots$, PRN5' of a user station, and PRN1, PRN2, PRN1' and PRN2' are inaccurate, mDGPS can correct PRN1' and PRN2' with the differential correction information of PRN1 and PRN2, respectively. On the other hand, in the case where PRN1, PRN2 and PRN1' are inaccurate but PRN2' is accurate, it is not favorable to correct PRN2' by PRN2. In this case, the residual sum of square after positioning should be larger than the threshold in the user station with the overcorrected PRN2'. Since the differential correction information includes which pseudoranges are inaccurate in the mobile base station, it is possible to evaluate which pseudoranges should be corrected in the user station in order to minimize the residual sum of square. Given the number n of inaccurate pseudoranges in the mobile base station, the number of its trial will be 2^n at most. However, this overhead can be acceptable because n is less than 3 in most cases.

In the case where there are only four visible satellites for a user station, mDGPS does not work well for a user station. To utilize the previous positioning results and the time series variation of each pseudorange, mDGPS could work well. This is one of our future work.

5 Evaluation of mDGPS

To evaluate the accuracy of mDGPS, we have conducted the following experiments with two GPS receivers.

5.1 Positioning Experiment

We have conducted another experiment with a receiver AEK-4T by u-box AG[22] at the same time as shown in Sec. 3.1. We put the receiver at the same triangulation station and made it receive signals every second. The receiver also received the signals from 8 visible satellites in good condition. We also use the same positioning algorithm as shown in Sec. 3.1.1.

In order to evaluate the effectiveness of mDGPS, the measured results by AEK-4T in this section is used as the results of a mobile base station, and the measured results by DELTA in Sec. 3.1 is used as the results of a user station. The reason why AEK-4T is used as a mobile base station is because the accuracy and precision of the single point positioning with AEK-4T is worse than that of DELTA. If the accuracy of the single point positioning at a mobile base station is much higher than that at a user station, the total positioning accuracy at the user station might be improved due to the accuracy of the single point positioning at the mobile base station. To distinguish the effectiveness of mDGPS from the above favorable influence by the accuracy of a mobile base station, we used AEK-4T as the mobile base station in this evaluation.

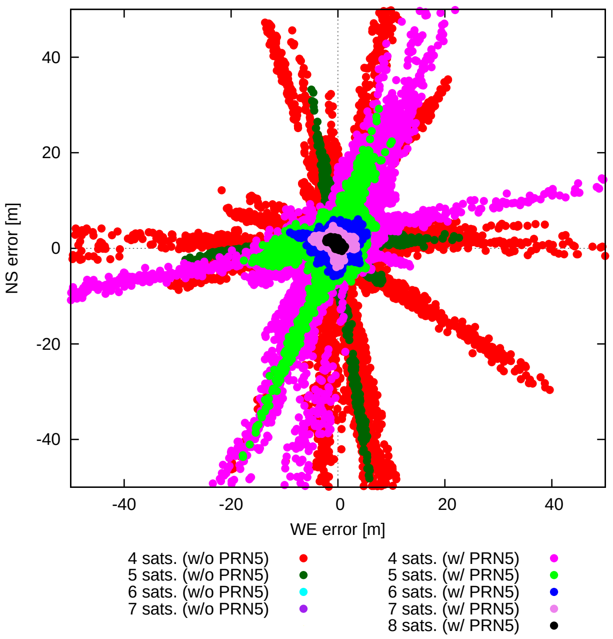


Figure 3: Correction Results with mDGPS (PRN5 contains +20m biased delay)

5.2 Correction Effect with mDGPS

In order to confirm that inaccurate pseudoranges are able to be corrected with mDGPS, the measured pseudorange PRN5 at the mobile base station (AEK-4T) and that at the user station (DELTA) were added +20 meters, an intentional gross error, before the position calculation. AEK-4T, a mobile base station, is based on the assumption that it can estimate its position accurately along Protocol 1. Thus, in this evaluation, it conducted the position determination without an inaccurate pseudorange PRN5, and it also generated the differential correction of each pseudorange, which includes PRN5. DELTA, a user station, conducted the position determination with the differential correction generated by AEK-4T instead of that from the navigation messages.

The above correction results is shown in Fig. 3. Comparing with the results shown in Fig. 2, the biased positioning errors have been corrected and all the positioning results become converged on the true position.

Table 3 shows the difference between the accuracy and precision of the single point positioning and that of mDGPS. Table 3(a) shows the positioning result without the inaccurate pseudorange PRN5, and it is good enough for the single point positioning. Table 3(b) shows the positioning result with the inaccurate pseudorange PRN5, the horizontal positioning error is large even if the number of visible satellites is larger.

Both its accuracy and precision of positioning are low. On the other hand, it has been confirmed that the correction by mDGPS improves the accuracy and precision.

According to the results of mDGPS shown in Tab. 3, the horizontal positioning error of 4 satellites including PRN5 is smaller than that excluding PRN5. This is because Satellite #5 is located at a position that makes DOP small. The proposed method, mDGPS, is especially effective when the number of visible satellites is small and DOP becomes large if eliminating inaccurate pseudoranges.

Table 3: Horizontal Positioning Error with or without mDGPS

(a) Positioning Results without PRN5

# of visible	Single PP [m]		mDGPS [m]	
satellites	Ave. μ	SD σ	Ave. μ	SD σ
4	6.164	15.77	14.15	40.62
5	1.972	1.322	3.232	4.118
6	1.478	0.781	1.727	0.904
7	1.374	0.385	1.374	0.385

(b) Positioning Results with PRN5

# of visible	Single PP [m]		mDGPS [m]	
satellites	Ave. μ	SD σ	Ave. μ	SD σ
4	53.69	118.1	8.457	23.98
5	21.05	10.34	2.885	2.878
6	14.91	3.986	1.956	1.074
7	11.84	1.381	1.591	0.806
8	10.14	0.253	1.335	0.600

Single PP stands for the single point positioning.

5.3 Our Contributions

The proposed method, mDGPS, supposes that a base station conducts the single point positioning by itself to obtain its position, and it does not need to obtain extra information, including its true position, from others. mDGPS has an aspect of relative positioning but it is not necessary to know the true distance between a mobile base station and a user station. In mDGPS, a mobile base station generates the differential correction for each pseudorange. Thus, a user station can correct its measured pseudoranges with received differential correction without regard to the combination of visible satellites.

In mDGPS, it is believed that the distance between a mobile base station and a user station is smaller than 100 meters. Within such short range, the multipath delay could be correlated between these two stations. Although the amount of multipath delay is related with the distance from the structure that cause the delay, mDGPS could work well when the receiver in cars that go through a road because roads are usually constructed parallel to such structures.

6 Conclusions

In this paper, we have proposed methods to improve the accuracy and precision of positioning utilizing a GPS receiver that estimates its position well by itself. Through the experiment under the situation that a few pseudoranges have been added 20 meters as the intentional measurement error, our methods have achieved the horizontal positioning error within 2 meters with 6 available satellites whereas that of the conventional one is more than 10 meters with 8 available satellites.

The proposed methods have focused on the fundamental part of the positioning with GNSS. Thus, in this paper, we do not adopt classic statistical methods to improve the accuracy of positioning such as averaging random noises and time-series processing. The proposed methods have room to be improved with these methods.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26330102.

REFERENCES

[1] Misra, P. and Enge, P.: GLOBAL POSITIONING SYSTEM, Signal, Measurements, and Performance, Second Edition, *Ganga-Jamuna Press* (2006)

[2] Sakai, T.: Practical Programing for GPS, *Tokyo Denki University Press* (2007) (in Japanese), the positioning program is available from <http://www.tdupress.jp/download/sonota-download/k_isbn978-4-501-32550-3.html> (accessed 2013-04-04).

[3] Van Dierendonck, A.J., Fenton, P., and Ford, T.: Theory and Performance of Narrow Correlator Spacing in a GPS Receiver, *Navigation Journal of the Institute of Navigation*, vol. 39, no. 3, pp. 265-283 (1992).

[4] Townsend, B., and Fenton, P.: A practical approach to the reduction of pseudorange multipath errors in a L1 GPS receiver, *Proc. ION GPS-94*, pp. 143–148 (1994).

[5] Veitsel, V.A., Zhdanov, A.V., and Zhodzishsky, M.I.: The mitigation of multipath errors by strobe correlators in GPS/GLONASS receivers, *GPS Solutions*, vol. 2, no. 2, pp. 38–45 (1998).

[6] Van Nee, R.D.J., Siereveld, J., Fenton, P.C., and Townsend, B.R., The multipath estimating delay lock loop: Approaching theoretical accuracy limits, *Proceedings of IEEE Position Location and Navigation Symposium*, pp. 246–251 (1994).

[7] Betz, J.W.: Binary offset carrier modulations for radio navigation, *Navigation, Journal of the Institute of Navigation*, vol. 48, no. 4, pp. 227–246 (2001).

[8] Hatch, R.: The synergism of GPS code and carrier measurements, *Proceedings of 3rd International Geodetic Symposium on Satellite Doppler Positioning*, pp. 1213–1231 (1993).

[9] Rougerie, S., Carrie, G., Vincent, F., Ries, L., and Monnerat, M.: A new multipath mitigation method for GNSS receivers based on an antenna array, *International Journal of Navigation and Observation*, vol. 2012, Article ID 804732 (2012).

[10] Suh, Y. and Shibasaki, R.: Evaluation of Satellite-based Navigation Services in Complex Urban Environments

using a Three-dimensional GIS, *IEICE Trans. Commun.*, vol. E90-B, no. 7, pp. 1816–1825 (2007).

[11] Meguro, J., Murata, T., Takiguchi, J., Amano, Y., and Hashizume, T.: GPS multipath mitigation for urban area using omni directional infrared camera, *IEEE Trans. ITS*, vol. 10, no. 1, pp. 22–30 (2009).

[12] Geodetic Observation Center, Geospatial Information Authority of Japan: GEONET: GNSS Earth Observation Network System, available from <`http://terras.gsi.go.jp/ja/terras_english.html`> (accessed 2013-04-04).

[13] Tang, S., Kubo, N., and Ohashi, M.: Cooperative Relative Positioning for Intelligent Transportation System, *Proceedings of International Conference on ITS Telecommunications*, pp. 506–511 (2012).

[14] Furukawa, R., Tang. S., Kawanishi, N., and Ohashi, M.: Evaluation and Analysis of Correlation in Reflected Signals and Its Application in Cooperative Relative Positioning, *ITS World Congress*, (2013) (to appear).

[15] Miyata, H., Noguchi, T., Sakitani, A., and Egashira, S.: Methods for Improving the Positioning Accuracy with Simplified DGPS, *Technical Report of IEICE*, vol. AP95-97, pp. 39–46 (1996), (in Japanese).

[16] Shibazaki, R. and Yang-Won, L.: Positioning Method and Positioning Device using Positioning Data from Satellite, *Patent of Japan*, Publication number:2011-013132, Jan. 20, 2011, Application number:2009-158501, Jul. 3, 2009.

[17] Skyhook, Inc.: Skyhook > home, available from <`http://www.skyhookwireless.com/`> (accessed 2013-02-15).

[18] Koozyt, Inc.: Wireless LAN · Location Information — Place Engine, available from <`http://www.placeengine.com/`> (accessed 2013-02-15).

[19] Geospatial Information Authority of Japan: Control Point Survey Results, (in Japanese), available from <`http://sokuseikagis1.gsi.go.jp/`> (accessed 2013-02-15).

[20] JAVAD GNSS Inc.: DELTA, available from <`http://www.javad.com/jgnss/products/receivers/delta.html`> (accessed 2013-02-15).

[21] Gurtner, W.: *RINEX: The Receiver Independent Exchange Format Version 2.10*, (2002).

[22] u-blox AG: leading provider of GPS/GNSS receiver modules and 2G/3G/GSM/UMTS/CDMA modem modules, GPS/GNSS receiver chips plus complete GPS/Wireless software, solutions and reference designs, available from <`http://www.u-blox.com/`> (accessed 2013-02-15).

Keynote Speech 2:
Dr. Ichiro Iida
(Fujitsu Laboratories Ltd.)

IWIN2014

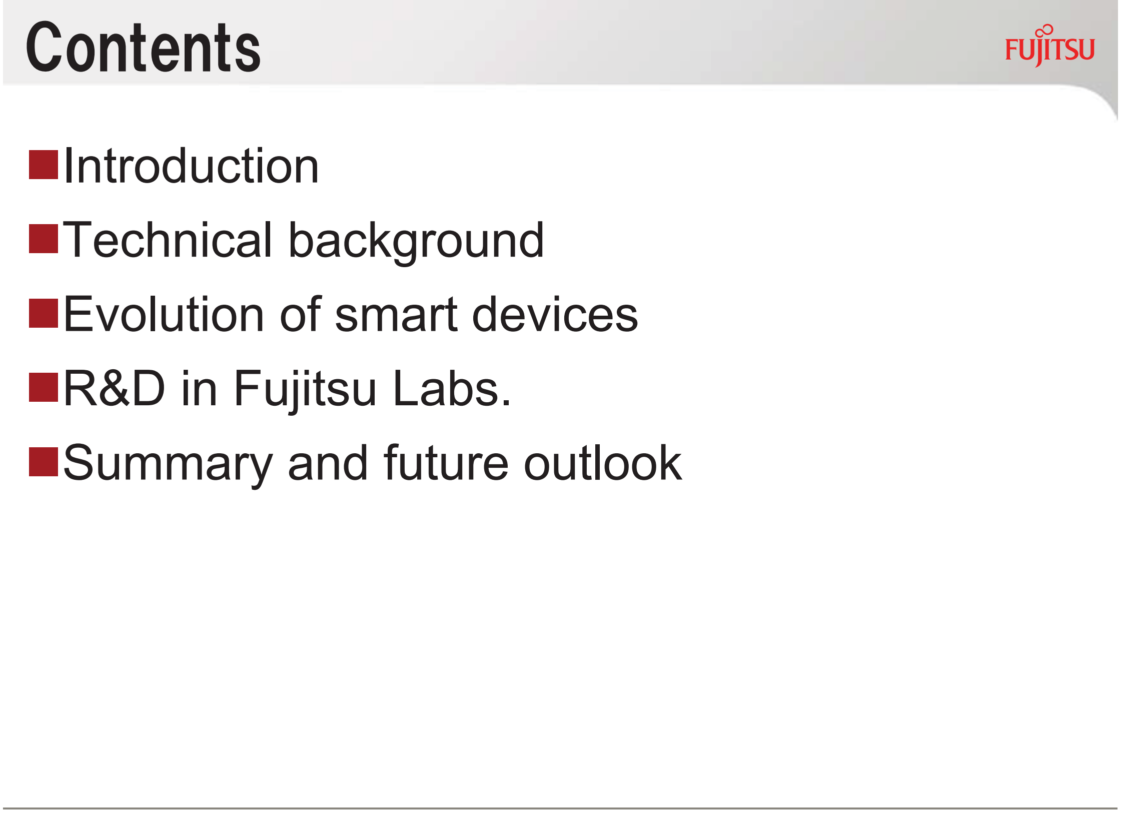
FUJITSU

shaping tomorrow with you

From "Web Computing" to "Front Computing"

The complex of
enterprise and consumer ICT

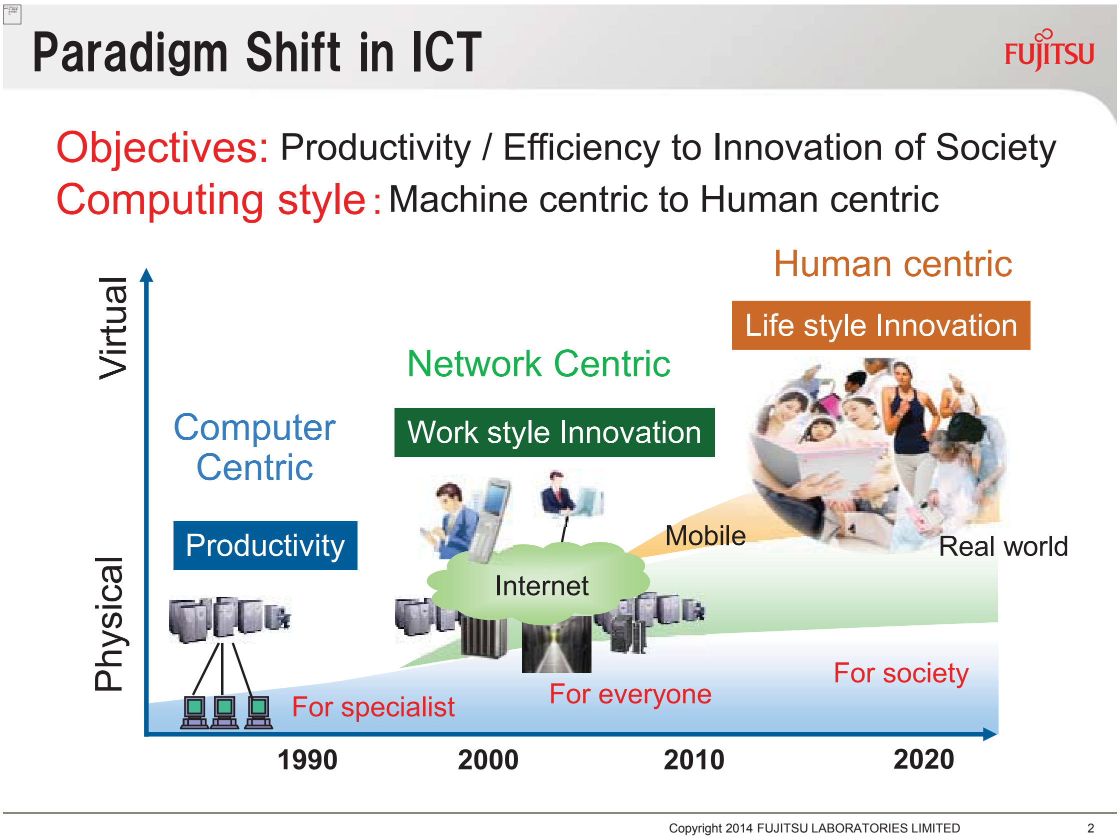
Sept.11 2014
Ichiro Iida
Fujitsu Laboratories Ltd.



Contents

FUJITSU

- Introduction
- Technical background
- Evolution of smart devices
- R&D in Fujitsu Labs.
- Summary and future outlook



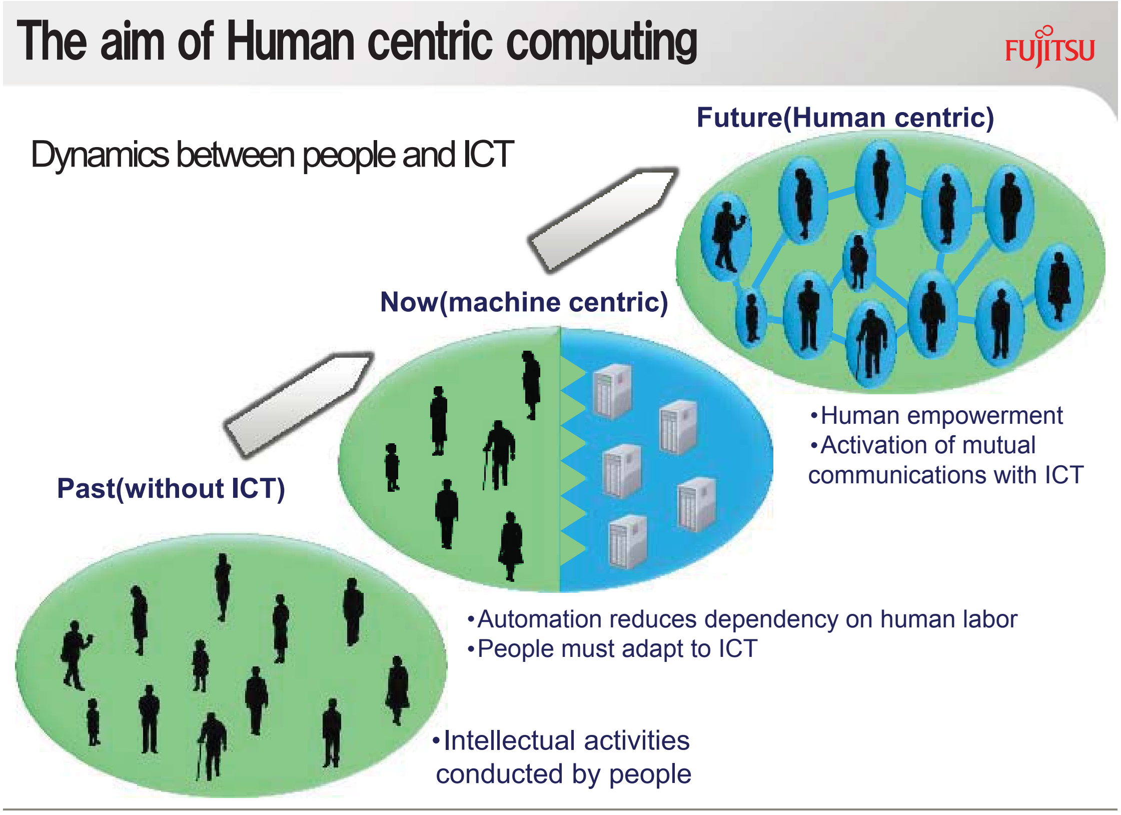
Paradigm Shift in ICT

FUJITSU

Objectives: Productivity / Efficiency to Innovation of Society

Computing style : Machine centric to Human centric

2



The aim of Human centric computing

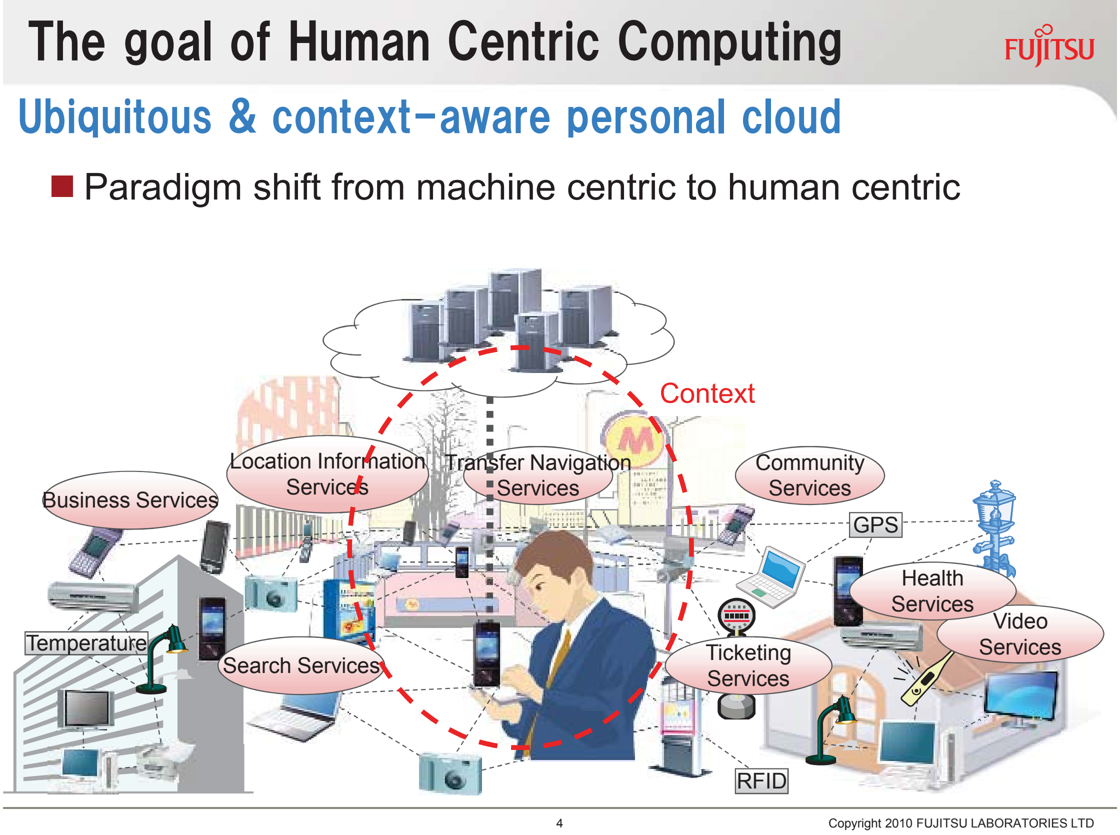
FUJITSU

Dynamics between people and ICT

- Human empowerment
- Activation of mutual communications with ICT

- Automation reduces dependency on human labor
- People must adapt to ICT

- Intellectual activities conducted by people



The goal of Human Centric Computing

FUJITSU

Ubiquitous & context-aware personal cloud

- Paradigm shift from machine centric to human centric

4

Copyright 2010 FUJITSU LABORATORIES LTD

Evolution of related technologies

FUJITSU

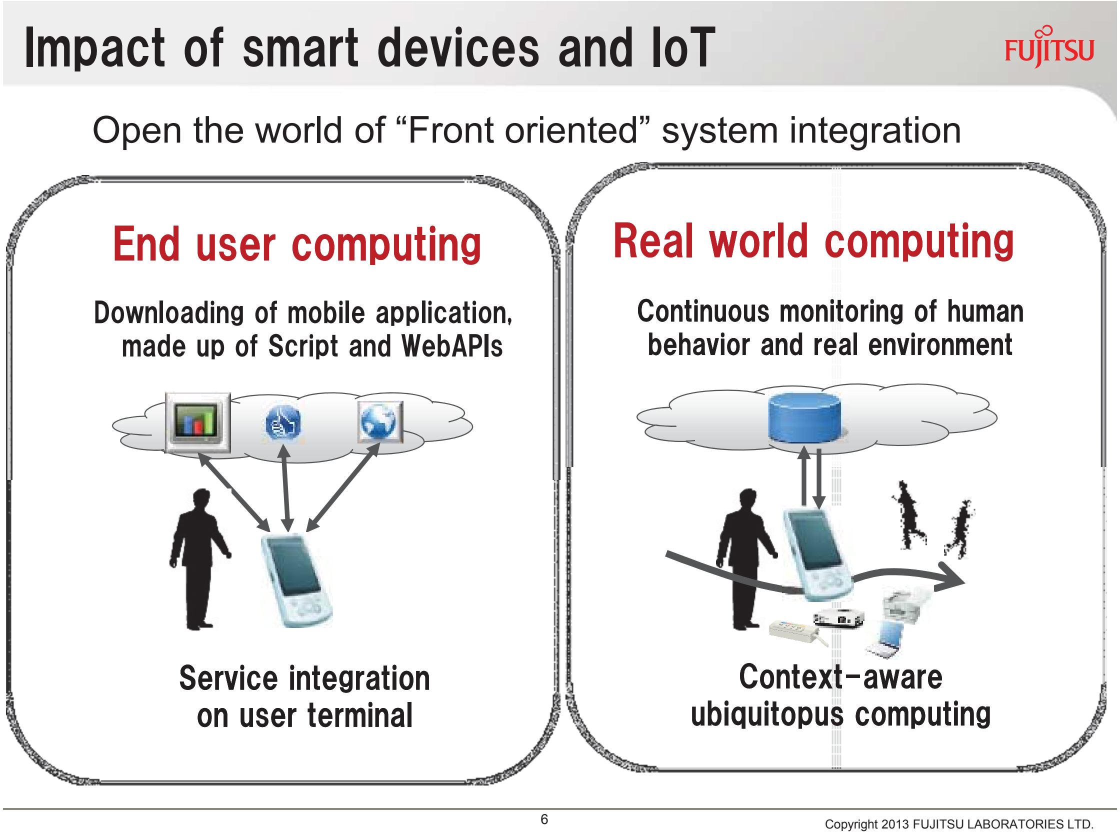
- The recent important innovative technologies

Mobile computing (Smart device)	The innovation in computing model (Web pages to Web APIs)
Cloud service (Service platform)	The virtualization of hardware (Location independency)
Internet of Things (Gadget & sensor)	Networking of sensors and gadget (Terminal access to ambient access)

These innovations come from the consumer fields

5

Copyright 2014 FUJITSU LABORATORIES LIMITED



Impact of smart devices and IoT

FUJITSU

Open the world of “Front oriented” system integration

End user computing

Downloading of mobile application, made up of Script and WebAPIs

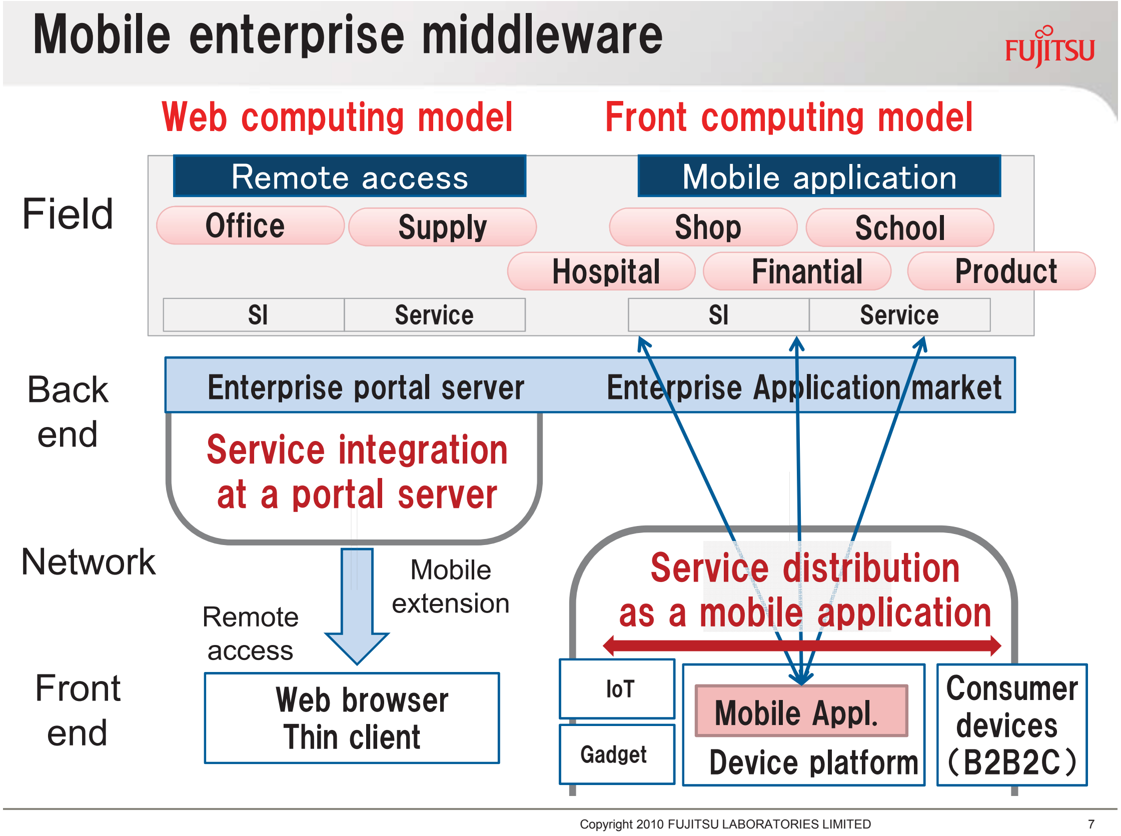
Service integration on user terminal

Real world computing

Continuous monitoring of human behavior and real environment

Context–aware ubiquitopus computing

6



Mobile enterprise middleware

FUJITSU

Web computing model

Front computing model

Field

Remote access

Office

Supply

SI

Service

Mobile application

Shop

School

Hospital

Finantial

Product

SI

Service

Back end

Enterprise portal server

Enterprise Application market

Service integration at a portal server

Network

Mobile extension

Remote access

Service distribution as a mobile application

Front end

Web browser Thin client

IoT

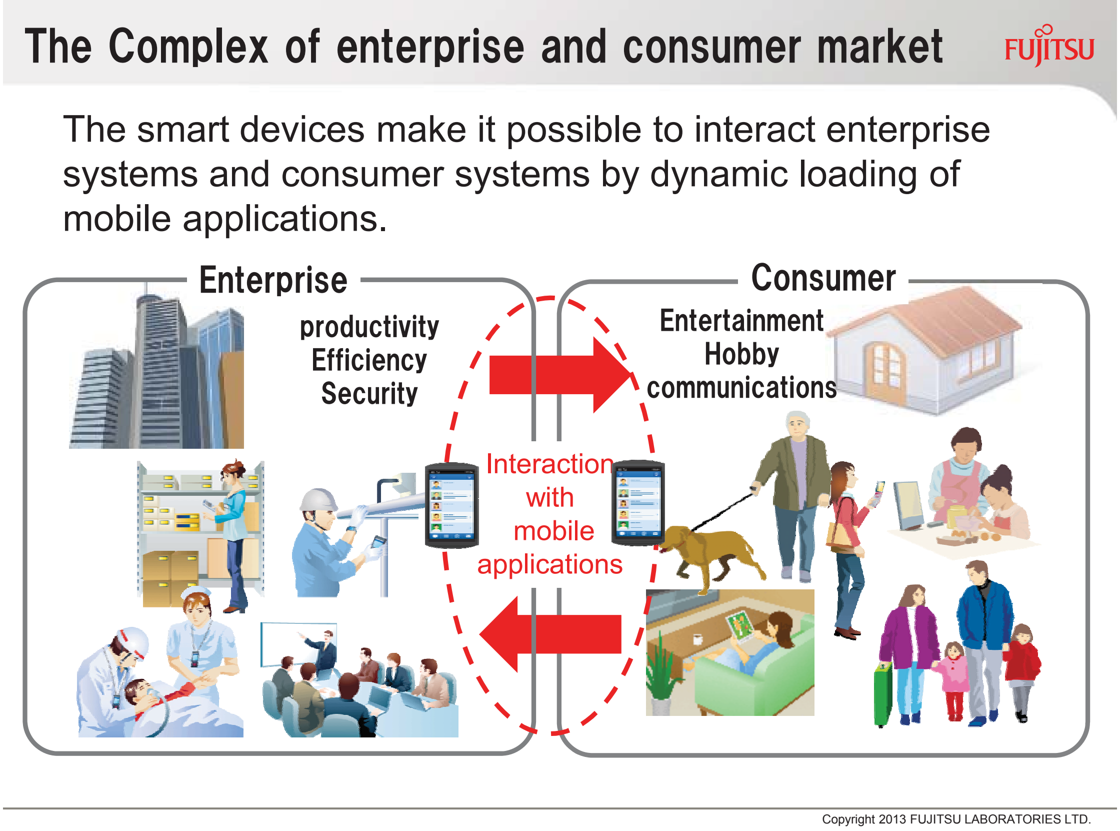
Gadget

Mobile Appl.

Device platform

Consumer devices (B2B2C)

7



The Complex of enterprise and consumer market

FUJITSU

The smart devices make it possible to interact enterprise systems and consumer systems by dynamic loading of mobile applications.

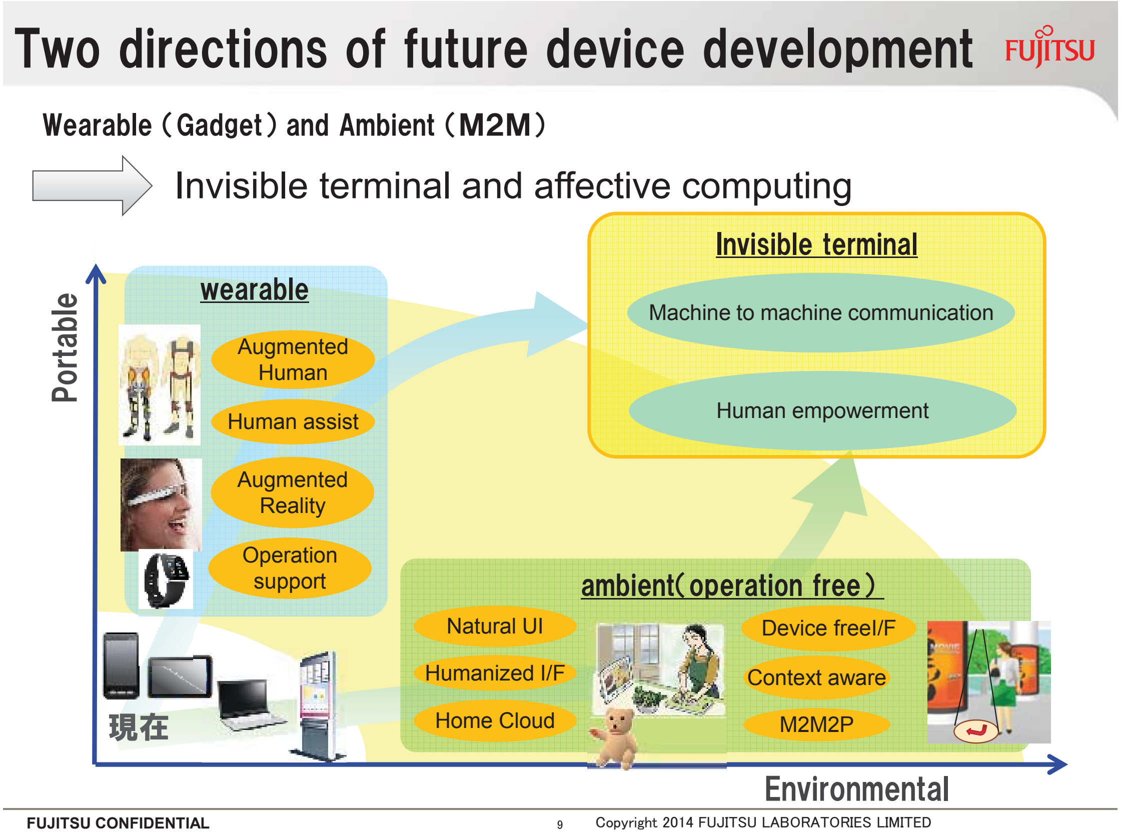
Enterprise

productivity
Efficiency
Security

Interaction with mobile applications

Consumer

Entertainment
Hobby
communications



Two directions of future device development

FUJITSU

Wearable（Gadget）and Ambient（M2M）

Invisible terminal and affective computing

Portable

wearable

Augmented Human

Human assist

Augmented Reality

Operation support

現在

Invisible terminal

Machine to machine communication

Human empowerment

ambient(operation free)

Natural UI

Humanized I/F

Home Cloud

Device freeI/F

Context aware

M2M2P

Environmental

FUJITSU CONFIDENTIAL

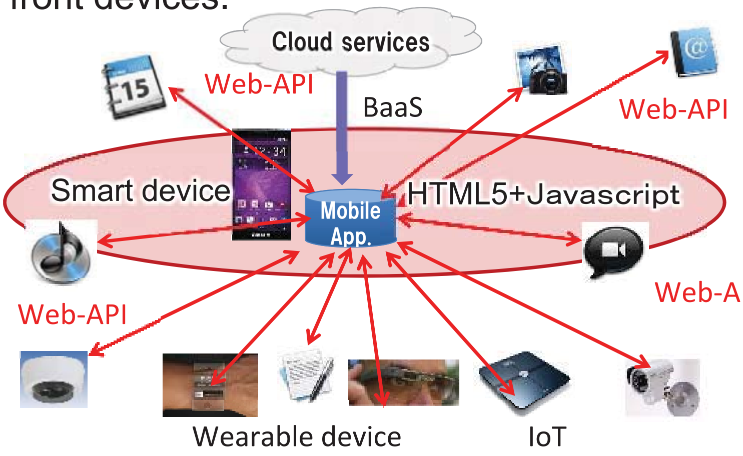
9

Real-virtual convergence

FUJITSU

Distributed Web-API network

- All network services and all devices can be accessed from mobile applications with Web-APIs
- Mobile application integrates various cloud services with various front devices.



10

From Web computing to Front computing

FUJITSU

- Human centric computing
- Impact of smart device on enterprise systems
- Web computing to front computing
- Systems become more and more dynamic
- Next target: Real-virtual integration
- New areas and new challenges in front area

Session 5: Networks, Applications and Web (Chair : Yoh Shiraishi)

A Selection Method of Optimal Channel in Wireless Network by the Dynamic Control of the Duty Cycle Threshold

Yoshitaka Nakamura[†], Yutaka Takahashi[‡], and Osamu Takahashi[†]

[†]Systems Information Sciences, Future University Hakodate, Japan
[‡]Graduate School of Systems Information Sciences, Future University Hakodate, Japan
{y-nakamr, osamu}@fun.ac.jp

Abstract - Recently, mobile terminals equipped with wireless LAN spread widely. Accordingly, access points such as public Wi-Fi spot or Pocket Wi-Fi come to be used widely. Most wireless LAN is communicated using 2.4GHz bandwidth, but it becomes the problem that the performance of the network decreases because plural terminals use the same channel. This paper intends that access point selects the channel which is most suitable for communication autonomously under the environment where there are a large number of wireless LAN terminals locally. Our proposed method predicts the congestion of the radio wave by analyzing Duty Cycle per each channel and detects the optimal channel. In addition, this method evaluates the quality of communication using expectation of SNR and demanded throughput from SNR. And our method reduces the cost of the channel selection by detecting optimal channel depending on the needs.

Keywords: Wireless LAN, IEEE802.11, FDMA, ISM Band, Duty Cycle, SNR, Cognitive Radio

1 INTRODUCTION

In recent years, many kinds of terminals such as personal computers, laptops, smart phones, tablet terminals, game consoles, and household appliances increasingly become equipped with 802.11 wireless LAN. Therefore Wi-Fi environment to use wireless LAN has been introduced into offices, home, and public accommodations. Furthermore, as measures to 3G line pressure by the spread of mobile phones such as smart phones, mobile phone carriers push the expansion of the public Wi-Fi spot. From these backgrounds, the demand of 2.4GHz bandwidth of radio wave spreads year by year. However, performance degradation by the radio wave interference becomes the problem when multiple wireless LAN apparatuses exist within certain areas.

Wireless LAN is equipped with the CSMA/CA which is a mechanism to avoid the radio wave interference. The more wireless LAN apparatuses exist in certain areas, the more frequently collision avoidance by CSMA/CA is carried out. Therefore, the transmission opportunity decreases, and problems such as the defectiveness of a throughput drop and the connection are caused. The detection of free channel is necessary to reduce this radio wave interference. And there are some existing wireless LAN access points with the free channel detection function. It may be said that the wireless LAN is the autonomous distributed wireless network. But in the autonomous distributed network, the case which all channels are using is assumed. In this case, if we can detect which channel is suitable for communication, we can stabilize and improve throughput of communication.

In this paper, we propose the method to predict the congestion degree and detect the optimal channel, by analyzing Duty Cycle which is the ratio of electric field strength beyond the threshold at the observation time for every channel. The proposed method controls the threshold to distinguish a radio signal from a noise autonomously depending on radio wave environment. However, at the time of the use by access points, the radio wave which the access points transmits and receives for sensing has an influence on the radio wave environment and may not obtain the most suitable result by the method using only Duty Cycle. Therefore, we evaluate the communication quality of access points using S/N ratio and the expectation of throughput calculated from S/N ratio, and detect the optimal channel using Duty Cycle if necessary.

2 RELATED WORK

2.1 Automatic Channel Selection of Wireless LAN Access Point

The item about the channel selection is not standardized in standards of IEEE802.11[1]. Therefore, it is entrusted implementation of each vendor under the present circumstances. In late years, almost all products have channel automatic setting function, and the products characterized by detecting free channel automatically also exist. However, the most of these functions are simple. Some method use a technique to detect the free channel around from channel information put on a beacon of the wireless LAN. This technique provides information such as Fig.1. From this information, an access point selects an unused free channel and uses the channel. However, this method may not select the optimal channel because there may not be free channel substantially, if the usage frequency bandwidth of wireless LAN is considered. Thus the free channel detection by beacon can perform relatively easily without the high-load complicated processing on access points. On the other hand, it becomes the problem that this technique cannot show maximum effect in the case that there are not enough free channels that channel interference does not occur.

2.2 Investigation into Utilization Status of Radio by the Site Survey

When the large-scale wireless LAN environment is built in the offices, the vender investigates the utilization status of

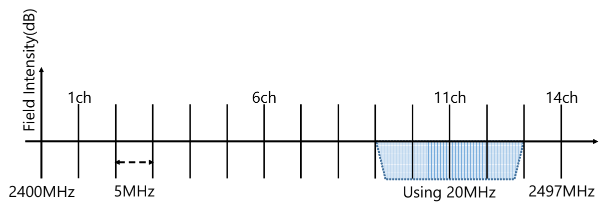


Figure 1: Channel division of the 2.4GHz bandwidth

neighboring radio beforehand to decide channel setting and the setting place of the access point may be used. This method called "site survey". There are some softwares such as Ref.[2] for the purpose of the site survey. By this software, we can analyze the signal intensity of the radio and can check the utilization status of radio for each area with a heat map form in detail. When we analyze the utilization status of channel using radio signal intensity by site survey, we can consider the radio signal of various standards except the wireless LAN in 2.4 GHz. In this way, we can find the optimal channel when all channels are used. However, the decision of the optimal channel depends on knowledge and the ability of the observer because it is entrusted to a human being performing site survey. In addition, it also becomes the problem that it is necessary to find the optimal channel for each time because the radio wave condition, the observation time, and the observation cost always changes after having decided the optimal channel once.

2.3 Effective Utilization of the White Space using DSA

There are some researches to utilize effectively of the radio resource by communicating with the channel which other radio systems do not use in 2.4GHz bandwidth temporarily and locally. These channels are called White Space. The technique to detect existing White Space discretely on a frequency axis, and to select and use the channel on White Space is called Dynamic Spectrum Access (following DSA). Reference[3] proposed the connection method, channel selection method, and data transmission method of the communication system using DSA. This research uses spectrum sensing for carrier sensing method to know the frequency utilization status. It is necessary to expand the carrier sense on not only the time axis, but also the frequency axis, because it is necessary to detect white space occurring discretely on a frequency axis in the DSA system instantly. Based on White Space provided in this way, the use efficiency of the radio resource rises with radio system changing frequency instantly, and using it. It is difficult to apply such DSA system to the wireless LAN of the IEEE802.11 standard that cannot instantly switch frequency in the specifications of the communication standard. Therefore the use with the software radio apparatus by the original communication standard is expected.

2.4 Learning Type Measuring of the Occupancy Rate

Reference[4] proposed the technique to select the optimal channel for radio transceivers establishing communication newly in multichannel autonomous distributed wireless network communication environment by paying its attention to the occupancy rate of the channel. This technique is effective when the occupancy rate of each channel is stable in time axis. However, the precision of this technique may decrease in the situation that the radio wave condition is easy to fluctuate and the situation that the occupancy rates of channels are distributed over uniformly. In addition, when we consider the application of this technique to existing communication standard including wireless LAN, access points and terminals also need expansion.

3 PROPOSED METHOD

In this paper, we propose the method to calculate the occupancy rate of each channel by calculating Duty Cycle and detect the optimal channel. Furthermore, I enable correspondence to the changeable radio wave environment by automatically controlling the threshold of Duty Cycle. And, we examine a method to apply the proposed method to the wireless LAN access point.

3.1 Duty Cycle

Duty Cycle expresses the ratio of electric field intensity beyond the threshold in the observation time by counting the appearance times of the spectrum beyond the threshold of the reception radio signal intensity set beforehand. It is thought to be able to predict the occupancy rate of each channel by calculating Duty Cycle of each channel. Duty Cycle(D) can be calculated as following formula from the number of the samples of a spectrum obtained in observation time(s) and the electric field intensity every chronological order($R[s]$).

$$D = \frac{\sum_0^s R[s]}{s} \qquad (1)$$

The proposed method detects the channel with the smallest D as the optimal channel. This method finds the optimal channel using the value that added up values of Duty Cycle of a leading channel and before and after channels in consideration of frequency bandwidth 20MHz, because the method is assumed an application to wireless LAN.

3.2 The Optimal Value of the Duty Cycle Threshold

3.2.1 The Threshold and the Detective Sensitivity

So as to make the threshold of Duty Cycle a large value, this method can detect the weaker electric field intensity. But, it is necessary to consider that it becomes easy to pick up a noise. On the contrary, the removal of the noise is enabled so as to make the threshold a small value, but the weak signal which is not a noise may be overlooked. In other words, the detective sensitivity of the signal and Duty Cycle threshold are in a trade-off relationship, and it is necessary to discover the optimal threshold. In addition, the optimal value of the Duty Cycle threshold varies according to the environment and the situation, because the signal strength changes by various factors such as the transmission output and the distance of the radio apparatus, and the obstacle.

3.2.2 The Optimal Threshold Selection Algorithm

The proposed method supports the changes of radio wave condition by controlling the optimal threshold of Duty Cycle dynamically depending on environment. Figure 2 shows the calculation algorithm of the optimal Duty Cycle threshold. This algorithm is carried out in the state that fixed quantities of measurement data are gathered. The standard deviation of Duty Cycle becomes higher so that variation of Duty Cycle is large in the threshold with every channel. Therefore the relative characteristic of each channel appears. We assume that it is in a condition to be able to separate a signal and the noise of the radio system when characteristics appear well. The proposed method adopts a Duty Cycle threshold with maximum standard deviation for the most suitable threshold.

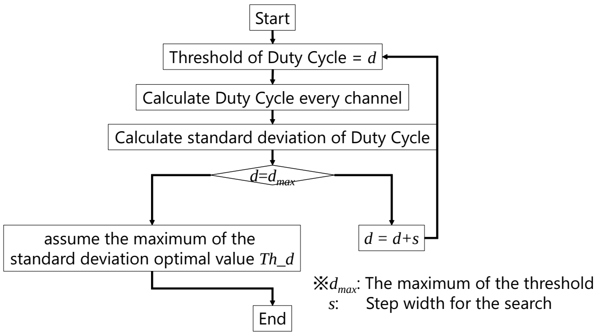

Start
Threshold of Duty Cycle = d
Calculate Duty Cycle every channel
Calculate standard deviation of Duty Cycle
d=d_{max}
assume the maximum of the standard deviation optimal value Th_d
d = d+s
End
※d_{max}: The maximum of the threshold
s: Step width for the search

Figure 2: Optimal value search algorithm of the Duty Cycle threshold

Figure 3 shows the relations of threshold and standard deviation of Duty Cycle on each channel when we measured in the 2.4GHz bandwidth that there are really plural radio systems. Duty Cycle is uniformly low in all channels in the range of the low threshold, and Duty Cycle of all channels becomes the maximum as far as the threshold is high. This shows the tradeoff relations of the threshold and the detective sensitivity. We can expect that the relations show similar tendencies in the different radio wave environment. Therefore, by setting the threshold based on standard deviation the measurement accuracy of the congestion degree by Duty Cycle of each channel rises and can detect the optimal channel even in a fluctuating radio wave environment.

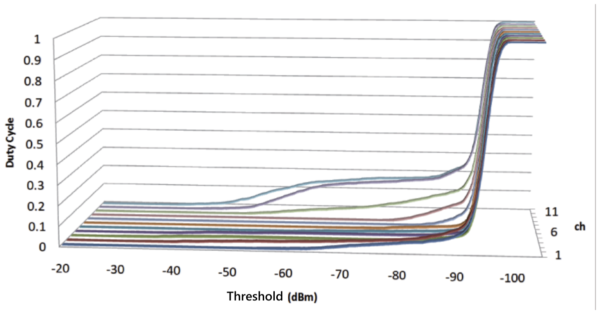


Figure 3: Duty Cycle of each channel in every threshold

3.2.3 The Optimal Channel Detection Method

Figure 4 shows the flow of the optimal channel detection method using the optimal threshold selection algorithm. At first this method performs sensing of the constant time and collects radio signal data. The method carries out the optimal value search algorithm of the Duty Cycle threshold. After obtaining the optimal value of the Duty Cycle threshold, the method calculates Duty Cycle of each channel. Afterward, in consideration of 20MHz that is a usable bandwidth of the wireless LAN, the method calculates the sum of Duty Cycle of a central channel and 2 neighbor channels, and assumes this value as the total value of Duty Cycles.

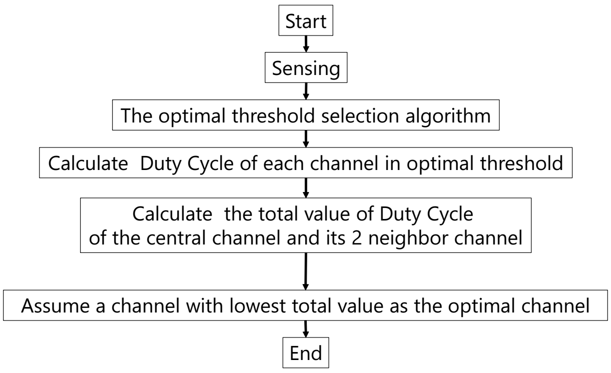

Start
Sensing
The optimal threshold selection algorithm
Calculate Duty Cycle of each channel in optimal threshold
Calculate the total value of Duty Cycle of the central channel and its 2 neighbor channel
Assume a channel with lowest total value as the optimal channel
End

Figure 4: Optimal channel detection method

3.3 Application to the Wireless LAN Access Points

3.3.1 Problems at the Time of the Application

There are some problems at the time of the proposed method application to the wireless LAN access point. At first, in the proposed method with the detection of the optimal channel by Duty Cycle, the own signal of the station is not considered. However, the own radio wave which the station transmits and receives is included in the signal data in the sensing when the method really detect the optimal channel and start communication. In addition, in order to calculate Duty Cycle, we need to observe signal data for some time and collect them. If observation time and frequency increase, the calculation cost also increases. Therefore, it is thought that the number of detections of optimal channel by Duty Cycle has to be suppressed to the minimum. From the above reasons, the following 2 points are the problems when we apply the proposed method to wireless LAN access points.

- The optimal channel detection in consideration of own signal of the base station
- Cost cutting of the optimal channel detection by Duty Cycle

3.3.2 SNR and Data Rate

SNR (Signal-to-Noise Ratio) expresses the ratio of the signal and the noise, and SNR is one index indicating the quality of wireless communication. If SNR is high, the influence of the noise is small. If SNR is low, the influence of the noise is large. When P_s represents signal power and P_n represents noise power, SNR is defined by following expressions.

$$\frac{S}{N} = \frac{P_s}{P_n} \tag{2}$$

If $P_s = -65(dBm)$, $P_n = -87(dBm)$ is provided, SNR can be calculated as follows.

$$\begin{aligned} \frac{S}{N} &= P_s - (P_n) \\ &= -65 - (-87) = 22 \end{aligned} \tag{3}$$

Each wireless LAN device defines necessary signal power for every data rate. Venders selling wireless LAN devices usually show the relations of receiving sensitivity and data rate. We can predict an approximate data rate under the measuring situation that can know the signal power because the data rate rises so that signal power is strong. However, only the signal power is not enough for the prediction of the data rate. This is because it may not satisfy the expected data rate if the noise power is strong even if signal power is strong enough. Therefore recommended SNR and lowest SNR which are required to satisfy the expected data rate are defined to consider the noise. Figure 5 shows each recommended SNR and each lowest SNR. The proposed method knows own quality of communication in the station by predicting data rate in reference to a value of recommended SNR in Fig. 5.

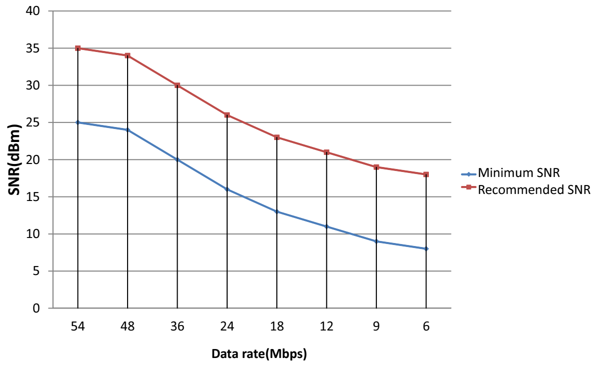


Figure 5: Relations of data rate and SNR

3.4 Channel Switching Algorithm

Figure 6 shows the flow of the channel switching algorithm to apply the proposed method to wireless LAN access points. At first, each base station measures the SNR for its own channel after constant waiting time passed. Then, the station calculates the data rate from provided SNR. Based on this provided data rate, the station judges whether effective throughput is provided in comparison with the real throughput. However, it is necessary to calculate the maximum data rate in the real environment because the data rate shown in wireless LAN is generally a theoretical value.

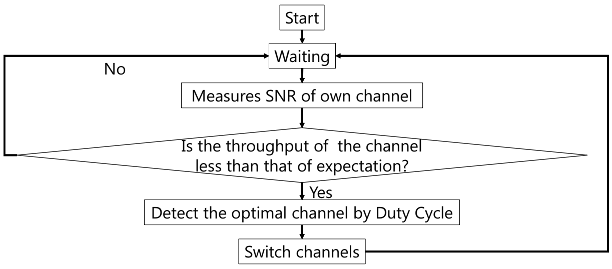

Start
Waiting
No
Measures SNR of own channel
Is the throughput of the channel less than that of expectation?
Yes
Detect the optimal channel by Duty Cycle
Switch channels

Figure 6: Channel switching algorithm

In the real environment, the access control by CSMA/CA acts firstly, and the waiting time of the ACK reply occurs every one data transmission. Reference [5] calculates effective speed of IEEE 802.11g in the real environment as follows. Reference [5] assumes the fixed waiting time as $35\mu s$, random waiting time as $67.5\mu s$ at the time of the packet transmission. The fixed waiting time of the ACK reply is also assumed as $16\mu s$. Under this condition, if the data packets of $1460byte$ are sent two times at data rate $54Mbps$, and the ACK reply is received one time, the effective speed is calculated by the following expressions.

$$\frac{146(byte) \times 2(packet) \times 8(bit)}{0.0009645(sec)} = 24.2(Mbps) \tag{4}$$

The effective speed becomes $24.2Mbps$ for theoretical value $54Mbps$. Therefore, the realistic effective speed becomes around 45% of the theoretical value. As the terminals using the access point, and other radio devices exisiting in the channel interference zone incerease, the struggles of access privileges by CSMA/CA are frequent, and the effective speed declined further. In this paper, we calculate expectation of the throughput by the following expressions.

$$\begin{aligned} &\text{Expected throughput}(Mbps) \\ &= \frac{\text{Data rate corresponding to SNR}(Mbps) \times 45\%}{\text{\# of the own terminals of the base station}} \end{aligned} \tag{5}$$

The access point compares the measured throughput with the expected throughput. If the measured throughput is more than of expected throughput, it enters the waiting state and measures SNR again. Afterwards it compares the measured throughput with the expected throughput. If the measured throughput is less than the expected throughput, it detects the optimal channel by Duty Cycle, switches channels and enters the waiting state. After that loops do this cycle until finishing the use of the access point. In this way, the proposed method reduces the processing cost of channel detection by repressing the number of optimal channel detection to a minimum by Duty Cycle while considering own signals of stations.

4 PERFORMANCE EVALUATION

4.1 Optimal Channel Detection by the Automatic Control of Duty Cycle threshold

We evaluated following 3 items of evaluation, a throughput evaluation with the optimal channel and other channels, an effective evaluation of automatic threshold control using the standard deviations, and a necessary sensing time evaluation before detecting the correct optimal channel. In addition, the target channels are assumed from channel 1 to channel 11. This is because the channel which wireless LAN can set as central channel is usually assumed. By the calculation processes of the proposed method, the frequency corresponds to channel -1, channel 0, channel 12 channel 13 are included internally. And we used PCATTCP[6] made in Printing Communications Associates for the measurement of the throughput. The parameters of experiments are shown in Table 1.

Table 1: Parameters of experiments

Observation time	$60sec$
Target frequency bandwidth	$2.4GHz$ bandwidth
Target channels	1 - 11
Communication standard	IEEE 802.11b/g
Communication capacity	$16Mbyte$
Number of trials to measure throughput	3
The channels which other devices are using	1,4,6,11

From the results of experiments, the optimal value of the threshold is $-93.5dBm$ and Duty Cycle of each channel is shown in Fig. 7, the throughput of each channel is shown in Fig. 8. The optimal channel of the proposed method is 5 from Fig. 7, and the highest throughput is accomplished in channel 5 from Fig. 8. The best channel of measured throughput accorded with the optimal channel detected by the proposed method and we can show the effectiveness of the optimal channel detection by the proposed method.

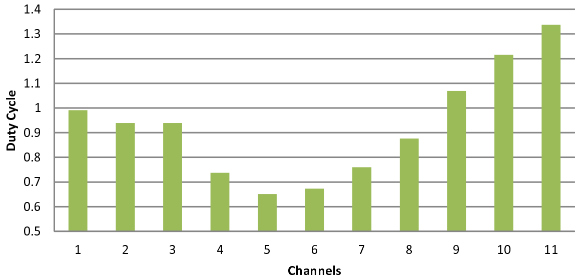


Figure 7: Duty Cycle of each channel at optimal threshold

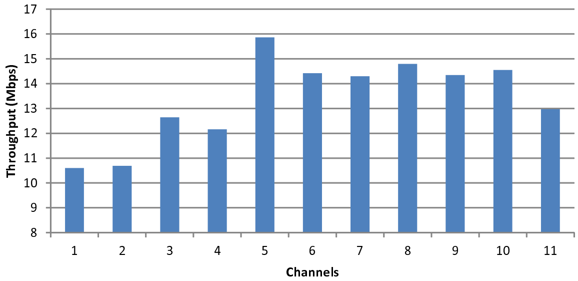


Figure 8: Throughput of each channel

Figure 9 shows the variation of the optimal channel every threshold using experimental data. The point where multiple optimal channels are plotted with the same threshold shows that the sum of Duty Cycle is the same, and the judgment of the optimal channel becomes impossible.

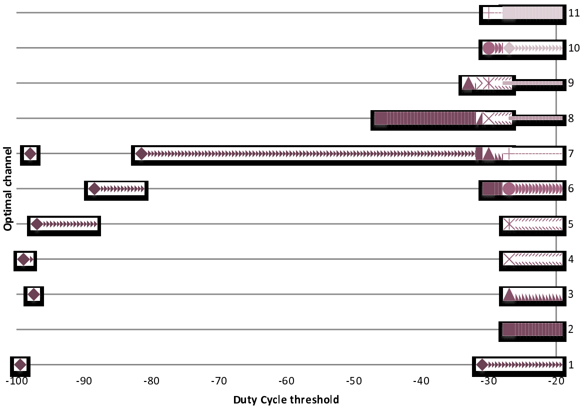


Figure 9: Variation of optimal channel at each threshold

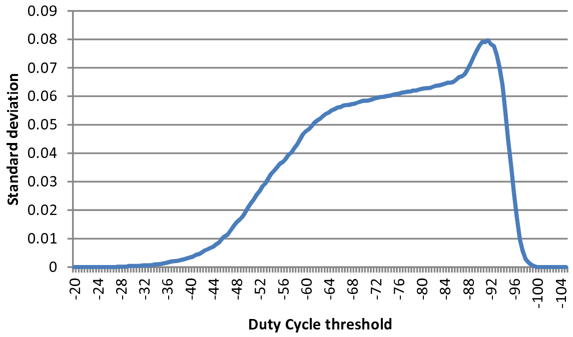


Figure 10: Variation of the standard deviation at each threshold

The signals were not detected in the threshold range lower than $-40dBm$. The proposed method can detect the optimal channel with 5, when the range of the threshold is from $-89dBm$ to $-97dBm$. Therefore, the range of threshold that the optical channel can be detected exists within the uniformity in succession, though there is the change to some extent by the environment.

From Fig. 10, the range where the proposed method can detect the optimal channel is equivalent to the part that standard deviation forms a peak. In other words, corresponding relationship is found between the standard deviation of Duty Cycle and the range of the threshold where the proposed method can detect the optimal channel.

4.2 Application to the Wireless LAN Access Points

Then, we evaluated the proposed method about the application to the wireless LAN access points by the simulations. We compared the average throughput in the simulation time, in a case to apply the proposed method to wireless LAN access point and in a case to detect the optimal channel by Duty Cycle only in the first time before the communication.

Firstly, we evaluated the case that one terminal connects to the access point with 60 minutes simulation under the parameter of Table 2. In this experiment, we used a traffic model to move channel to 2, 4, 6, 8, and 10 every ten minutes. In addition, we used a radio wave environment model based on data of the real environment observed by the experiment of the Sec. 4.1.

Table 2: Simulation parameters

Simulation time	$60min$
Waiting time	$1sec$
Simulation waiting time	$100msec$
Own Duty Cyce	80%
Own signal power	$-40dBm$
# of own terminals	1
Traffic generation channel	$2 \rightarrow 4 \rightarrow 6 \rightarrow 8 \rightarrow 10$
Traffic generation interval	$10min$
Duty Cycle of traffic	30%
Signal power of traffic	$-50dBm$

From the result of an experiment, Fig. 11 shows the variations of the channel by the channel switching algorithm. Figure 12 shows the moment throughput at the time of applying the channel switching algorithm in each time, and Fig. 13 shows the moment throughput at the time of the algorithmic non-application. In addition, Fig. 14 shows the average throughput in the case of applying the channel switching algorithm and in the case of not applying the algorithm.

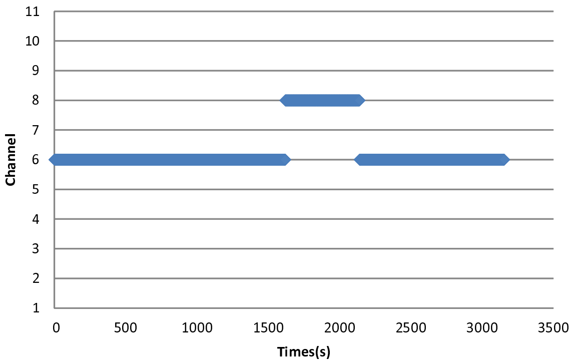


Figure 11: Variation of the channel

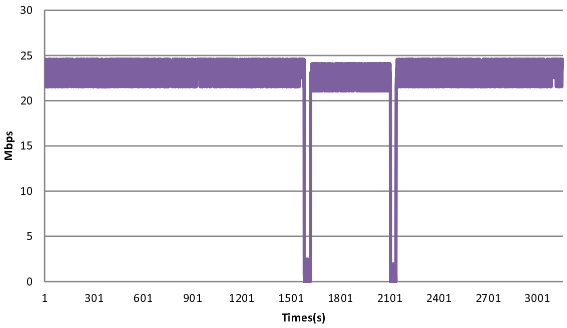


Figure 12: Moment throughput with channel switching

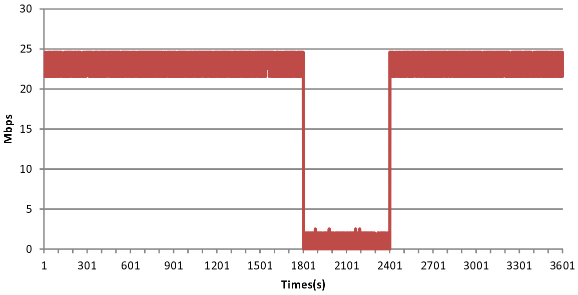


Figure 13: Moment throughput without channel switching

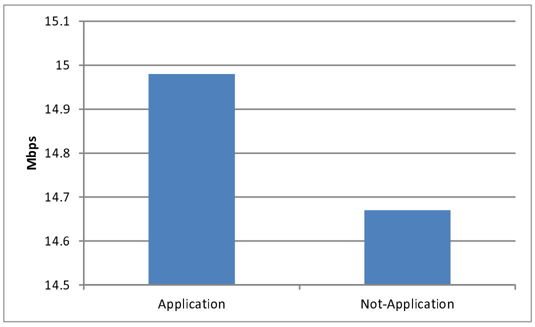


Figure 14: Average throughput

Channel 6 is selected by the first optimal channel detection in both cases, and the change of the channel occurred around $1800sec$ when additional traffic generated on channel 6. At this time, the moment throughput decreases by time lag required for channel switching and by the interference of the traffic increase. And the throughput also decreases sharply at the time when traffic increased on channel 6 in the case of not applying the channel switching algorithm. The average throughput of applying the channel switching algorithm is $0.32Mbps$ more than the throughput of not applying the algorithm.

5 CONCLUSIONS

In this paper, we intended to detect the optimal channel when a wireless LAN system communicated in the radio environment such as 2.4GHz bandwidth. We proposed the detecting method of the optimal channel by measuring the congestion degree of every channel by our performing spectrum sensing and calculating of Duty Cycle. We compared the optimal channel detected by the proposed method with the actual value of the throughput, and showed the effectiveness of the proposed method by the experiments in the real environment. From the relations of the carrier sense time and the detected optimal channel, it is found that the optimal channel might be stable to constant value, when it exceeds the fixed periods of time.

In addition, we applied the proposed method to wireless LAN access point and examined the method to detect and switch the optimal channel when a own channel of the station caught the interference while communicating. We compared the throughput at the time of the non-application of the channel switching algorithm with throughput at the time of the application of the algorithm by simulation under multiple conditions. In each case, the throughput at the time of the application of the algorithm that shows higher values and we can confirm the effectiveness of the proposed method.

For future work, we have to inspect the action in the environment with multiple wireless LAN access points applying proposed method, or in the inferior environment that traffic occurs with multiple channels complicatedly. And it will be necessary to work effectively in such an environment.

REFERENCES

[1] IEEE: "Part11: Wireless lan medium access control (mac) and physical layer (phy) specifications" (2007).

[2] "Ekahau". `http://www.dit.co.jp/products/ekahau_ss/`.

[3] N. Nakamoto, K. Yano, Y. Suzuki, S. Aikawa, M. Uno and M. Ueba: "Impact of spectrum sensing strategy for dsa system on wlan and bluetooth", IEICE Technical Report, Vol. 110, pp. 69–76 (2011).

[4] O. Takyu, T. Kisi, T. Fujii, Y. Umeda and K. Kinoshita: "High speed rendezvous channel based on recursive update measurement method for channel occupancy ratio", IEICE Technical Report, Vol. 111, pp. 19–24 (2011).

[5] "Technology scope: High-speed wireless lan", NIKKEI COMMUNICATIONS, Vol. 2002/07/15, pp. 126–133 (2002).

[6] PCAUSA: "Test tcp utility". `http://www.pcausa.com/Utilities/pcattcp.htm`.

Adopted Transfer Learning to Item Purchase Prediction on Web marketing

Noriko Takaya[†], Yusuke Kumagae[†], Yusuke Ichikawa[†], and Hiroshi Sawada[†]

[†]NTT Service Evolution Laboratories, NTT Corporation, Japan
{takaya.noriko, kumagae.yusuke, ichikawa.yusuke, sawada.hiroshi}@lab.ntt.co.jp

Abstract - The transfer learning method will be modified more effectively for the item purchase prediction on web marketing. Acquiring a various related site information, it would give more accurate prediction than a single site analysis. These multiple EC sites have two problems that 1) some item purchase data are inconsistent to another data set and indeed lower the prediction accuracy and 2) the item's information of brands, categories, prices, and item names in multiple EC sites are sparse and imbalance. Analyzed these characteristics, we propose an ensemble-based approach that effectively aggregates weak classifiers by efficiently avoiding the negative learning effect. Furthermore, we convert the item information to an abstract form. These methods are validated by the actual purchase logs over several Japanese fashion EC sites.

Keywords: Transfer learning, Marketing, Machine learning, EC

1 INTRODUCTION

Administrator of an EC site has been constructing a model to predict their customers purchase. Knowing the information about user's purchase behaviors in other EC sites, they would get more precise insight. Fig. 1[1] shows that the information of the customer behaviors on multiple site will help more precise analysis of the prediction model. The *Transfer Learning* method is known as one of the effective approach for analyzing these transfer situation. In this method, the target domain for which predictions are to be made is called *Target* and the different domain used for learning is called *Source*. The purpose of transfer learning is to utilize the knowledge acquired from the source to improve prediction performance in the target domain. The negative transfer is pointed as a known problem for adopting this method[1]. Fig. 2 describes this negative transfer where the attributions of target data are different from that of source data. In this case, using the source data in learning phase degrades the accuracy. Rosenstein showed a specific example[1]. While their goal was predicting whether the target person would attend or not a specific meeting, the training data were drawn from two people with different attributions (academic and military). Using this training data decreased prediction accuracy. Naturally, there are target/source pairings which improve or degrade accuracy. Therefore, in transfer learning, it is a problem that how we avoid negative transfer effect and how we find similar data. Our proposal is a purchase-based model to predict item sales. This *OptTrBagg* (Opt Transfer Bagging) model is more tolerant to negative transfer. The algorithm is based primarily on *TrBagg*, which is an extension of bagging method, and efficiently drops inadequate base classifiers in aggregation phase.

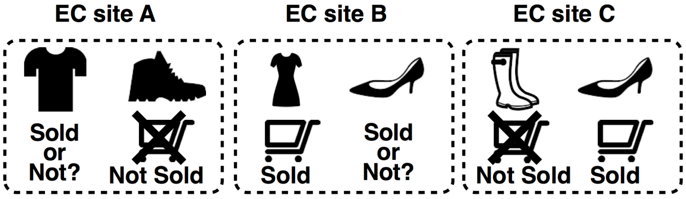


Figure 1: An illustration of our research settings. Traditional item purchase behavior research focuses on single EC site's information (in the above figure, EC site A). Our proposal collects purchase of multiple EC sites (in the above figure, EC sites A, B, and C) and constructs a model to predict whether the item would sell or not.

The inconsistency of each brand or category attribution on the multiple site would cause a identity difficulty for the model. Adopting the abstract description will give the answer to avoid this problem. The effectiveness of our approach is validated by experimentation actual purchase data.

In Section 2, we explain related work on modeling for purchase behavior, ensemble learning, and transfer learning. In Section 3, we introduce TrBagg as the baseline, and propose OptTrBagg, our approach. In Section 4, we explain construction of features across multiple EC sites. In Section 5, we explain our actual purchase information datasets and show the results of experiments. Moreover, in this section, we explain how transfer learning changed the prediction models. Finally, in Section 6, we summarize this paper.

2 RELATED WORK

2.1 Modeling Purchase Behavior

In the area of modeling purchase behavior in e-commerce, there are two approaches, item based prediction and session based prediction. Item based prediction construct models that use the item's own information such as price, category, and item name to predict if the item would be sold. On the other hand, session based prediction construct models that process the user's activity information such as how long the user peruses an item, how many times the user clicked a link, and what queries the user input to predict whether the user will purchase the item in the current session. In item based prediction, Wu and Bolivar discussed the problem of prediction of item purchase behavior[2]. Within eBay[2], which is the

[1] All pictograms used in this paper are from The Noun Project (http://thenounproject.com/). Boots designed by Luis Prado.

[2] http://www.ebay.com/

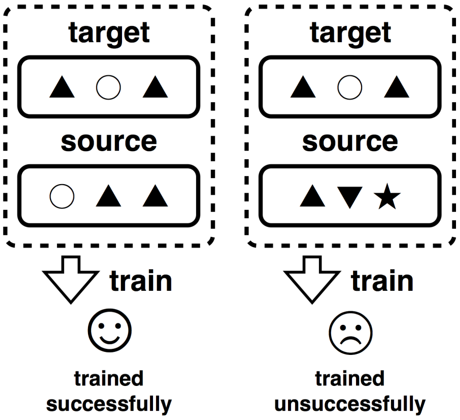


Figure 2: An example of negative transfer. In the left image, the target and source data have similar distribution and model training is successful. In contrast, the right image shows different distributions triggering negative transfer and the failure of model training.

largest Internet auction site, they assigned features to items posted on eBay and predicted the result of purchase by logistic regression. Our research resembles theirs in that it assigns features to each item and prediction of the result, but differs in that is uses information from several EC sites. They also discussed item based prediction but with the goal of predicting item rarity[3]. For session based prediction, Kim uses neural networks that have different hidden layers and aggregate their classification results to predict item purchase behavior in an EC site[4]. Our research resembles this work in that it uses ensemble learning, but differs in that it uses transfer learning. Moe and Fader assign features, such as page transitions and split time period, to the user session on an EC site and predict whether the user will purchase any items in this session or not[5]. Poel and Buckinx also discussed this problem[6]. Guo and Agichtein predict whether the user is now trying to purchase an item or just browsing[7]. They used a Markov chain model and compared the transition probabilities between user purchase and other. In addition, there is similar research in the area of display advertising in websites[8]–[10]. These works used user behavior information extracted from click-through logs as features, whereas we use user behavior-independent information.

Limayem analyzed user purchase behavior on the Internet using factor models[11]. The models they use are based on hypothesis and statistical test between two factors, such as there being a positive relationship between Personal Innovativeness and Intention. Bellman investigated lifestyle and purchase behavior of the user who was called Wired those days[12].

2.2 Ensemble Learning and Transfer Learning

The concept of ensemble learning is to generate several weak learners to reduce variance and improve accuracy. Bagging[13] generates several base classifiers (decision tree is used as base classifier in the original paper) and aggregates their classification results by simply majority voting (in regression problems, values of each weak learner are averaged). Freund proposed AdaBoost[14], which does not aggregate base classifier's results naively. AdaBoost weights each base classifier by empirical error and final prediction is yielded by weighted voting.

Transfer learning is widely used in link prediction[15], displaying advertise[16], object detection in image processing[17], regression[18], video summarization[19], text classification[20]. Kamishima proposed TrBagg[21], which applies bagging to transfer learning (see in Section 3.1). Dai proposed TrAdaBoost which applies AdaBoost to transfer learning[22]. Rosenstein proposed ExpBoost which also applies AdaBoost to transfer learning[23]. Pararoe expanded TrAdaBoost and ExpBoost to cover regression problem[18]. Given that TrBagg offers ease of implementation and tuning, the possibility of parallel computation, and superior accuracy, we propose create our algorithm. Daume proposed a transfer learning method that converts both target and source features simply[24]. For example, the F dimension feature vector $\mathbf{x} \in \mathbb{R}^F$ in target domain $\mathcal{D}_T$ is converted to new feature vector $\Phi^T(\mathbf{x}) =< \mathbf{x}, \mathbf{0}, \mathbf{x} >\in \mathbb{R}^{3F}$, where $\mathbf{0}$ is empty vector $< 0, \cdots, 0 > \in \mathbb{R}^F$. F dimension feature vector $\mathbf{x} \in \mathbb{R}^F$ in source domain $\mathcal{D}_S$ is also converted to new feature vector $\Phi^S(\mathbf{x}) =< \mathbf{x}, \mathbf{x}, \mathbf{0} >$. The converted vectors are used to train a model. We adopt this approach as our baseline in the experiments.

3 ENSEMBLE TRANSFER LEARNING

First, we explain TrBagg as baseline method, and next we propose OptTrBagg algorithm.

3.1 Baseline Method: TrBagg

TrBagg is the extension of bagging, which is proposed by Kamishima[21](Algorithm 1). The inputs are target data $\mathcal{D}_T$, source data $\mathcal{D}_S$, and the number of initial base classifiers N. In training, the output is a set of base classifiers $\mathcal{F}^* = \{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$. The number of output is n and is not greater than the number of initial base classifiers N. The algorithm is as follows.

First, we generate merged data sets $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_S$, the union set of target and source data. We get classifier $\hat{f}_0$ learned from $\mathcal{D}_T$ in step 4. In the iteration of t, we generate training data $\mathcal{D}'_t$ by bootstrap sampling (random sampling which allows duplication) from $\mathcal{D}$ and get base classifier $\hat{f}_t$ learned from $\mathcal{D}'_t$. By repeating this iteration N times, we get the set of base classifiers $\mathcal{F} = \{\hat{f}_0, \hat{f}_1, \cdots, \hat{f}_N\}$.

Next, we filter the base classifiers $\mathcal{F}$ by Algorithm 2. In step 3, we sort $\mathcal{F}$ in ascending order of their empirical errors on the target set $\mathcal{D}_T$. From step 7, we check each base classifier f_t according to the empirical error. That is, we check whether the empirical error of majority voting is improved or not by the addition of f_t on $\mathcal{D}_T$ to the set of base classifiers $\mathcal{F}'$. The result of prediction $\hat{c}$ by majority voting on unknown data $\mathbf{x}$ by the set of models $\mathcal{F}'$ is determined by

$$\hat{c} = \arg\max_{c \in \mathcal{C}} \sum_{\hat{f}_t \in \mathcal{F}'} \mathrm{I}[c = \hat{f}_t(\mathbf{x})], \tag{1}$$

Algorithm 1 TrBagg

```
function Training
INPUT $\mathcal{D}_T$, $\mathcal{D}_S$, $N$
$\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_S$
$\mathcal{F} = \{\hat{f}_0\}$; $\hat{f}_0$: learned from $\mathcal{D}_T$
for all $t = 1$ to $N$ do
    $\mathcal{D}'_t \leftarrow$ generated by bootstrap from $\mathcal{D}$
    $\hat{f}_t$: learned from $\mathcal{D}'_t$
    $\mathcal{F} = \mathcal{F} \cup \hat{f}_t$
end for
$\mathcal{F}^* = \mathbf{Filtering}(\mathcal{F}, \mathcal{D}_\mathbf{T})$
OUTPUT $\mathcal{F}^* : \{\hat{f}_0, \hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$ $(n \leq N)$
```

Algorithm 2 Filtering

```
function Filtering
INPUT $\mathcal{F} : \{\hat{f}_0, \hat{f}_1, \hat{f}_2, \cdots, \hat{f}_N\}$, $\mathcal{D}_T$
$< f_0, f_1, \cdots, f_N >$: sort $\mathcal{F}$ by empirical error on $\mathcal{D}_T$ in ascending
$\mathcal{F}' = \{f_0\}$
$\mathcal{F}^* = \{f_0\}$
$e \leftarrow$ empirical error of $\hat{f}_0$ on $\mathcal{D}_T$
for all $t = 1$ to $N$ do
    $\mathcal{F}' = \mathcal{F}' \cup f_t$
    $e' \leftarrow$ empirical error of majority voting $\mathcal{F}'$ on $\mathcal{D}_T$
    if $e' \leq e$ then
        $\mathcal{F}^* = \mathcal{F}^* \cup \mathcal{F}'$
        $e' = e$
    end if
end for
OUTPUT $\mathcal{F}^* : \{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$ $(n \leq N)$
```

Algorithm 3 OptTrBagg

```
 1: function Training
 2: INPUT $\mathcal{D}_T$, $\mathcal{D}_S$, $N$
 3: $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_S$
 4: $\mathcal{F} = \{\hat{f}_0\}$; $\hat{f}_0$: learned from $\mathcal{D}_T$
 5: for all $t = 1$ to $N$ do
 6:     $\mathcal{D}'_t \leftarrow$ generated bootstrap from $\mathcal{D}$
 7:     $\mathcal{D}'_{t,T} = \mathcal{D}'_t \cap \mathcal{D}_T$
 8:     $\hat{f}_{t,T+S}$: learned from $\mathcal{D}'_t$
 9:     $e_{T+S} \leftarrow$ empirical error of $\hat{f}_{t,T+S}$ on $\mathcal{D}_T$
10:     $\hat{f}_{t,T}$: learned from $\mathcal{D}'_{t,T}$
11:     $e_T \leftarrow$ empirical error of $\hat{f}_{t,T}$ on $\mathcal{D}_T$
12:     if $e_T \leq e_{T+S}$ then
13:         $\hat{f}_t = \hat{f}_{t,T}$
14:     else
15:         $\hat{f}_t = \hat{f}_{t,T+S}$
16:     end if
17:     $\mathcal{F} = \mathcal{F} \cup \hat{f}_t$
18: end for
19: $\mathcal{F}^* = \mathbf{Filtering}(\mathcal{F}, \mathcal{D}_\mathbf{T})$
20: OUTPUT $\mathcal{F}^* : \{\hat{f}_0, \hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$ $(n \leq N)$
```

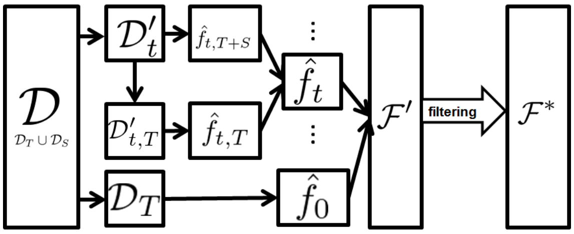


Figure 3: An overview of OptTrBagg. $\mathcal{D}'_t$ is generated by bootstrap sampling from both target and source data $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_S$. $\mathcal{D}'_{t,T}$ is extracted from $\mathcal{D}'_t$ by intersection of target data $\mathcal{D}_T$. Classifier $\hat{f}_t$ is selected from $\hat{f}_{t,T+S}$ learned from $\mathcal{D}'_t$ and $\hat{f}_{t,T}$ learned from $\mathcal{D}'_{t,T}$ by its empirical error.

where I[$cond$] is an indicator function that returns 1 if condition $cond$ is true, and $\mathcal{C}$ is the set of classes. If empirical error is improved, we set all $\mathcal{F}'$ to $\mathcal{F}^*$.

Finally, we get the set of base classifiers $\mathcal{F}^*$. The aim of this filtering is to prevent negative transfer, since transfer learning is not always assured to be effective. That is to say, in filtering iteration i, using just the target data may yield higher performance. We propose a method that is more effective in avoiding negative transfer.

3.2 Proposed Method: OptTrBagg

The method is based on the idea of not using source data that degrade prediction accuracy; OptTrBagg overcome this problem by filtering out the base classifiers. Algorithm 3 and Fig. 3 describe the procedure. The difference between OptTrBagg and TrBagg is the learning process involving base classifier $\hat{f}_t$. In iteration t, our approach pays attention to target data $\mathcal{D}'_{t,T}$ which are contained in training data $\mathcal{D}'_T$ (in step 6 and 7). We get the base classifier $\hat{f}_{t,T+S}$ learned from $\mathcal{D}'_t$ and another base classifier $\hat{f}_{t,T}$ learned from $\mathcal{D}'_{t,T}$ (in step 8 and 10). Incidentally $\hat{f}_{t,T+S}$ is denoted as $\hat{f}_t$ in Algorithm 1. Using empirical error on $\mathcal{D}_T$ to comparing $\hat{f}_{t,T+S}$ with $\hat{f}_{t,T}$, we use the model as $\hat{f}_t$ in iteration t (in step 17).

The base classifier $\hat{f}_{t,T+S}$ is learned from both the target and source data because $\mathcal{D}'_t$ contains target and source data. If learning process with source data is effective, the empirical error of $\hat{f}_{t,T+S}$ which is learned from target and source data may be smaller than the empirical error of $\hat{f}_{t,T}$ which is learned from target data. In this case, $\hat{f}_{t,T+S}$ is adopted as $\hat{f}_t$ and this result is equal to TrBagg's process. On the contrary, if learning by source data fails, the empirical error of $\hat{f}_{t,T+S}$ may be larger than empirical error of $\hat{f}_{t,T}$. In this case, OptTrBagg adopt $\hat{f}_{t,T}$ as $\hat{f}_t$.

In eliminating negative transfer, both TrBagg and OptTrBagg filter base classifiers by majority voting. In addition to this filtering, OptTrBagg checks whether each base classifier degrades accuracy or not. Hence, OptTrBagg more efficiently the negative transfer classifiers that can slip into aggregation of base classifiers. That is to say, OptTrBagg can be interpreted as extension of TrBagg where source data that degrade accuracy are removed.

3.3 Difference Between OptTrBagg And TrBagg Viewing From bagging

We explained OptTrBagg as an extension of the TrBagg algorithm in the previous section, but OptTrbagg can also be interpreted as an extension of bagging[13]. In iteration t, bagging generates training data $\mathcal{D}'_t$, and base classifier $\hat{f}_t$ learned from $\mathcal{D}'_t$. Finally, bagging gets the set of base classi-

Table 1: Definition of symbols used in constructing features.

Symbol	Definition
item_i	A fashion item.
r_i	Price of item_i.
c_i	Category of item_i (e.g. Polo shirt, Denim jeans, and Scarf).
b_i	Brand of item_i (e.g. LOUIS VUITTON[3] and Burberry[4]).
t_i	Item name of item_i (e.g. *"Oxford Button-Down Shirt"*).
s_i	Boolean indicating whether item_i was sold (= 1) or not (= 0).
$\mathcal{I}$	Set of all items.
$\mathcal{I}_{c,i}$	Set of items whose category is c_i, $\{\forall \text{item}_j \in \mathcal{I}, c_j = c_i\}$.
$\mathcal{I}_{b,i}$	Set of items whose brand is b_i, $\{\forall \text{item}_j \in \mathcal{I}, b_j = b_i\}$.
$\mathcal{I}_s$	Set of items sold, $\{\forall \text{item}_i \in \mathcal{I}, s_i = 1\}$.
$\mathcal{I}_{K,i}$	Top K items that have similar item name to item_i.
$\lvert\mathcal{I}_*\rvert$	Size of set of items $\mathcal{I}_*$ (* indicates some conditions).
$\langle r_i \rangle_{i \in \mathcal{I}_*}$	Average price of a set of items $\mathcal{I}_*$ (* indicates some conditions).

fiers $\mathcal{F}^* = \{\hat{f}_1, \cdots, \hat{f}_N\}$.

Both algorithms, bagging and OptTrBagg, generate training data $\mathcal{D}'_{t,T}$ from $\mathcal{D}_T$ by bootstrap and get base classifier $\hat{f}_{t,T}$. Although bagging uses $\hat{f}_{t,T}$ as $\hat{f}_t$, OptTrBagg decides $\hat{f}_t$ by comparing $\hat{f}_{t,T}$ with $\hat{f}_{t,T+S}$, which is learned from the data bootstrapped from source data $\mathcal{D}_S$ and $\mathcal{D}_T$. As explained in Section 3.2, if source data are effective in the learning phase, $\hat{f}_{t,T+S}$ is adopted as $\hat{f}_t$. On the contrary, if source data may cause negative transfer, $\hat{f}_{t,T}$ is adopted as $\hat{f}_t$ and this result is equivalent to that of bagging.

That is to say, OptTrBagg is interpreted as extension of bagging to transfer learning where source data that can improves accuracy is used.

4 FEATURES ACROSS MULTIPLE EC SITES

Constructed a model to predict the selling, eight features were proposed based on the purchase prediction model of Wu and Bolivar[2]. There are two kinds of features, six attribution are based on item attributes of item's price, category and brand, and two name based features were constructed by item name. The symbols used in explaining features are defined in Table 1.

4.1 Attribution Based Features

Attribution based features are constructed from the information about price of sold items. The detail is as follows;

- We cannot use the price of item_i, r_i as a feature directly because it differs by brand and/or category of item_i. Comparing the prices of different categories, such as underwear and suit priced, is nonsense. It is important to consider the item price as the different from an average price.
- The popularity of a brand or category on each site differ. Therefore, the direct comparison of these attributions is not fair. These information will be transform the abstract form.

The proposal methods are;

- **Category Averaged Price** : $r_i - \langle r_i \rangle_{i \in \mathcal{I}_{c,i}}$, difference between price of item_i, r_i and average price of category c_i.
- **Category Averaged Sold Price** : $r_i - \langle r_i \rangle_{i \in (\mathcal{I}_{c,i} \cap \mathcal{I}_s)}$, difference between price of item_i, r_i and average price of sold items in category c_i.
- **Category Hotness** : $\frac{|\mathcal{I}_{c,i} \cap \mathcal{I}_s|}{|\mathcal{I}_{c,i}|}$, the selling rate of items items whose category is c_i.
- **Brand Averaged Price** : $r_i - \langle r_i \rangle_{i \in \mathcal{I}_{b,i}}$, difference between price of item_i, r_i and average price of brand items b_i.
- **Brand Averaged Sold Price** : $r_i - \langle r_i \rangle_{i \in (\mathcal{I}_{b,i} \cap \mathcal{I}_s)}$, difference between price of item_i, r_i and average price of sold items whose brand is b_i.
- **Brand Hotness** : $\frac{|\mathcal{I}_{b,i} \cap \mathcal{I}_s|}{|\mathcal{I}_{b,i}|}$, the selling rate of items whose brand is b_i.

4.2 Name Based Features

The name based features is follows;

- The price, category, and brand information annotated in item captures item purchase tendency, it does not contain other information such as color, shape, and feel. For example, for the item named "Cute Mori-Girl[5] style! Over knee high socks with Natural color made by Paralleled Yarn" existing features can represent only the category "knee high". However, using name based features allows the learning phase to refer to attributes that directly impact the user's sense of fashion and preference, such as "Mori-Girl, Natural color, and paralleled yarn."
- The item name causes sometimes misunderstanding because of the sparsity problem and item name brevity. Similar to category and brand, we have to convert item name information into abstract form to be able to use it as features.
- For abstracting item name information, we adopt the hypothesis that similar items have similar purchase tendencies. To calculate item similarity, we regard item name as a set of characters (e.g. we regard "Oxford Button-Down Shirt" as {o, x, f, r, d, ' ', b, u, t, n, '-', w, s, h, i}) and we employ the Jaccard coefficient to measure the similarity of the names of two items.

Based on these assumptions, we construct name based features.

- **Name Averaged Sold Price** : $r_i - \langle r_i \rangle_{i \in (\mathcal{I}_{K,i} \cap \mathcal{I}_s)}$, difference between price of item_i, r_i and average price of items with similar top K item names.
- **Name Hotness** : $\frac{|\mathcal{I}_{K,i} \cap \mathcal{I}_s|}{K}$, the selling rate of items with similar top K item names.

[3] http://www.louisvuitton.com

[4] http://www.burberry.com

[5] "Mori-Girl" is a Japanese fashion trend for young women invoking a soft, forest-like tone.

5 EXPERIMENTS: PURCHASE PREDICTION

In this experiment, we construct model that predicts of the sales. Our model takes, as inputs, the attributes of the item of price, category, brand, and item name, and its output is binary value indicating the sales results. First, we explain our actual purchase dataset gathered from multiple EC sites and how we collected it.

5.1 Multiple EC Site DataSet And the Crawling Scheme

Our project was working with about 1,000 EC site users (called "Panel") and collects online purchase behavior activity logs over a long period of time by the log crawling software (called "Client"). Fig. 4 shows our data collecting system. User behavior logs collected by Client were annotated by Support Vector Machine[25] based model, and converted into certain format that was suitable for analysis. We use this purchase information in the following experiments. The main purpose of this experiment is making the prediction on each customer. The dataset was collected over the 23 Japanese EC sites listed in Table 2. We made sampling in order to ensure balance in terms of positive/negative data. For the EC sites of MUJI, Amazon, and RAKUTEN, we filtered out unrelated data to fashion items.

We performed two experiments. In Experiment 1 (Section 5.2), we used all the 23 EC sites. In Experiment 2 (Section 5.3), we selected 8 EC sites, 4 for target EC sites and the other 4 for source EC sites.

First, we explain about each EC site which were used as target data in Experiment 2. OUTLET PEAK is a fashion EC site and lays in a stock of items from fashion brands directly. MUJI is the largest commodity brand in Japan. Their EC site handles only one brand, "No Brand Quality Goods". Nissen originally handles mail order. Now they handle various fashion items such as woman's shirt, men's jacket, and baby's pajamas. By contrast, PEACH JOHN handles mainly woman's underwear with their own brand, "PJ".

Next, we explain about each EC site which were used as source data in Experiment 2. GLAMOUR SALES is the EC site focused on deal of the day, their sales have 157 hours limitation. They handle over 1,200 brands. ZOZOTOWN is the largest fashion EC site of Japan. They handle over 2,000 brands and over 130,000 items. They also manage a social networking service, ZOZOPEOPLE. UNIQLO is not only an EC site but also the largest casual fashion brand in Japan, such as MUJI. Their EC site handles very few brands; UNIQLO (their main brand), g.u (lower price items), and UT (tee-shirts only). RAKUTEN is a kind of online shopping mall. They are largest EC site of the business type called B2B2C (Business to Business to Consumer) in Japan. They do not deal with consumers directly (Business to Consumer), but also provide an E-Commerce platform where other companies are able to build up their own EC sites (called mall) in RAKUTEN. By opening their own mall on RAKUTEN, companies deal with consumers directly. Currently RAKUTEN has about 40,000 malls and various items over 0.1 billion from fashion clothes to real estate. Thus, there are many brands and categories. These EC sites on our experiment have several characteristics depending on their origin and business model.

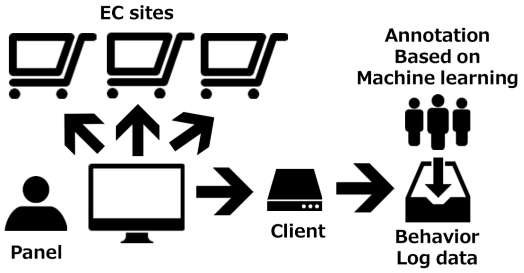


Figure 4: Data crawling scheme used by the project. Client software installed on Panel member's personal computers captures information about purchase behavior such as items that the Panel bought or queries they entered in Google, and sends them to our server. These data are converted into format suitable for analysis by statistical methods. After format conversion, the information is annotated by Support Vector Machine based method and checked by humans.

5.2 Experiment 1: Single Source Settings

5.2.1 Parameter Settings

For Experiment 1, we prepared to set 3 parameters; the number of initial base classifiers N, top K size using name based features, and the data size of bootstrap $|\mathcal{D}'_t|$. The number of initial base classifiers N was fixed to 100. In name based features, the top $K = 10$ items were used to identify similar items. The data size of each bootstrap $|\mathcal{D}'_t|$ equaled source data size, $|\mathcal{D}_S|$, 17,398.

We selected standard bagging[13], Frustratingly Easy Domain Adaptation[24], and TrBagg[21] as our verification methods. The base classifier used in each algorithm was C5.0[26] decision tree, which is an extension of the C4.5[27] algorithm. Abbreviations of method names are defined in Table 3. We performed a five-fold cross-validation test and used the average values.

5.2.2 Results

We tested whether the proposed method was superior to other methods in terms of accuracy. In experiment 1, we used all EC sites except RAKUTEN as target and RAKUTEN as source. The results are shown in Table 4. In Table 4, the columns list the target EC sites. Each row lists, from left to right, target EC site's name, the number of items in the site, and the remaining cells show the prediction accuracy for all methods. The values are the average of output by the five-hold cross-validation test.

We assessed these results from two view points;

1. whether OptTrBagg was superior to TrBagg
2. whether transfer learning worked effectively in item purchase prediction

Table 2: A list of target EC sites and their size of items.

Site Name	URL	# of items
FLAG SHOP	flagshop.jp	116
0101	0101.jp	124
SELECT SQUARE	selectsquare.com	128
Wacoal	wacoal.jp	146
SELESONIC	selesonic.com	156
SHOP CHANNEL	shopch.jp	220
ELLE SHOP	elleshop.jp	222
fashionwalker.com	fashionwalker.com	242
i LUMINE	i.lumine.jp	282
YOOX	yoox.com/jp	284
WORLD ONLINE	store.world.co.jp	300
OUTLET PEAK	outletpeak.com	322
MAGASEEK	magaseek.com	358
MUJI	muji.net/store	384
BRANDELI	brandeli.com	524
Nissen	nissen.co.jp	556
GILT	gilt.jp	584
Javari	javari.jp	676
PEACH JOHN	peachjohn.co.jp	712
Amazon	amazon.co.jp	988
GLAMOUR SALES	glamour-sales.com	1,382
ZOZOTOWN	zozo.jp	2,130
UNIQLO	uniqlo.com/jp	3,338
RAKUTEN	rakuten.co.jp	17,398

Table 3: Definition of abbreviations of method names.

Abbreviation	Method Name
DT	C5.0 decision tree[26]
BG	bagging[13]
FRUST	Frustratingly Easy Domain Adaptation[24]
TB	TrBagg[21]
OPT	OptTrBagg (proposed)

First, OptTrBagg showed better performance than TrBagg in all target data. In predicting MAGASEEK, OptTrBagg outperformed TrBagg as 5.9 points. Compared to Frustratingly Easy Domain Adaptation, OptTrBagg was superior for 22 of the 23 data sets. This confirms the validity of our OptTrBagg.

Second, we compare OptTrBagg to the standard learning methods. In comparison with standard learning methods, bagging outperformed standard C5.0 decision tree. This indicates the effectiveness of ensemble learning. Comparing OptTrBagg to bagging, OptTrBagg showed superiority to bagging in 12 EC sites. This means that prediction results by transfer learning were effective in some situations, but were not in 12 EC sites. We tried to find some tendencies when transfer learning did not work effectively.

Fig. 5 shows the tendency between the number of item size for each EC site and the improvement of accuracy achieved by OptTrBagg. X axis indicates the number of items in each EC sites and Y axis indicates the increase of accuracy offered by transfer learning (improved score between of OptTrBagg accuracy and bagging accuracy). This figure shows that the effectiveness of transfer learning. EC sites with over 1,500 items, such as ZOZO and UNIQLO, have sufficient items for constructing the prediction model. Transfer learning improved the accuracies of EC sites which have less than 200 items. This result indicates that these EC sites have insufficient data to construct a prediction model. Transfer learning effectively worked to train a model in these situations. On the other hand, the accuracies of transfer learning were inferior of bagging in some EC sites such as MUJI (-3.131 points) and OUTLET PEAK (-2.481 points). Next experiment was conducted to determine the most appropriate pairing and to examine the prediction results.

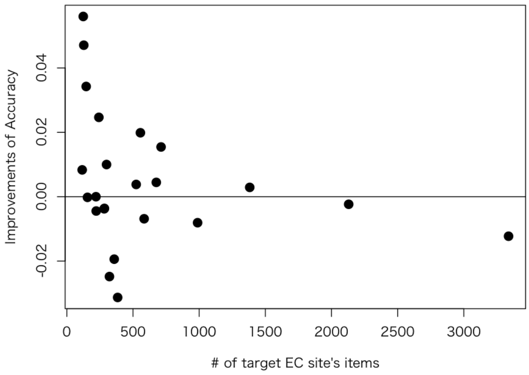


Figure 5: A scatter plot of the item size of target EC sites (X axis) and the accuracy improvements offered by OptTrBagg (Y axis).

5.3 Experiment 2: Multiple Source Settings

In Experiment 2, we intended to identify appropriate target/source pairs that improve accuracy. Then we checked the similarity of price distribution between source EC site and target EC sites in this experiment:

5.3.1 Parameter Settings

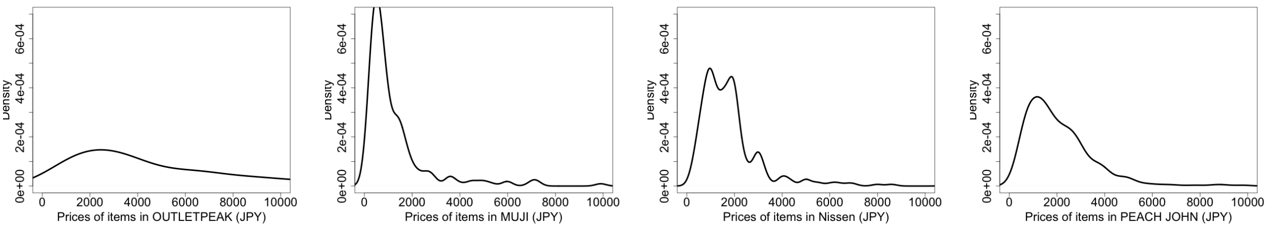
For Experiment 2, we prepared to set 3 parameters; the number of initial base classifier N, the top K size using name based features, and the data size of bootstrap $|\mathcal{D}'_t|$. At first, the number of initial base classifiers N was fixed to 100. In name based features, the top $K = 10$ items was used to identify similar items. The data size of each bootstrap $|\mathcal{D}'_t|$ equaled to the size of each source data size $|\mathcal{D}_S|$. The base classifier used each algorithm was C5.0 decision tree.

We selected OUTLET PEAK, MUJI, Nissen, and PEACH JOHN as target EC sites having the number of items around 500, because the prediction accuracies of OUTLETPEAK and MUJI were decreased by transfer learning, and Nissen and PEACH JOHN were increased by transfer learning. As source data in addition to RAKUTEN, we added 3 sites; GLAMOUR SALES, ZOZOTOWN, and UNIQLO, having the number of records around 1000.

Fig. 6 shows the price density distribution of items in each target EC site, OUTLET PEAK, MUJI, Nissen, and PEACH JOHN. Fig. 7 shows the price density distribution of items in each source EC site, GRAMOUR SALES, ZOZOTOWN, UNIQLO, and RAKUTEN. Fig. 6 and Fig. 7 shows difference of price distributions in each EC site clearly. If the prediction performance depended on the similarity of features associated with price, the transfer learning with similar price distribution would improve the accuracy.

Table 4: Results of experiment 1. Values are average accuracy of five-hold cross-validation. Bold number indicates the best accuracy among all learning methods and italic number indicates the best accuracy among transfer learning methods.

		Standard Learning		Transfer Learning		
target	# of items	DT	BG	FRUST	TB	OPT
FLAG SHOP	116	0.8442	0.9221	0.9047	0.9047	***0.9304***
0101	124	0.9033	0.9357	0.9837	***0.9917***	***0.9917***
SELECT SQUARE	128	0.8043	0.8988	0.8677	0.8911	***0.9458***
Wacoal	146	0.7945	0.8149	0.8147	0.8428	***0.8492***
SELESONIC	156	0.7567	**0.7823**	0.7052	0.7498	*0.7821*
SHOP CHANNEL	220	0.7409	**0.7818**	0.7409	0.7682	***0.7818***
ELLESHOP	222	0.7524	**0.8157**	0.7567	0.7747	*0.8112*
fashionwalker.com	242	0.9340	0.9547	0.9710	0.9628	***0.9793***
i LUMINE	282	0.7555	**0.7768**	0.7410	0.7236	*0.7731*
YOOX	284	0.9543	**0.9648**	0.9541	0.9506	*0.9612*
WORLD ONLINE	330	0.7700	0.8033	0.7700	0.8100	***0.8133***
OUTLET PEAK	322	0.7484	**0.8042**	0.7607	0.7517	*0.7794*
MAGASEEK	358	0.7655	**0.8210**	0.7375	0.7431	*0.8016*
MUJI	384	0.8098	**0.8358**	0.7810	0.7940	*0.8045*
BRANDELI	542	0.8016	0.8149	0.7825	0.8112	***0.8188***
Nissen	556	0.7428	0.7769	0.7356	0.7788	***0.7968***
GILT	584	0.7757	**0.8049**	0.7912	0.7364	*0.7981*
Javari	676	0.9275	0.9512	***0.9556***	0.9542	***0.9556***
PEACH JOHN	712	0.6756	0.6699	0.6489	0.6517	***0.6854***
Amazon	988	0.8290	**0.8320**	0.8057	0.8229	*0.8239*
GLAMOUR SALES	1,382	0.7771	0.7916	0.7663	0.7728	***0.7945***
ZOZOTOWN	2,130	0.9915	**0.9930**	0.9901	0.9901	*0.9906*
UNIQLO	3,338	0.6690	**0.6773**	*0.6699*	0.6606	0.6651
# of best accuracy among transfer learning		-	-	2	1	22
# of best accuracy among all methods		0	12	1	1	12



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

Figure 6: Price distribution of items in each target EC site used in Experiment 2, OUTLET PEAK (Fig. 6(a)), MUJI (Fig. 6(b)), Nissen (Fig. 6(c)), and PEACH JOHN (Fig. 6(d)).

5.3.2 Accuracy of target / source pair and their price distribution

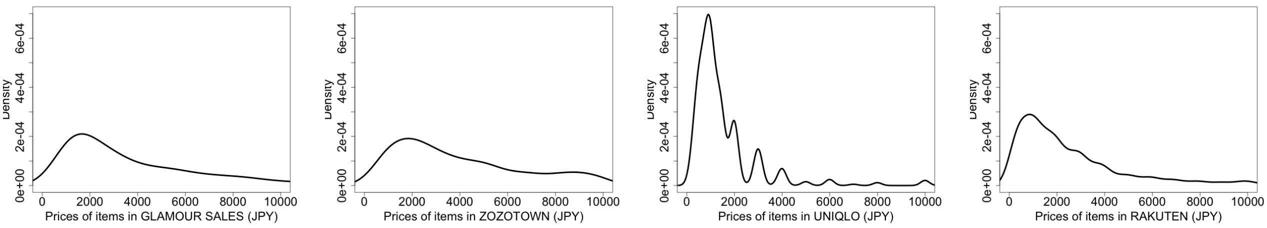
In Table 5, abbreviations of method names are defined. The results are shown in Table 5. The values in each cell were averaged accuracy of the five-hold crossvalidation. The **bold** number indicates the best accuracy among all learning methods and the *italic* number indicates the best accuracy among transfer learning methods.

First, the improvement of accuracy does depend on the target/source pairing. In OUTLET PEAK and MUJI, which transfer learning failed to predict in Experiment 1, transfer learning yielded better accuracy than standard bagging. Overall, none of the source data sets yielded the best accuracy for all targets (*silver bullet*) and none of the source data sets yielded the worst accuracy for all targets. It indicates the importance for transfer learning to select source data when constructing prediction models.

Second, we focused on the similarity of price distribution (Fig. 6 and Fig. 7) of each EC site. In each target and source pairing which yield best accuracy, the price distribution of the source EC site was similar with the target EC site. For example, the price distribution of OUTLET PEAK was similar with that of GLAMOUR SALES. These two price distribution of MUJI and UNIQLO ware also skewed. These observations suggest that the validity of transfer learning is determined by the similarity of features between the source EC data and the target EC data.

6 CONCLUSION

In this paper, we focused on prediction of the sales results using multiple EC site's purchase information. In order to construct the effective model, we converted the item's information such as brand, category, price, and item name into suitable formulation. We also intend to develop the effective method for finding the optimal pair of target and source data sets in transfer learning. The proposed “ OptTrbagg ” was a new method adopting transfer learning on EC marketing. We examined many Target and Source pairing and confirmed su-



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

Figure 7: Price distribution of items in each source EC site used in Experiment 2, GRAMOUR SALES (Fig. 7(a)), ZOZOTOWN (Fig. 7(b)), UNIQLO (Fig. 7(c)), and RAKUTEN (Fig. 7(d)).

Table 5: Results of experiment 2. Values are average accuracy of five-hold cross-validation. Bold number indicates the best accuracy among all learning methods and italic number indicates the best accuracy among transfer learning methods. In each row, bagging is standard learning method which uses target EC site's information only, and others use source EC site's information.

target	BG	source	FRUST	TB	OPT
OUTLET PEAK	0.8042	GLAMOUR SALES	0.7984	***0.8232***	0.8200
		ZOZOTOWN	0.7577	0.7794	0.7950
		UNIQLO	0.7640	0.8075	0.8104
		RAKUTEN	0.7607	0.7517	0.7794
MUJI	0.8358	GLAMOUR SALES	0.8047	***0.8463***	0.8411
		ZOZOTOWN	0.7865	0.8099	0.8334
		UNIQLO	0.8307	0.8150	***0.8463***
		RAKUTEN	0.7810	0.7940	0.8045
Nissen	0.7769	GLAMOUR SALES	0.7464	0.7716	0.7698
		ZOZOTOWN	0.7375	0.7662	0.7824
		UNIQLO	0.7534	0.7732	0.7606
		RAKUTEN	0.7356	0.7788	***0.7968***
PEACH JOHN	0.6699	GLAMOUR SALES	0.6742	0.6798	0.6741
		ZOZOTOWN	0.6798	0.6742	0.6699
		UNIQLO	0.6784	0.6811	***0.6867***
		RAKUTEN	0.6489	0.6517	0.6854

periority of our method . Experiments on the actual EC site data indicated that OptTrBagg outperformed TrBagg or the other transfer learning methods. Our composition was more tolerant against negative transfer by exploiting the sparsity structures of features of item. OptTrBagg could contribute to find most appropriate pairs of EC sites to determine the optimal pairing for transfer learning.

REFERENCES

[1] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich, To transfer or not to transfer, in In NIPS'05 Workshop, Inductive Transfer: 10 Years Later, 2005.

[2] X. Wu and A. Bolivar, Predicting the conversion probability for items on c2c ecommerce sites, in Proceedings of the 18th ACM conference on Information and knowledge management, ser. CIKM'09. ACM, 2009, pp. 1377–1386.

[3] D. Shen, X. Wu, and A. Bolivar, Rare item detection in e-commerce site, in Proceedings of the 18th international conference on World wide web, ser. WWW'09. ACM, 2009, pp. 1099–1100.

[4] E. Kim, W. Kim, and Y. Lee, Combination of multiple classifiers for the customer's purchase behavior prediction, Decis. Support Syst., vol. 34, pp. 167–175, January 2003.

[5] W. W. Moe and P. S. Fader, Dynamic conversion behavior at ecommerce sites, Manage. Sci., vol. 50, no. 3, pp. 326–335, Mar. 2004.

[6] D. V. D. Poel and W. Buckinx, Predicting online-purchasing behaviour, European Journal of Operational Research, vol. 166, pp. 557–575, 2005.

[7] Q. Guo and E. Agichtein, Ready to buy or just browsing?: detecting web searcher goals from interaction data, in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR'10. New York, NY, USA: ACM, 2010, pp. 130–137.

[8] S. Pandey, M. Aly, A. Bagherjeiran, A. Hatch, P. Ciccolo, A. Ratnaparkhi, and M. Zinkevich, Learning to target: what works for behavioral targeting, in Proceedings of the 20th ACM international conference on Information and knowledge management, ser. CIKM'11. New York, NY, USA: ACM, 2011, pp. 1805–1814.

[9] A. Bagherjeiran, A. Hatch, A. Ratnaparkhi, and R. Parekh, Large-scale customized models for advertisers, in Proceedings of the 2010 IEEE International Conference on Data Mining Workshops, ser. ICDMW'10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1029–1036.

[10] J. Li, P. Zhang, Y. Cao, P. Liu, and L. Guo, Efficient behavior targeting using svm ensemble indexing, in ICDM, M. J. Zaki, A. Siebes, J. X. Yu, B. Goethals, G. I. Webb, and X. Wu, Eds. IEEE Computer Society, 2012, pp. 409–418.

[11] M. Limayem, M. Khalifa, and F. Anissa, What makes consumers buy from internet? a longitudinal study of online shopping, IEEE Transactions on Systems, Man, and Cybernetics, vol. 30, pp. 421–432, 2000.

[12] S. Bellman, G. L. Lohse, and E. J. Johnson, Predictors of online buying behavior, Commun. ACM, vol. 42, pp. 32–38, December 1999.

[13] L. Breiman, Bagging predictors, Mach. Learn., vol. 24, pp. 123–140, 1996.

[14] Y. Freund and R. E. Schapire, Experiments with a new boosting algorithm, in Thirteenth International Conference on Machine Learning. Morgan Kaufmann, 1996,

pp. 148–156.

[15] B. Cao, N. N. Liu, and Q. Yang, Transfer learning for collective link prediction in multiple heterogenous domains, in Proceedings of the 27th International Conference on Machine Learning, ser. ICML'10, J. Fürnkranz and T. Joachims, Eds. Omnipress, 2010, pp. 159–166.

[16] T. Chen, J. Yan, G. Xue, and Z. Chen, Transfer learning for behavioral targeting, in Proceedings of the 19th international conference on World wide web, ser. WWW'10. New York, NY, USA: ACM, 2010, pp. 1077–1078.

[17] P. Huang, G. Wang, and S. Qin, Boosting for transfer learning from multiple data sources, Pattern Recogn. Lett., vol. 33, no. 5, pp. 568– 579, Apr. 2012.

[18] D. Pardoe and P. Stone, Boosting for regression transfer. in Proceedings of the 27th international conference on Machine learning, ser. ICML '10, J. Frnkranz and T. Joachims, Eds. Omnipress, 2010, pp. 863–870.

[19] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu, Video summarization via transferrable structured learning, in Proceedings of the 20th international conference on World wide web, ser. WWW'11. ACM, 2011, pp. 287–296.

[20] G.-R. Xue, W. Dai, Q. Yang, and Y. Yu, Topic-bridged plsa for cross-domain text classification, in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR'08. ACM, 2008, pp. 627–634.

[21] T. Kamishima, M. Hamasaki, and S. Akaho, Trbagg: A simple transfer learning method and its application to personalization in collaborative tagging, in Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, ser. ICDM'09, 2009, pp. 219–228.

[22] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, Boosting for transfer learning, in Proceedings of the 24th international conference on Machine learning, ser. ICML'07. ACM, 2007, pp. 193–200.

[23] A. Rettinger, M. Zinkevich, and M. Bowling, Boosting expert ensembles for rapid concept recall, in Proceedings of the 21st national conference on Artificial intelligence - Volume 1, ser. AAAI'06. AAAI Press, 2006, pp. 464–469.

[24] H. Daume III, Frustratingly easy domain adaptation, in Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Association for Computational Linguistics, 2007, pp. 256–263.

[25] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov., vol. 2, no. 2, pp. 121–167, Jun. 1998.

[26] J. R. Quinlan, C5.0, 2011, http://mloss.org/software/view/329/.

[27] R. Quinlan, C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning), 1st ed. Morgan Kaufmann, Oct. 1992.

Proposal for Knowledge Model using RDF-based Service Control for Balancing Security and Privacy in Ubiquitous Sensor Networks

Makoto Sato[*], Yoshimi Teshigawara[**] , and Ryoichi Sasaki[*,**]

[*]Graduate School of Advanced Science and Technology, Tokyo Denki University, Japan
[**]Cyber Security Laboratory, The Research Institute of Science and Technology, Tokyo Denki University, Japan
{sato_m, teshiga}@isl.im.dendai.ac.jp, sasaki@im.dendai.ac.jp

Abstract - In ubiquitous sensor networks, various sensors and tag readers automatically collect information in space and relevant information is acquired. Efficient utilization of the acquired information is important for providing high-quality services that meet the users' privacy requirements. We use RDF triples that represent spatial information at the granularity of the requested security levels. In earlier work, we created a knowledge model that considers privacy by representing the user information hierarchically, and we verified its feasibility by a simulator that we developed. Then, we extended this knowledge model. In this paper, we discuss our newly proposed extended knowledge model, its applicability to various spaces, and again evaluate the feasibility of the model in the simulator.

Keywords: Security and Privacy, Knowledge Model, RDF, Semantic Sensor Network Ontology, Sensor Network.

1 INTRODUCTION

In ubiquitous sensor networks, various sensors and tag readers automatically collect information in space and relevant information is acquired. Efficient utilization of the acquired information is important for providing high-quality services that meet the users' privacy requirements. It is expected that the amount of information of the sensor network space will further increase due to advances of these networks. It is also expected that the personal information of users will be presented with various granularities. In this regard, it is possible to identify users' personal information by combining sensor information with user information that seems to be trivial by itself. Therefore, the risk of an indirect violation of privacy makes it difficult to provide high-quality services, because protecting the user's privacy means limiting the information obtained.

We have been developing a platform that integrates all the information in a space by using the Resource Description Framework (RDF). An RDF represents information about a resource (subject, predicate, object) in the form of an RDF triple [1]. An RDF triple was represented by a directed graph in Figure 1. The RDF expresses the subject of the resources associated with the object through the predicate. By combining inference rules and a set of vocabulary, it is possible to connect different types of data and to make the aggregation of over the partial sums. RDF triples are represented with the granularity of any spatial information. Therefore, service control information or privacy information is represented flexibly. For this reason, using the information efficiently to provide a flexible service requires organizing the RDF triples of the control information and the service state information of the space required by each service.

On the other hand, protecting personal information requires collecting this information with restrictions and proper control. In our previous study, we discussed only the collecting restrictions, because the use of restrictions is outside our work scope [2]. Thus, we defined privacy protection as follows. Sensors were not permitted to collect unintended information of users. Services were allowed to use only intended information of users.

In this study we previously proposed a knowledge model that can be applied to a platform using RDF-based practical services [3]. We created the knowledge model, which is a set of vocabulary required for expressing services provided by the RDF and analyzing the RDF obtained at that time. Because our knowledge model considers privacy by representing the user information hierarchically, we were able to control the user information by adding a function that reflected user requests [4]. In addition, we verified the feasibility of the knowledge model by developing a test simulator.

The knowledge model is represented by simple and common logic. For this reason, the service provider benefits by verifying whether the personal information is properly used when the service is under development. The user benefits by limiting personal information in accordance with the user's intention.

Our current work provides a level of service management for a particular space. That is, we extended the same service to a different service management system. In this paper, we discuss the applicability of our extended knowledge model to various services and various spaces, and we evaluate the feasibility of the extended model by using the simulator.

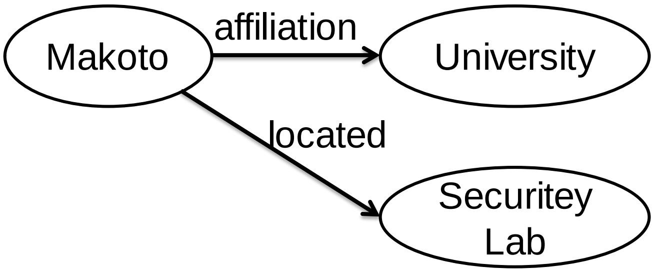


Figure 1 Example of RDF Triple

2 RESEARCH BACKGROUND

2.1 Related Work

Various integrated management method of the sensor network have proposed[5]. Some of the studies represents sensor network information by using RDF. Fujinami et al. represent a physical environment model by a location model and an object model using RDF[6]. The location model is represented by relationships between an unit space such as a room and a building and unit territories such as entrance and kitchen. The object model is represented by object information such as specifications information and operating condition. By using these models, developers can handle directly required information for a variety of applications. Held et al. represent the user specific information such as user preferences by using RDF [7]. By evaluating context information and managing the user profiles, it allows for personalized, context-aware service mediation and content adaptation. Noguchi et al. managed sensor information by using the RDF to realize intelligent support systems in a room at a home [8]. Therefore, the system needed a mechanism for automatically understanding information such as sensor configurations of rooms. Therefore, they proposed an RDF sensor description to inclusively portray the sensor information. It not only could describe the characteristics of the sensors, but also easily realize an extension of the description in collaboration with other knowledge information, including new information. With these features, it allowed unified processing of sensor data. An example of the applied RDF description is the implementation of applications, such as component discovery in the middleware. In our study, the service execution rules and the user requirements are centrally managed in the same way as the sensor information. Therefore, our model is expected to provide both high-quality service and protection of privacy.

Some researches discuss access control on the Web. Sacco et al. propose the Privacy Preference Ontology that enable fine-grained access control[10]. This ontology is contained some properties such as the type of restriction and which resource to restrict access. By using this ontology, access control to privacy information are restricted by the properties which a requester must satisfy. Carminati et al. propose an access control framework for social networks by specifying privacy rules using the SWRL (Semantic Web Rule Language). Additionally, user/resource relationships are modeled by using RDF/OWL (Web Ontology Language)[9]. Because there are a lot of privacy information on the Web, access control is effective as a method of privacy protection. Similarly, privacy protection using RDF in the sensor network is studied. Jagtap et al. investigated privacy protection by using the RDF [11]. They proposed a model for representing the user's environment, position, and their activities. An important element of their study was the use of collaborative information among sharing devices, which share and integrate knowledge about the contexts of the collaborative information. Therefore, mechanisms for privacy and security were required. They used the RDF to specify the high-level declarative policy describing the settings for sharing user information.

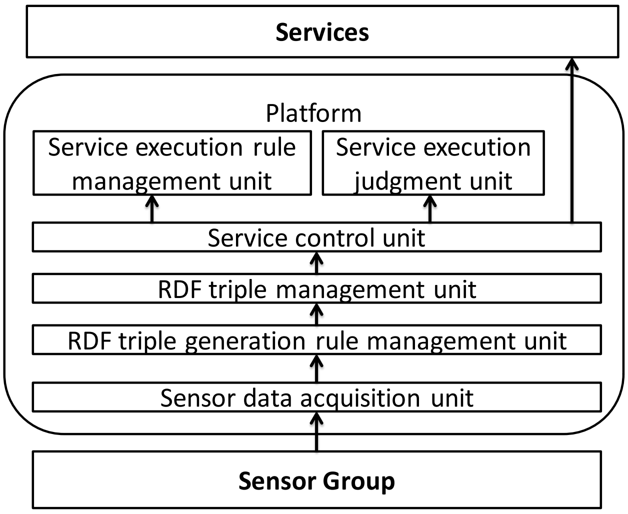

Services
Platform
Service execution rule management unit
Service execution judgment unit
Service control unit
RDF triple management unit
RDF triple generation rule management unit
Sensor data acquisition unit
Sensor Group

Figure 2: Overview of system functions

Our study presents a framework to provide users with an appropriate level of privacy for a mobile device and to protect the personal information gathered, including personal information that can be inferred from other information. Our study assumes an environment where the mobile devices are owned by individuals, and sensors, such as camera sensors and positioning sensors, are placed in each location. Therefore, our model is expected to protect privacy while providing a variety of services.

2.2 Development of Platform

As described in Section 1, we have been studying a method to integrate all the information in a space by describing the sensor information, the user information, and the service states for using the RDF [12]. We aim to control services in the sensor network space by using RDF triples to provide services and information corresponding to the users' requests. Furthermore, to provide high-quality service and to protect the privacy information of the user by reflecting the user requirements into the usage rights of RDF triples, we have been developing a platform that uses the appropriate information that satisfies the users' requirements. Figure 2 shows an overview of the functions of this platform. The functions of the platform are to generate RDF triples from the spatial information acquired from sensors and to select provided services based on the RDF triples. These processes are carried out in "the RDF triple generation rule management unit" and "the service execution rule management unit". In "the RDF triple generation rule management unit", RDF triples are generated from the acquired sensor information based on the RDF triples from the RDF triple generation rules. Here, the generating rule for the RDF triples is managed as a set of rules, or triggers, for generating new and already generated RDF triples. Table 1 shows an example of a generated RDF triple rule. The service execution rule management unit selects the services that can be provided by checking the RDF triples passed from the service control unit to the service execution rules.

Table 1: Generated RDF triple rule

Rule	Generated RDF triple
userA is located at (x,y)	(userA, locate, (x,y))

Table 2: Service execution rule

Service	RDF triple			Excutive instruction
	Subject	Predicate	Object	
Lighting control	User	locate	Room	Light on

In addition, the service providing service execution rules, the RDF triples that trigger the service, and the service execution instruction are managed as a single set of rules. Table 2 shows an example of a service execution rule.

2.3 Creation of Knowledge Model

As described in Section 2.2, spatial information is represented by an RDF. We define the vocabulary and the relations of spatial information as a knowledge model expressed by the RDF. To create a knowledge model that can provide a service, it must be created after stipulating the service requirements envisioned. However, the service that runs on this platform is not yet defined. Therefore, an effective approach is to create a primary knowledge model first and then extend it gradually.

The primary knowledge model is created with a clear description of the technical issues for practical use, while considering and evaluating the services as a prototype. Specifically, the resources required for the services are assumed. Next, the state transition of the resources is expressed by an RDF graph (a set of RDF triples). Then, the resources within the RDF graph are classified into sets of the same type. A knowledge model is created to represent the relationship between the sets. For example, the primary knowledge model is applied to a service of the same type. A new service concept is introduced when one is lacking. Thus, by extending the knowledge model, a more general knowledge model is created.

In such a manner, we created the knowledge model shown in Figure 3, which is intended for a university. In this figure, an ellipse represents a resource, and an arrow expresses a predicate. A feature of this knowledge model is that the domain corresponds to a subject, and the range corresponds to an object, shown as (*predicate_property*, domain, *domain_name*), (*predicate_property*, range, *range_name*). This RDF triple expresses a resource that is the subject of the relationship and the object of the relationship. Thus, when RDF triples are added to the RDF graph, the inference is that resources belong to a classification with a focus on the predicate [3]. In addition, by using a hierarchical representation of the affiliation information of the user, it becomes possible to restrict the use of private information [4].

We examined the flexibility of the service execution rules by an experiment using this knowledge model in a simulator [3].

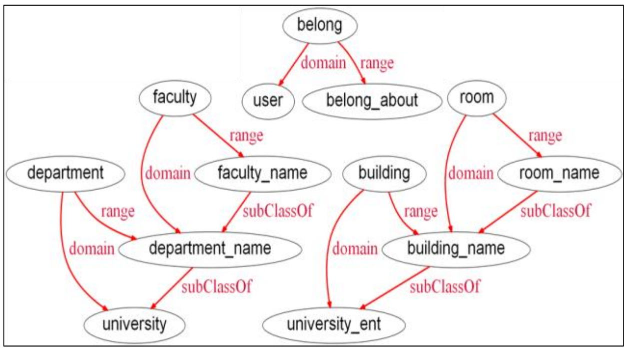


Figure 3: An example of the created knowledge model

2.4 Development of Simulator

No real system has been developed to provide services by using the knowledge model created in Section 2.3. Because the platform includes ambiguous parts, such as the storage method of the service execution rules, we cannot clearly verify the feasibility of the knowledge model. Therefore, we developed a simulator to apply the knowledge model [13].

In the simulator, we developed several functions, such as input of RDF triples, introduction of new RDF triples by inferring, reflection of user requests, and selection of executable services. Jena was used for development of the simulator [13]. Jena provides a framework for processing RDFs, and an inference engine. Graphviz is used to visualize the RDF graph. We demonstrated the operation of each function and verified the feasibility of a service control by the knowledge model [4].

One of the beneficial features of the simulator is a function for reflecting user requests in order to limit the information used in the sensor network space. In Figure 2, this function is executed in the RDF triple management unit. The user requests are managed in the form of inference rules. Specifically, RDF triples representing the restrictions (*user information*, permit, no) are added by using the inference rules, and only usable information is outputted based on these added RDF triples.

2.5 The Need for the Sensor Concept and Collecting Restrictions

The main resources in the sensor network space can be divided into space, user and service. Space is divided into "environment" and "sensor". "Environment" is a place for providing services. For example, the environment is stations, a university or a home. "Sensor" obtains the spatial information. In the previous study, we also focused on the service control using RDF, but it was limited to only "environment" in spatial information. We considered service control information as information directly related to the service. We did not discuss the sensor. Therefore, as a next step, it is necessary to incorporate the concept of the sensor to create a more general knowledge model.

In addition to introducing the concept of the sensor, it is also necessary to again consider the restrictions of collecting information, as discussed in our previous study [4]. To meet the requirements of more users, the RDF triples acquired from a sensor only use information allowed by the user. In addition, RDF triples are not applied except for those uses. The results of the study are as follows.

3 EXTENDED KNOWLEDGE MODEL

3.1 Assumed Service

The assumed service is the entry management in a university campus according to the affiliation information of a user. For example, one service is unlocking the entrance door if the user is enrolled in the affiliated faculty. We make the following assumptions. The service manager is able to attach the affiliation information of the user (faculty), and the user is able to specify the affiliation information for which the service is available. We considered the "environment" as a cafeteria, three buildings, five rooms in each building and the main gate to the university campus. Figure 3 illustrates the relationship of entry permission and user affiliation. A user must belong to the university in order to pass through the main gate. Similarly, the faculty must belong to the university in order to enter a building. The faculty must also belong to a corresponding department in order to enter a classroom. For example, users can enter room A1110 if they have the affiliation information of faculty. They can enter the cafeteria if they have the affiliation information of the university. Each entry has a keycard system in conjunction with the entry permit information outputted from the service, and the RDF triples are assumed to have been inputted into the system in advance by the service administrator.

As described in Section 2.3, our RDF-based service extracts the information required for the service provision from the assumed service. Service execution rules are created by analyzing the information to trigger the service execution from the service contents. The service is then provided when the affiliation information of the user is inputted and room entry is allowed. For example, an RDF triple indicates whether a user can pass through the main gate of the university (university, entrypermit, maingate). In another example, the trigger for room entry is represented by the RDF triples (user, affiliation, Department of Information Media), (user, permit, Room A1110). Table 3 shows the service execution rule of room entry.

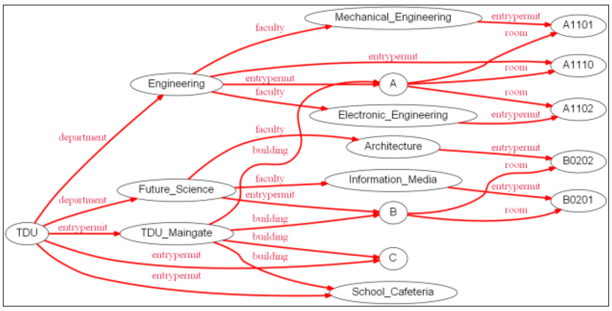


Figure 4: RDF graph for entry permit information for faculty user

Table 3: Service execution rule of room entry

Service	RDF triple			Executive instruction
	Subject	Predicate	Object	
Entry Management	User	permit	Room	Open
	User	affiliation	Affiliation information	

3.2 Privacy of Extended Knowledge Model

We created an extended knowledge model based on the assumed service described in Section 3.1 [4]. Specifically, the model contains the affiliation information of the user and the model of the university's sensor network space corresponding to "environment".

For the "sensor" that obtains spatial information, we used the Semantic Sensor Network Ontology (SSN) proposed by the W3C [15]. This ontology describes sensors, observations, and related concepts. For example, Sensor, Sensor Output, Sensor Input, and Device are basic resources of the sensor [16]. Therefore, we consider that these resources are sufficient as a primary model of the sensor.

It is necessary to incorporate the collecting restrictions, as described in Section 2.5. To indicate the availability of the sensor, we added a predicate "hasAvailability". The predicate is the relationship between the user and the sensor. In addition, a word with the prefix "ssn." indicates that it is a vocabulary of the SSN. Figure 5 shows the extended knowledge model based on this information.

Moreover, we needed to introduce new inference rules on the collecting restrictions. Therefore, if the user does not have the "hasAvailability" predicate, the simulator does not use the affiliation information of the user (Formula 1). If the RDF triple (*user*, hasAvailability, *sensor*) is not added, the sensor stops working (Formula 2). The collecting restrictions are realized by these formulas.

$$\mathrm{Affiliatio\ n}(?user, ?affiliationof) \wedge$$
$$\mathrm{noValue}(?user, \mathrm{ssn.hasAva\ ilability})$$
$$\rightarrow \mathrm{Permit}(?affiliationof,\ \mathrm{no}) \qquad \ldots(1)$$

$$\mathrm{noValue\,of}(\mathrm{ssn.hasAvailabilityBy}, ?Sensor)$$
$$\rightarrow \mathrm{Permit}(?Sensor, \mathrm{down}) \qquad \ldots(2)$$

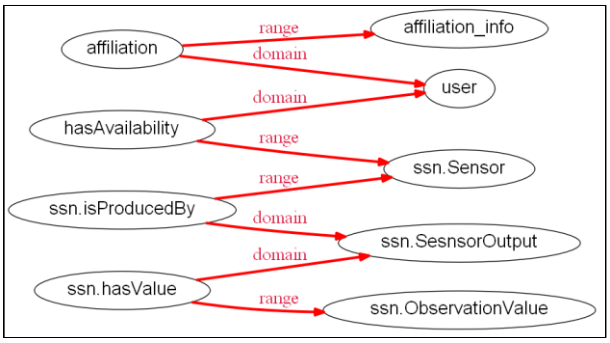


Figure 5: A part of the extended knowledge model

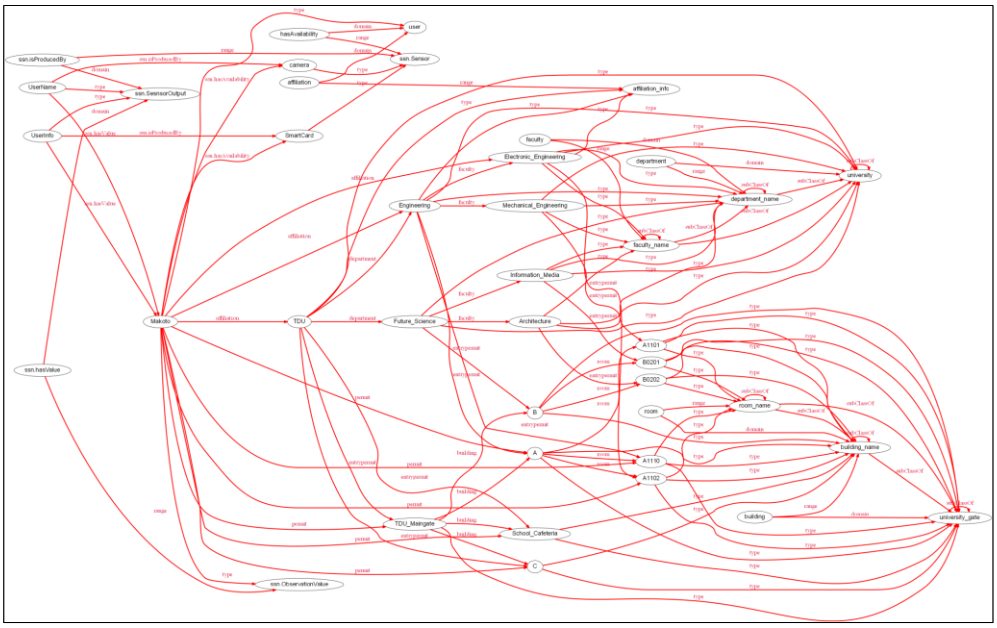


Figure 6: RDF graph after inference processing (Scenario A)

4 SIMULATION EXPERIMENTS

This section describes simulation experiments. This experiments were executed to verify the extended knowledge model to be able to protect privacy information by using user affiliation information.

4.1 Experimental Environment

The environment is the same as that described in Section 3.1. We made the following assumpt ions for the experiment. Each user has a smartcard. The physical location of the university is the sensor network space. The physical location can be uniquely identified by the geographical coordinates of latitude and longitude. All locations in the university have the names of identified strings assigned by public authorities. Sensors are installed in the vicinity of the door or the gate for each location. Sensors used in this space are a camera sensor and a smartcard reader. The camera sensor obtains the name information of the users present in the space. The smartcard reader obtains the affiliation information of the users. The individual can then be authenticated by comparing the information in the smartcard and the information acquired by the camera sensor. The acquired sensor information is converted to RDF triples automatically. The difference from the previous experiment [4] is the sensor information and increased number of relationships. The relationships of entry permission and affiliation information of the user are shown in Table 4.

We assumed the following two scenarios in the experiments. The difference between the scenarios is that the input on the sensor information is added only in scenario A (Steps 2 and 3). Then, service execution rules and the knowledge model are assumed to be stored in the database in advance. The user is engineering faculty at this university in the Department of Electronic Engineering.

Scenario A:

1) Input the initial state of the assumed space. Specifically, the service administrator enters RDF triples indicating the building permit and the space information into the simulator.
2) Enter the user requirements. The user selects the available user information and the available sensors and enters them into the simulator. All sensor and user information is supposed to be available at this time.
3) The acquired sensor information, such as affiliation information, is inputted in the simulator.
4) Generate new RDF triples to perform inference processing by using the input information.
5) Determine whether the entry management service can be performed by using an RDF graph.
6) Suggest the possible entry locations.

Scenario B:

A similar scenario is carried out, but this one does not use the available sensors in Step 2 above.

We evaluated the feasibility and flexibility of the extended knowledge model. In addition, we evaluated the feasibility of collecting restrictions to meet the user requirements.

Table 4: Some of the inputted RDF triples

Subject	Predicate	Object
University	department	Engineering
University	department	Future Science
Future Science	faculty	Information Media
A	room	A1101
University	entrypermit	Maingate
Engineering	entrypermit	A
Information Media	entrypermit	A1110

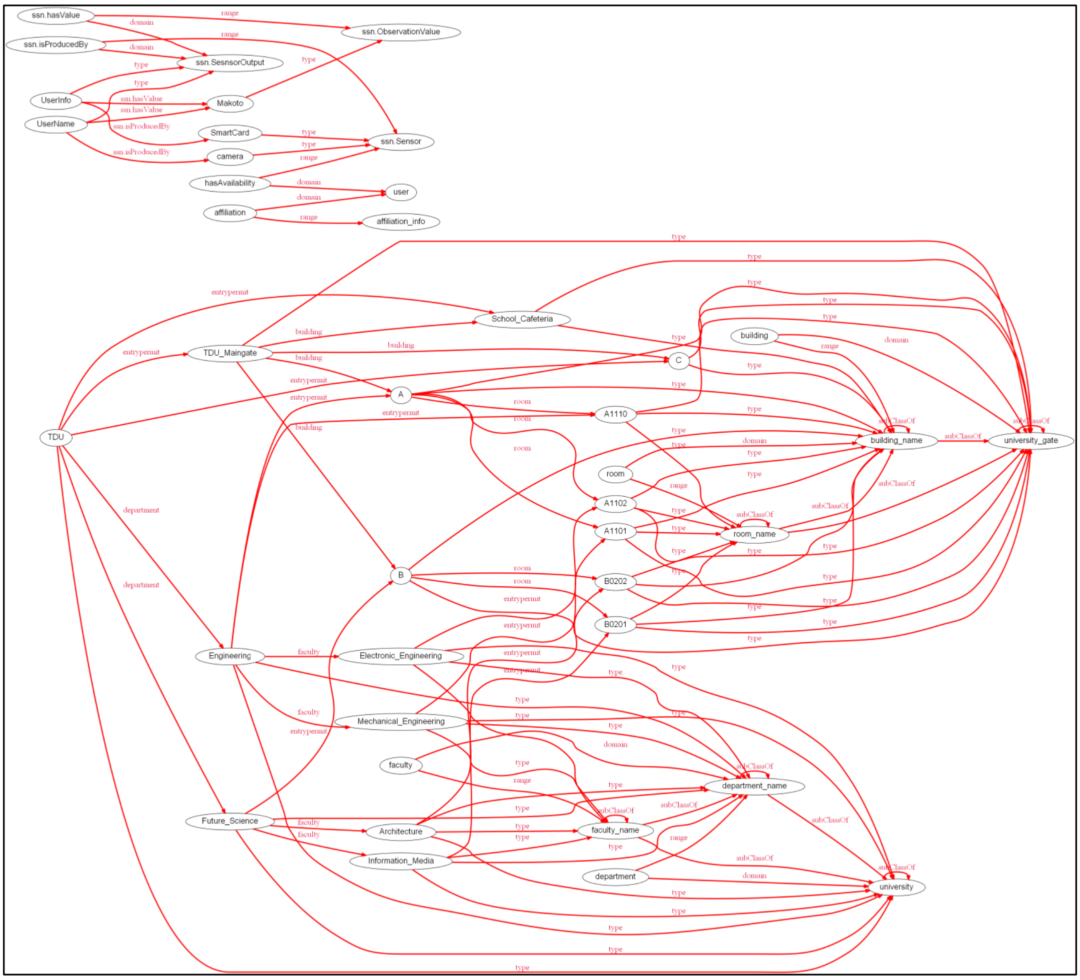


Figure 7: RDF graph after inference processing (Scenario B)

4.2 Experimental Results

Here we present the results of the experiments described in Section 4.1, where the scenarios were executed by the simulator. Figure 5 illustrates a part of the result of Step 1. It can be seen that the smartcard and the camera sensor were associated with the SSN.

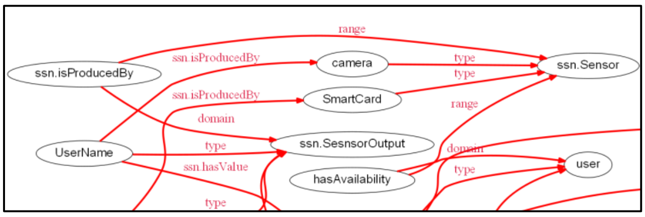
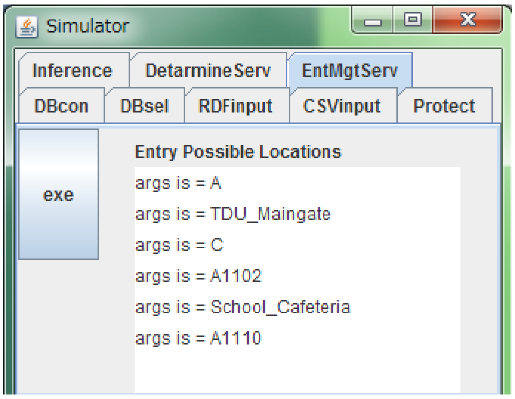


Figure 8: A part of the knowledge model

Figure 6 illustrates the result of Step 4 in scenario A. From this figure, it can be seen that by inference processing, many RDF triples were automatically generated. Figure 7 shows a list of the rooms where the user is permitted to enter. As compared with Figure 3, all the permitted rooms are listed.



```
Entry Possible Locations
args is = A
args is = TDU_Maingate
args is = C
args is = A1102
args is = School_Cafeteria
args is = A1110
```

Figure 9: Screenshot of the list of locations that user is permitted to enter (Scenario A)

Figure 8 shows the results of applying the inference rule collecting restrictions on the user request inputted in Step 2 in scenario B. The available information was eliminated because the user does not have the "hasAvailability" predicate. Figure 9 illustrates the result of Step 4. As compared with Figure 6, Figure 9 shows that the RDF graph is divided into two groups. This is because the affiliation information of the user is not bound to the entry permission information. As the result, the entry management service is not provided, as shown in Figure 10.

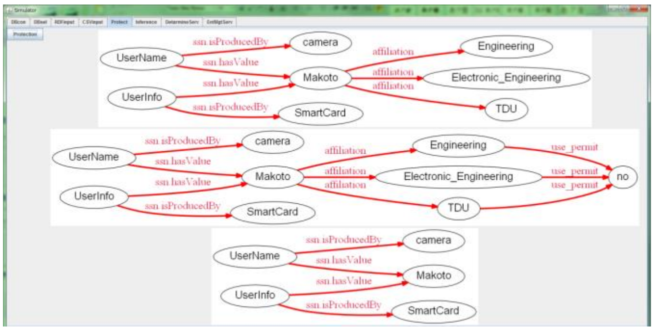
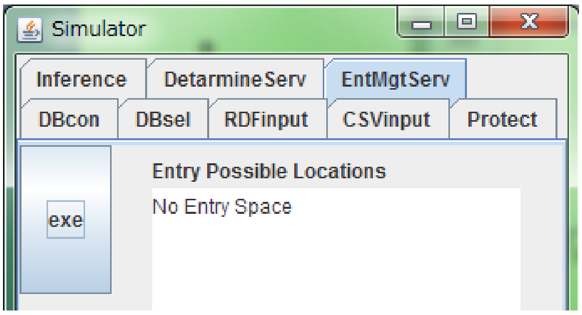


Figure 10: Screenshot of the result of collecting restrictions (Scenario B)



Entry Possible Locations

No Entry Space

Figure 11: Screenshot of a list of rooms that the user is permitted to enter (Scenario B)

4.3 Discussion

The collecting restrictions were executed in accordance with the user by using the extended knowledge model. Entry in permitted rooms was properly listed in the newly defined space. In the two scenarios, we confirmed that the provision of services can be automatically executed. This result shows the feasibility of the sensor network space by using the knowledge model in a variety of spaces.

The RDF graph in the middle part of Figure 10 can be derived from the inference rule in formula 1 and the RDF graph in the upper part of Figure 10. The derived RDF graph shows the user affiliation information cannot be used. Therefore, it was considered that restriction of privacy information that satisfies the users' requirements was fulfilled. However, the resource indicating the user name was remained in the RDF graph of the lower part of Figure 10. Therefore, it seems that there is a need to be erased this privacy information.

The results of this study shows a possible resolution to the security issue in privacy protection.

5 CONCLUSION

The purpose of this study was to confirm whether an RDF-based service implementation method keeps a balance of service provisions that efficiently employ state information and privacy protection at the same time in a ubiquitous sensor network. In this paper, we extended the knowledge model by introducing the concept of a sensor by Semantic Sensor Network Ontology. In addition, we expressed the collecting restrictions by adding inference rules. We also verified the feasibility of the extended knowledge model and collecting restrictions by experiments using the simulator that we developed. As a result, we see a possible resolution to the issue of balancing security and privacy.

REFERENCES

[1] W3C, RDF Primer (online), from <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.

[2] IPA, Survey on IT Technology and Personal Information protection (in Japanese), IPA (2012), from <http://www.ipa.go.jp/security/fy23/reports/pdata/ >.

[3] M. Sato, K. Awazu, K. Kato, and Y. Teshigawara, A Study on RDF Based Service Implementation in Ubiquitous Sensor Networks (in Japanese), Proc. of Multimedia Distributed Cooperative and Mobile Symposium (DICOMO2011), pp. 749-756 (2011).

[4] M. Sato, and Y. Teshigawara, A Proposal of a Knowledge Model in Consideration of Privacy for the RDF-based Service Control in Ubiquitous Sensor Network (in Japanese), Proc. Computer Security Symposium (CSS2012), pp. 246-253 (2012).

[5] Y. Hirota, H. Kawashima, T. Umezawa, and M. Imai, Design and Implementation of Real World Oriented Metadata Management System MeT for Semantic Sensor Network (in Japanese), The IEICE Transactions, Vol.J89-A, No 12, pp. 1090-1103 (2006)

[6] K. Fujinami, and T. Nakajima, An Information Management Infrastructure for Sentient Artefact-based Smart Spaces (in Japanese), IPSJ Transactions on Computing System, Vol. 47, No. SIG12(ACS 15), pp. 399-410 (2006)

[7] A. Held, S.Buchholz, and A. Schill, Modeling of Context Information for Pervasive Computing Applications, In Procceding of the World Multiconference on Systemics, Cybernetics and Informatics, Springer(2002)

[8] H. Noguchi, K. Tanaka, T. Mori, T. Sato, Room Situation Search System Based on RDF Describing Room Object as Target of Human Behavior, Technical Report of IEICE, Vol. 104, No. 725, pp. 31-36 (2005).

[9] B. Carminati, E. Ferrari, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, A Semantic Web Based Framework for Social Network Access Control, Proceedings of the 14th ACM symposium on Access control models and technologies, pp. 177-186 (2009)

[10] O. Sacco and A. Passant, A Privacy Preference Ontology (PPO) for Linked Data, Procs of the 4th Workshop about Linked Data on the Web(LDOW-2011) (2011)

[11] P. Jagtap, A. Joshi, T. Finin, and L. Zavala, Preserving Privacy in Context-aware Systems, 2011 Fifth IEEE International Conference, pp. 149-153 (2011).

[12] K. Awazu, D. Hirashima, K. Kato, and Y. Teshigawara, A Study on Dynamic Space Administration and Service Control by Using RDF in Consideration of Privacy in Ubiquitous Sensor Networks (in Japanese), Proc. Multimedia Distributed

Cooperative and Mobile Symposium (DICOMO2010), pp. 1318-1325 (2010).

[13] M. Sato, K. Awazu, and Y. Teshigawara, A Proposal of a Simulator for the RDF Based Service Control in Ubiquitous Sensor Networks (in Japanese), Proc. Multimedia Distributed Cooperative and Mobile Symposium (DICOMO2012), pp. 921-928 (2012).

[14] J.J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson, Jena: Implementing the Semantic Web Recommendations, Proc. 13th Int'l World Wide Web Conf. Alternate Track Papers and Posters, pp. 74-83 (2004).

[15] W3C Semantic Sensor Network Incubator Group, Semantic Sensor Network Ontology (online), from <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn >.

[16] M. Compton et al., The SSN Ontology of the W3C Semantic Sensor Network Incubator Group, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 17, pp. 25-32 (2012).

Computing Evaluation Scores with An Arbitrary Aspect from Evaluation Texts in Review Sites

Satoru Hosokawa[†], Etsuko Inoue[‡], Takuya Yoshihiro[‡*], Masaru Nakagawa[‡]
[†]Graduate School of Systems Engineering, Wakayama University, Japan
[‡]Faculty of System Engineering, Wakayama University, Japan
[*]tac@sys.wakayama-u.ac.jp

Abstract - Recently, evaluation sites have become to be popular in which we can share evaluation comments over various objects including restaurants, shops, and commercial products. In such sites, users can write evaluation comments as evaluators, and also refer to the comments of others to grasp the evaluation of the object that the users are interested in. To grasp the evaluation of an object in such sites is, however, very laborious because users have to look over too many evaluation comments. In this paper, to reduce the labor to grasp the evaluation, we propose a method to compute numerical scores of objects from a set of evaluation comments with an arbitrary given aspect, which can be determined according to users' own preferences. With this method, users can refer to numerical scores of various objects with their own free aspects in order to reduce the objects to compare, so they can reduce the labor in grasping evaluation by reading evaluation comments for only high-score objects.

Keywords: Evaluation Analysis, Aspects, Evaluation Scores

1 Introduction

The Internet has grown rapidly in the several decades, and which enabled people to state their opinions or comments in public. As an example of the public statements, several review sites appeared, in which people write evaluation values or review comments for various evaluation entities such as restaurants, shops, and products for sale. This kind of web sites plays an important role to share so called word-of-mouth information among users of the Internet; Some part of people write their evaluation values and review comments into the site, and others refer them. These sites actually are useful for people to select shops or products to buy, or for companies in their marketing activities.

In review sites, however, people generally have to read so many evaluation comments as to grasp the real evaluation for each entity because the reviews are often quite different depending on individuals and further sometimes include unreliable or irresponsible comments. The problem here is that it requires considerable labor to refer and examine these review comments. One direct approach for this problem is to summarize the review comments so as to reduce the amount of comments to read. However, such a simple summarization rarely works well because in many cases the number of evaluated entities and the review comments are too large, and also currently the quality of summarizing techniques are not sufficiently high.

As another approach to reduce the labor of users, it is possible to sort the evaluated entity in the order of evaluation scores, and users only see the review comments of high-rank (e.g., top-10) entities for their selection of entities to buy or use. A history of studies to compute evaluation scores for entities is seen in the literature. As seminal work, Turney [1] proposed a method to classify review articles into two levels of polarity, *positive* and *negative*. Koppel et al. [2] extend the method to classify them into three polarity levels, *positive*, *neutral*, and *negative*. Later, they lead to methods compute finer grained numerical scores, say, rating of entities [3][4]. However, their methods are not based on particular "aspects" of evaluation, so they cannot follow variation and difference of users' viewpoints or tastes. The viewpoints or tastes in evaluating entities are usually different depending on individuals, so these methods would not meet the requirements that users would like to know the evaluation results from various practical aspects.

On the other side, there are several studies on analysis of review comments considering various aspects in evaluation. References [5] and [6] consider typical evaluation aspects, and summarize review comments with each evaluation aspect through retrieving sentences related to each evaluation aspect. Here, the typical evaluation aspects for hotels, for example, would be "cleanness," "location," "service of staffs," etc. These studies assume that such an evaluation aspect is given as a few words in advance. Their methods actually consider several aspects in evaluation, but they only treat typical evaluation aspects given by simple words. Therefore, they do not sufficiently cover the requirements of users because users' requirements have large diversity of aspects reflecting on wide variety of users' viewpoints and tastes in evaluating entities.

As for the diversity of evaluation aspects, several studies [8][9][10] try to retrieve words as "topics" that represent evaluation aspects. If we retrieve topics using these methods and summarize review comments with each topic, we may cover larger diversity of users' requirements. Also, we can compute evaluation scores instead of summarizing texts. Then, users will achieve efficient use of review sites by reducing their labor via referring evaluation comments of only high-rank entities. However, these methods do not cover all possible aspects of users, and the range of "topics" is still limited within a word.

In this paper, we propose a method to compute evaluation scores of entities to reduce labor of users to grasp evaluations in review sites, while covering all possible evaluation aspect of users. In our study, we assume that an aspect for evaluation is given as a form of text description, and we compute the evaluation score with the given aspect. For example, in case

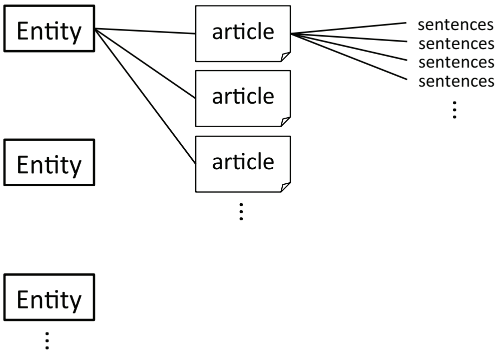


Figure 1: The Structure of Review-site Data

of restaurants, "Good restaurant for family with reasonable cost" can be a typical practical aspect description. With our method, users can obtain an ordered list of evaluation entities with respect to the computed evaluation scores, and so they can focus on high-rank entities based on their own evaluation aspect, which enable them efficient use of review sites. To the best of our knowledge, this is the first method to compute scores based on a text-style aspect description.

This paper is organized as follows: In Section 2, we describe the proposed method that computes evaluation scores with respect to the given aspect description. In Section 3, we give an evaluation results for the proposed method, and show that the method computes the evaluation scores that fit the feeling of the users of review sites. Finally, we conclude the work in Section 4.

2 Computing Evaluation Scores with An Arbitrary Aspect Description

2.1 Overview of the Proposed Method

In this paper, we compute a numerical score for each entity based on the given text description of an evaluation aspect. The structure of the review-site data is shown in Fig. 1; For each entity to be evaluated, we have text evaluation articles that consist of many sentences, which forms a tree structure. The proposed method, which computes a score for each entity from this form of data, consists of three folds:

(a) Learning a dictionary of evaluation words.

(b) Computing the evaluation score for each sentences included in each review article.

(c) Computing the evaluation score for each entity using the scores of the sentences computed in step (2).

Fig. 2 illustrates the overview of the proposed method. First, (a) we learn a dictionary of evaluation words from a small data set of review sites with human annotations, and then (b)(c) compute the evaluation score for each entity. Here, the data set to learn the dictionary could be different from the full dataset of the review-sites, could be rather small data set, but the sort of entity to be evaluated should be the same as the review-data with which we compute the evaluation score. (Namely, if you want to evaluate restaurants, then the data set to learn the dictionary should include the evaluation articles for restaurants.) Further, note that the data set to learn the dictionary should include human annotations; for each sentence in the evaluation articles, a reliable person should perform the following.

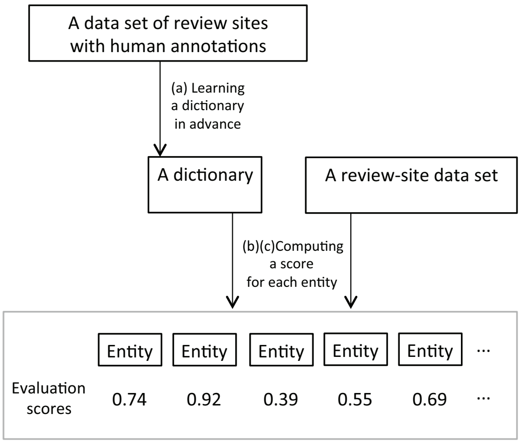

A data set of review sites with human annotations
(a) Learning a dictionary in advance
A dictionary
A review-site data set
(b)(c)Computing a score for each entity
Entity
Entity
Entity
Entity
Entity
...
Evaluation scores
0.74
0.92
0.39
0.55
0.69
...

Figure 2: The Process of The Proposed Method

(1) We judge whether the sentence is surely related to the evaluation with the given evaluation aspect or not. If yes, the sentence is called an *evaluation sentence* under an aspect a.

(2) For each *evaluation sentence* under a, we further add the polarity, i.e., *positive*, *neutral*, or *negative*, to each sentence.

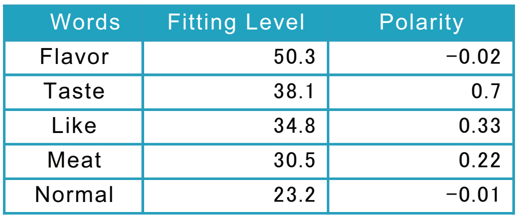
From this manipulated data set with human annotations, the dictionary is constructed. The method to construct the dictionary is described in Section 2.2.

Facing on (b)(c) computing evaluation scores, our basic strategy is to first compute the score for each sentence (not for each article) based on the dictionary, and collect them to compute the score for each entity. We adopt this strategy because we expect the averaging effect; It is important to collect a sufficient number of units for evaluation to perform statistical computation, so we choose a "sentence" as a unit of evaluation because a relatively large number of sentences are included in each evaluation article, while in many cases each sentence includes sufficient number of words to judge their polarity roughly. Specifically, in our method, we first compute a polarity value, i.e., *positive*, *neutral*, or *negative* for each *evaluation sentence* using the dictionary, and merge them to compute finer-grained rating score for each entity according to the ratio of *positive* and *negative* sentences.

2.2 (a) Learning A Dictionary of Evaluation Words

2.2.1 Structure of The Dictionary

The evaluation dictionary is a dictionary that is used to compute the polarity of each sentence in review articles, and is a set of tuples $(w, F_{w,a}, P_{w,a})$, where w is a word for evaluation, $F_{w,a}$ is the *fitness level* of the word w with aspect a, and



Words	Fitting Level	Polarity
Flavor	50.3	-0.02
Taste	38.1	0.7
Like	34.8	0.33
Meat	30.5	0.22
Normal	23.2	-0.01

Figure 3: An Example of Evaluation Dictionary

$P_{w,a}$ is the *polarity level* of w with aspect a. The *fitness level* $F_{w,a}$ represents the degree how much w is important in evaluating sentences w.r.t. an aspect a, and takes a higher value if the importance is higher. The *polarity level* $P_{w,a}$ represents the degree of *positive* or *negative* feeling of the word in evaluation w.r.t. a, which takes a value in $[1, -1]$ such that $P_{w,a}$ is nearer to 1 when w gives more positive evaluation, and nearer to -1 in case of more negative evaluation.

2.2.2 Retrieving Words for Evaluation

To construct the dictionary, we first retrieve the evaluation words, which are the words that we use in evaluation, from the data set for learning. We construct the dictionary with the words retrieved as evaluation words from the data set, while other words in the data set are just ignored. To retrieve evaluation words, we apply the morphological analysis to the data set and as evaluation words we choose nouns, verbs, adjectives, adverbs, interjections, and symbols.

In our method, to judge polarity correctly, a small preprocessing of words is required. Specifically, the negate words such as “not” and “never” in English would reverse the polarity of evaluation words. Thus, if we find these negate words with a verb or a adjective, we treat the verb or the adjective as a new word that includes negative meaning. Namely, one verb or adjective word may be included in the dictionary as two different words, i.e., with and without negative meaning.

2.2.3 Computing Fitness Levels of Words

As described above, the *fitness level* $F_{w,a}$ is the real value that represents how important a word w is in evaluation w.r.t. an aspect a. We designed a formula to compute the *fitness level* based on the well-known *tf-idf* index. The *tf-idf* is an index value that takes high value for words peculiar to a given document; For a given document included in a document set, *td-idf* is the product of *tf* and *idf*, where *tf* is the *term frequency* that represents the frequency of the term (word) in the document, and *idf* is the *inverse document frequency* that represents how common the term appears in all documents in the document set. Namely, the *tf-idf* index takes higher value for the words peculiar to the document, while it takes lower value for the words that commonly appears in all documents.

In the proposed method, as the value corresponding to *tf*, we use the frequency of a word w in the *evaluation sentences* under an aspect a i.e., the number of the sentences that includes w among *evaluation sentences* under aspect a in the learning data set. On the other side, as the value corresponding to *idf*, we use the ratio of sentences including w among all the sentences in the learning data set. Thus, *idf* takes larger value when the number of sentences including w is small.

Now we give a formal description of the *fitness level*. Let S be the set of all sentences in the learning data set, S_a be the *evaluation sentence*, i.e., the set of sentences judged to be related to the aspect a, and $S_{\bar{a}}$ be those judged not to be related to a. Naturally, $S_a \cap S_{\bar{a}} = \emptyset$ and $S = S_a \cup S_{\bar{a}}$ hold. Also, let $n_{w,a}$ and $n_{w,\bar{a}}$ be the frequency of w in S_a and $S_{\bar{a}}$, respectively. Let $|\{s|w \in s \text{ and } s \in S\}|$ be the number of sentences that include w in S. Then, the definition of $F_{w,a}$ for a given word w and an aspect a is written as follows:

$$F_{w,a} = \mathrm{tf}_{w,a} \cdot \mathrm{idf}_w,$$

where

$$\mathrm{tf}_{w,a} = \frac{n_{w,a}}{n_{w,a} + n_{w,\bar{a}}},$$

$$\mathrm{idf}_w = \log \frac{|S|}{|\{s|w \in s \text{ and } s \in S\}|}.$$

2.2.4 Computing Polarity Levels of Words

Polarity level $P_{w,a}$ is the real value in range $[1, -1]$ that represents the degree of positive or negative that a word w is used to evaluate entities under an aspect a. $P_{w,a}$ takes a value near 1 when w contributes to positive evaluation, and near -1 when negative.

The polarity level of a word w with an aspect a is computed based on the ratio of positive and negative sentences among all *evaluation sentences* that includes w. We designed the formula to compute $P_{w,a}$ where the polarity takes 1 when w appears in only positive sentences, and takes -1 when w appears in only negative ones.

The formal definition of $P_{w,a}$ is given in the following. Let S_a^p and S_a^n be the sets of *evaluation sentences* in S_a that are annotated as *positive* and *negative*, respectively. Also, let $f_{w,a}^p$ and $f_{w,a}^n$ be the frequency of w appearing in the sentences in S_a^p and S_a^n, respectively. Then, the polarity level $P_{w,a}$ for a word w and an aspect a is defined as follows:

$$P_{w,a} = \frac{f_{w,a}^p}{f_{w,a}^p + f_{w,a}^n} - \frac{f_{w,a}^n}{f_{w,a}^p + f_{w,a}^n}$$

2.3 (b) Computing Polarities for Sentences

2.3.1 Retrieving Evaluation Sentences under Aspect a

For each sentence in the evaluation articles in the review site, we first judge whether the sentence should be used to compute the score of the entity, i.e., whether each sentence in review articles is *evaluation sentence* or not. The *evaluation sentence* should surely evaluate the entity under the aspect a. Thus, in this process, we judge this point using the fitting levels of the words included in the sentence.

The basic strategy is as follows. From a sentence s to be judged, we first retrieve words whose fitting level is sufficiently high, which are the words that have ability to evaluate entities. We next compute the average of the fitting levels, and if the average is sufficiently high, the sentence s has ability to evaluate entities, so judged to be *evaluation sentence*.

We present the formal description of this process. Let s be the sentence to be judged. Let F_{min} be the threshold of fitting level for *evaluation words*. With threshold F_{min}, we define the set of words that has sufficiently high fitting levels as $W_f^s = \{w|w \in s \text{ and } F_{w,a} \geq F_{min}\}$. Thus, the average of the fitting levels of *evaluation words* in s is written as

$$F_s = \frac{1}{|W_f^s|}\Sigma_{w \in W_f^s} F_{w,a}.$$

If F_s is equal to or larger than threshold T_s, i.e., $F_s \geq T_s$, then the sentence s is judged to be *evaluation sentence*, which is used in evaluating entities.

Fig. 4 illustrates an example of the process to choose the *evaluation sentences* for entities. In this figure, we judge whether the sentence is an *evaluation sentence* or not under an aspect a. Here, the fitting levels of all (four) words used for evaluation are larger than threshold F_{min}, we compute the average of the fitting levels among them. Because the average value is larger than threshold $T_s = 10$, this sentence is selected as an *evaluation sentence*.

2.3.2 Computing Polarities

For the each *evaluation sentences* s, we further compute the polarity of the sentence s. Since a single sentence usually includes not many words, we choose to use the 3-graded polarity value, i.e., *positive*, *neutral*, and *negative*, rather than finer-grained polarity such as real values.

The basic strategy to compute the polarity of sentence s is to examine the total polarity of the evaluation words included in s. We first retrieve the words that has sufficiently strong polarity, and examine the total balance of the polarity of them.

Specifically, let P_{min} be the threshold to select words of strong polarity. With P_{min}, we define the set of words that has strong polarity as $W_p^s = \{w|w \in s \text{ and } |P_{w,a}| \geq P_{min}\}$. Using this set of words, we define the polarity of sentence s as follows:

$$P_s = \begin{cases} \textit{positive}, & (\text{if } \; T_p < \frac{1}{|W_p^s|}\Sigma_{w \in W_p^s} P_{w,a}), \\ \textit{neutral}, & (\text{if } \; -T_p \leq \frac{1}{|W_p^s|}\Sigma_{w \in W_p^s} P_{w,a} \leq T_p), \\ \textit{negative}, & (\text{if } \; \frac{1}{|W_p^s|}\Sigma_{w \in W_p^s} P_{w,a} < T_p). \end{cases}$$

Fig. 5 illustrates an example of polarity computation of a sentence. Since the sentence is determined as *evaluation sentence* in Fig. 4, we next compute the polarity of this sentence. We first retrieve the words whose polarity values are equal to or larger than P_{min} in absolute value, and compute the average of the polarity of the selected words. Because the average is larger than threshold $T_p = 0.35$, the polarity of this sentence is determined to be *positive*.

2.4 (c) Computing Evaluation Scores for Entities

Finally, we compute the evaluation score of an entity i using the *evaluation sentences* selected under aspect a. The evaluation score of i, which is referred as $Score(i)$, is computed based on the ratio of *positive* and *negative* evaluation sentences in the review articles of i. Formally, the evaluation score is computed as

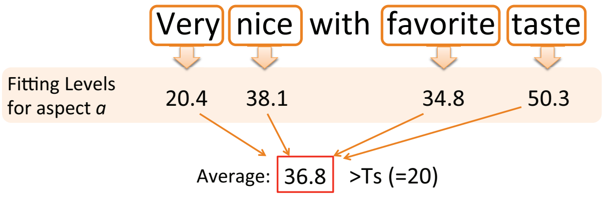


Figure 4: Judging Aspect for An Evaluation Sentence

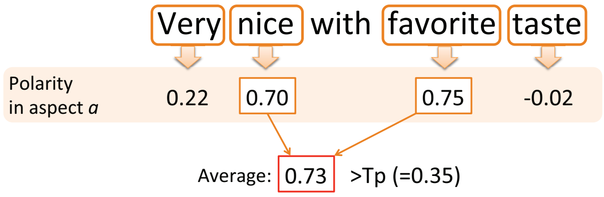


Figure 5: Judging Polarity for An Evaluation Sentence

$$Score(i) = \frac{|\{s|P_s = \textit{positive} \text{ and } s \in E_i\}|}{|\{s|P_s = \{\textit{positive} \text{ or } \textit{negative}\} \;\text{ and } s \in E_i\}|}$$

where E_i represents the set of *evaluation sentences* that evaluate i selected with the process shown in Sec. 2.3.1, and s represents a sentence.

3 Evaluation

3.1 The Viewpoints

In this paper, we propose a method to compute the evaluation scores for each entity with respect to an arbitrary text description of evaluation aspects. In other words, this method intends to predict the human evaluation scores for each entity that readers of the evaluation articles would make. Therefore, in our evaluation, we requested several persons to read evaluation articles and make a 10-grade score for each entity. We evaluated the difference between the human scores and the computed scores to measure the precision of the proposed method.

Note that, however, human scores in general vary depending on individuals, especially in the average or the standard deviation of the scores. (Imagine that some person may make relatively low scores in average, while other person may prefer high rating.) To take this diversity into account, we standardized the human scores for each person (namely, the average and the standard deviation of the scores made by a person are adjusted to be the same), and made a ranking of entities with their average scores. If the two rankings based on human scores and computed scores are similar to each other, it implies that the performance of the proposed method to predict human scores is good. Thus, we used Spearman's rank correlation coefficient between human and computed rankings as evaluation criterion to measure the precision of the proposed method.

We conducted two evaluations using different aspects. We supposed the following two different cases in determining the

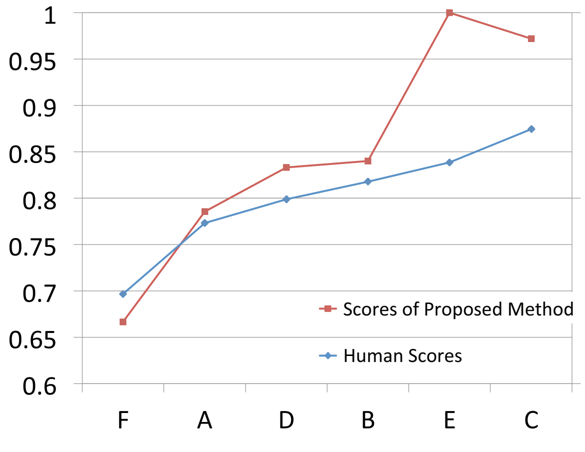


Figure 6: Human and Computed Scores for Restaurants under Aspect “Taste” (Experiment 1)

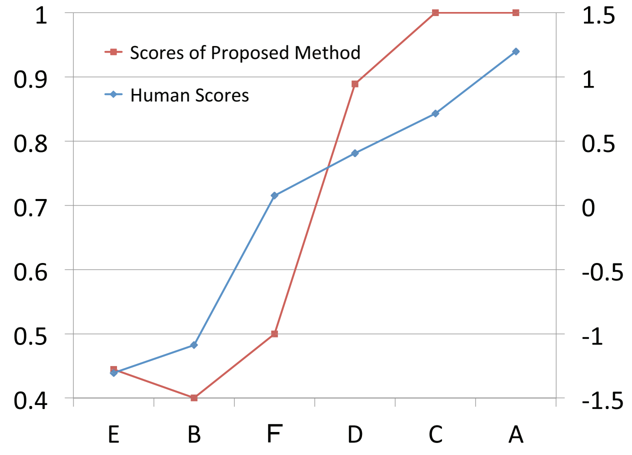


Figure 7: Human and Computed Scores for Restaurants under Aspect “Price” (Experiment 1)

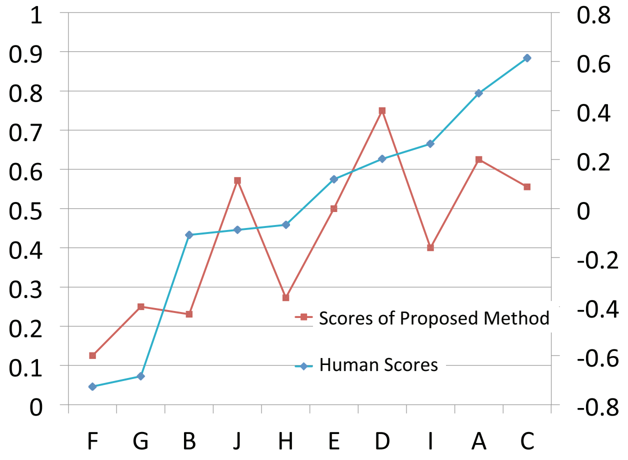


Figure 8: Human and Computed Scores for Ra-men Restaurants under Aspect “Quality of Noodles” (Experiment 2)

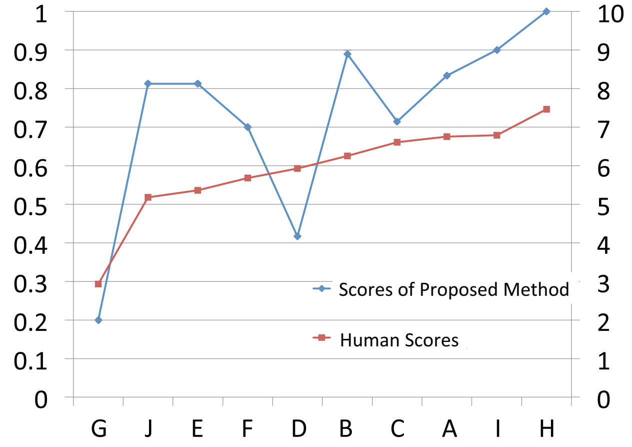


Figure 9: Human and Computed Scores for Ra-men Restaurants under Aspect “Taste of Soup” (Experiment 2)

aspects.

Experiment 1: In case of general evaluation aspects.

Experiment 2: In case of specific evaluation aspects that reflects on personal viewpoints of individuals.

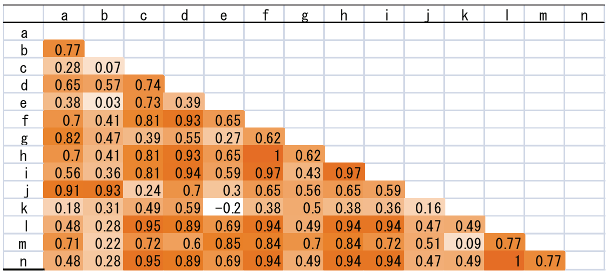
In Experiment 1, we used general evaluation aspects that we often see in review sites. We selected “restaurants” as evaluation entities, and used two evaluation aspects “taste” and “price.” In Experiment 2, we used a little specific evaluation aspects that requires several words to describe. We selected “Ra-men restaurants” as evaluation entities, and used two evaluation aspects “quality of noodles” and “taste of soup.”

3.2 Evaluation Methods

For Experiment 1, we selected 6 restaurants as the reviewed entities from a popular Japanese review site called “Tabelog” [11]. To guarantee fair evaluation, these 6 restaurants are chosen from high-rated restaurants in Tabelog, placed in Tokyo, where users’ ratings are the same as a whole. We selected 3 review articles for each restaurants mainly under the criteria that (1) the length is almost the same as 50-60 sentences, (2) review date is not old, (3) they do not include any direct description of numerical scores, and (4) sentences are relatively tidy. As written above, the evaluation aspects are “taste” and “price,” and we told the participants of our experiments (i.e., subjects in our experiments) that “taste” means how good the taste of dishes is, and “price” means how reasonable the price of dishes is.

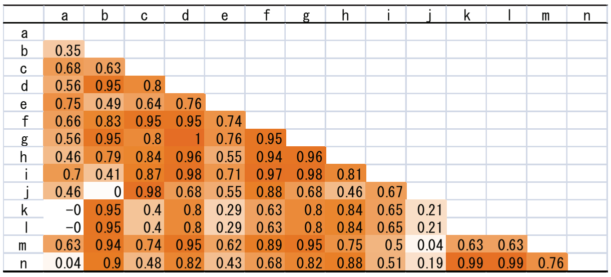
For Experiment 2, as the reviewed entities, we selected 10 Japanese Ra-men restaurants placed in Wakayama city also from Tabelog. Note that they are all high-rated Ra-men restaurants in Wakayama city. We used 7 review articles for each restaurant, where each article consists of about 30 sentences. The criteria to select those review articles are the same as the case of Experiment 1. The evaluation aspects are “quality of noodles” and “taste of soup.”

In advance of the experiments, we constructed the dictionary under the given four aspects. As a set of learning data, we collected review articles from Tabelog. For Experiment 1, we collected 1,500 review articles for restaurants that include about 6,000 sentences. As the result of our annotation, we obtained 4,600 evaluation words for the aspect “taste,” and 1,500



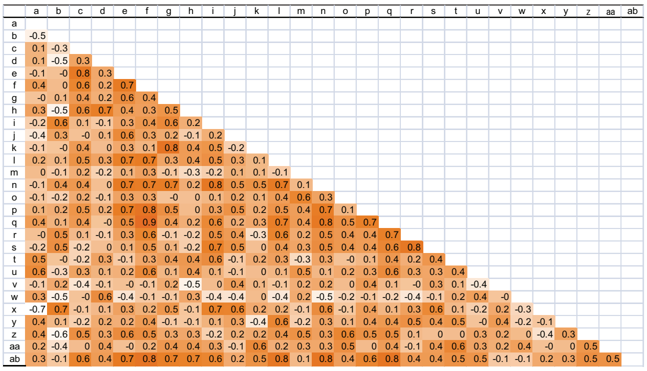
	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a														
b	0.77													
c	0.28	0.07												
d	0.65	0.57	0.74											
e	0.38	0.03	0.73	0.39										
f	0.7	0.41	0.81	0.93	0.65									
g	0.82	0.47	0.39	0.55	0.27	0.62								
h	0.7	0.41	0.81	0.93	0.65	1	0.62							
i	0.56	0.36	0.81	0.94	0.59	0.97	0.43	0.97						
j	0.91	0.93	0.24	0.7	0.3	0.65	0.56	0.65	0.59					
k	0.18	0.31	0.49	0.59	−0.2	0.38	0.5	0.38	0.36	0.16				
l	0.48	0.28	0.95	0.89	0.69	0.94	0.49	0.94	0.94	0.47	0.49			
m	0.71	0.22	0.72	0.6	0.85	0.84	0.7	0.84	0.72	0.51	0.09	0.77		
n	0.48	0.28	0.95	0.89	0.69	0.94	0.49	0.94	0.94	0.47	0.49	1	0.77	

Figure 10: Rank Correlation Coefficients between Users. (Aspect “Taste” in Experiment 1)



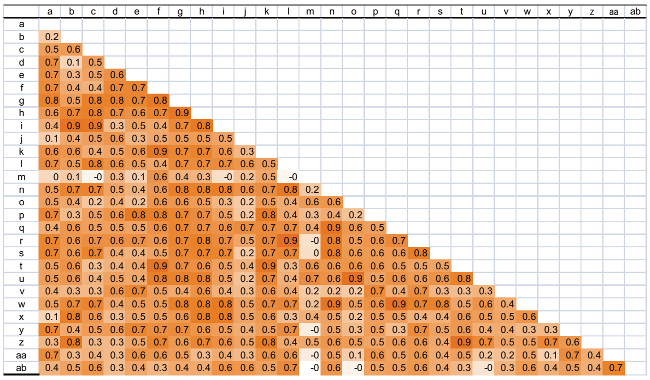
	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a														
b	0.35													
c	0.68	0.63												
d	0.56	0.95	0.8											
e	0.75	0.49	0.64	0.76										
f	0.66	0.83	0.95	0.95	0.74									
g	0.56	0.95	0.8	1	0.76	0.95								
h	0.46	0.79	0.84	0.96	0.55	0.94	0.96							
i	0.7	0.41	0.87	0.98	0.71	0.97	0.98	0.81						
j	0.46	0	0.98	0.68	0.55	0.88	0.68	0.46	0.67					
k	−0	0.95	0.4	0.8	0.29	0.63	0.8	0.84	0.65	0.21				
l	−0	0.95	0.4	0.8	0.29	0.63	0.8	0.84	0.65	0.21				
m	0.63	0.94	0.74	0.95	0.62	0.89	0.95	0.75	0.5	0.04	0.63	0.63		
n	0.04	0.9	0.48	0.82	0.43	0.68	0.82	0.88	0.51	0.19	0.99	0.99	0.76	

Figure 11: Rank Correlation Coefficients between Users. (Aspect “Price” in Experiment 1)



	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab
a																												
b	-0.5																											
c	0.1	-0.3																										
d	0.1	-0.5	0.3																									
e	-0.1	-0	0.8	0.3																								
f	0.4	0	0.6	0.2	0.7																							
g	-0	0.1	0.4	0.2	0.6	0.4																						
h	0.3	-0.5	0.6	0.7	0.4	0.3	0.5																					
i	-0.2	0.6	0.1	-0.1	0.3	0.4	0.6	0.2																				
j	-0.4	0.3	-0	0.1	0.6	0.3	0.2	-0.1	0.2																			
k	-0.1	-0	0.4	0	0.3	0.1	0.8	0.4	0.5	-0.2																		
l	0.2	0.1	0.5	0.3	0.7	0.7	0.3	0.4	0.5	0.3	0.1																	
m	0	-0.1	0.2	-0.2	0.1	0.3	-0.1	-0.3	-0.2	0.1	0.1	-0.1																
n	-0.1	0.4	0.4	0	0.7	0.7	0.7	0.2	0.8	0.5	0.5	0.7	0.1															
o	-0.1	-0.2	0.2	-0.1	0.3	0.3	-0	0	0.1	0.2	0.1	0.4	0.6	0.3														
p	0.1	0.2	0.5	0.2	0.7	0.8	0.5	0	0.3	0.5	0.2	0.5	0.4	0.7	0.1													
q	0.4	0.1	0.4	-0	0.5	0.9	0.4	0.2	0.6	0.2	0.3	0.7	0.4	0.8	0.5	0.7												
r	-0	0.5	0.1	-0.1	0.3	0.6	-0.1	-0.2	0.5	0.4	-0.3	0.6	0.2	0.5	0.4	0.4	0.7											
s	-0.2	0.5	-0.2	0	0.1	0.5	0.1	-0.2	0.7	0.5	0	0.4	0.3	0.5	0.4	0.4	0.6	0.8										
t	0.5	-0	-0.2	0.3	-0.1	0.3	0.4	0.4	0.6	-0.1	0.2	0.3	-0.3	0.3	-0	0.1	0.4	0.2	0.4									
u	0.6	-0.3	0.3	0.1	0.2	0.6	0.1	0.4	0.1	-0.1	0	0.1	0.5	0.1	0.2	0.3	0.6	0.3	0.3	0.4								
v	-0.1	0.2	-0.4	-0.1	-0	-0.1	0.2	-0.5	0	0.4	0.1	-0.1	0.2	0.2	0	0.4	0.1	-0	0.3	0.1	-0.4							
w	0.3	-0.5	-0	0.6	-0.4	-0.1	-0.1	0.3	-0.4	-0.4	0	-0.4	0.2	-0.5	-0.2	-0.1	-0.2	-0.4	-0.1	0.2	0.4	-0						
x	-0.7	0.7	-0.1	0.1	0.3	0.2	0.5	-0.1	0.7	0.6	0.2	0.2	-0.1	0.6	-0.1	0.4	0.1	0.3	0.6	0.1	-0.2	0.2	-0.3					
y	0.4	0.1	-0.2	0.2	0.2	0.4	-0.1	-0.1	0.1	0.3	-0.4	0.6	-0.2	0.3	0.1	0.4	0.4	0.5	0.4	0.5	-0	0.4	-0.2	-0.1				
z	0.4	-0.6	0.5	0.3	0.6	0.5	0.3	0.3	-0.2	0.2	0.2	0.4	0.5	0.3	0.6	0.5	0.5	0.1	0	0	0.3	0.2	0	-0.4	0.3			
aa	0.2	-0.4	0	0.4	-0	0.2	0.4	0.4	0.3	-0.1	0.6	0.2	0.3	0.3	0.5	0	0.4	-0.1	0.4	0.6	0.3	0.2	0.4	-0	0	0.5		
ab	0.3	-0.1	0.6	0.4	0.7	0.8	0.7	0.7	0.6	0.2	0.5	0.8	0.1	0.8	0.4	0.6	0.8	0.4	0.4	0.5	0.5	-0.1	-0.1	0.2	0.3	0.5	0.5	

Figure 12: Rank Correlation Coefficients between Users. (Aspect “Quality of Noodles” in Experiment 2)



	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab
a																												
b	0.2																											
c	0.5	0.6																										
d	0.7	0.1	0.5																									
e	0.7	0.3	0.5	0.6																								
f	0.7	0.4	0.4	0.7	0.7																							
g	0.8	0.5	0.8	0.8	0.7	0.8																						
h	0.6	0.7	0.8	0.7	0.6	0.7	0.9																					
i	0.4	0.9	0.9	0.3	0.5	0.4	0.7	0.8																				
j	0.1	0.4	0.5	0.6	0.3	0.5	0.5	0.5	0.5																			
k	0.6	0.6	0.4	0.5	0.6	0.9	0.7	0.7	0.6	0.3																		
l	0.7	0.5	0.8	0.6	0.5	0.4	0.7	0.7	0.7	0.6	0.5																	
m	0	0.1	-0	0.3	0.1	0.6	0.4	0.3	-0	0.2	0.5	-0																
n	0.5	0.7	0.7	0.5	0.4	0.6	0.8	0.8	0.8	0.6	0.7	0.8	0.2															
o	0.5	0.4	0.2	0.4	0.2	0.6	0.6	0.5	0.3	0.2	0.5	0.4	0.6	0.6														
p	0.7	0.3	0.5	0.6	0.8	0.8	0.7	0.7	0.5	0.2	0.8	0.4	0.3	0.4	0.2													
q	0.4	0.6	0.5	0.5	0.5	0.6	0.7	0.7	0.6	0.7	0.7	0.7	0.4	0.9	0.6	0.5												
r	0.7	0.6	0.7	0.6	0.7	0.6	0.7	0.8	0.7	0.5	0.7	0.9	-0	0.8	0.5	0.6	0.7											
s	0.7	0.6	0.7	0.4	0.4	0.5	0.7	0.7	0.7	0.2	0.7	0.7	0	0.8	0.6	0.6	0.6	0.8										
t	0.5	0.6	0.3	0.4	0.4	0.9	0.7	0.6	0.5	0.4	0.9	0.3	0.6	0.6	0.6	0.6	0.5	0.5	0.5									
u	0.5	0.6	0.4	0.5	0.4	0.8	0.8	0.8	0.5	0.2	0.7	0.4	0.7	0.6	0.9	0.5	0.6	0.6	0.6	0.8								
v	0.4	0.3	0.3	0.6	0.7	0.5	0.4	0.6	0.4	0.3	0.6	0.4	0.2	0.2	0.2	0.7	0.4	0.7	0.3	0.3	0.3							
w	0.5	0.7	0.7	0.4	0.5	0.5	0.8	0.8	0.8	0.5	0.7	0.7	0.2	0.9	0.5	0.6	0.9	0.7	0.8	0.5	0.6	0.4						
x	0.1	0.8	0.6	0.3	0.5	0.5	0.6	0.8	0.8	0.5	0.6	0.3	0.4	0.5	0.2	0.5	0.5	0.4	0.4	0.6	0.5	0.5	0.6					
y	0.7	0.4	0.5	0.6	0.7	0.7	0.7	0.6	0.5	0.4	0.5	0.7	-0	0.5	0.3	0.5	0.3	0.7	0.5	0.6	0.4	0.4	0.3	0.3				
z	0.3	0.8	0.3	0.3	0.5	0.7	0.6	0.7	0.6	0.5	0.8	0.4	0.5	0.6	0.5	0.5	0.6	0.6	0.4	0.9	0.7	0.5	0.5	0.7	0.6			
aa	0.7	0.3	0.4	0.3	0.6	0.6	0.5	0.3	0.4	0.3	0.6	0.6	-0	0.5	0.1	0.6	0.5	0.6	0.4	0.5	0.2	0.2	0.5	0.1	0.7	0.4		
ab	0.4	0.5	0.6	0.3	0.4	0.3	0.4	0.4	0.6	0.6	0.5	0.7	-0	0.6	-0	0.5	0.5	0.6	0.4	0.3	-0	0.3	0.6	0.4	0.5	0.4	0.7	

Figure 13: Rank Correlation Coefficients between Users. (Aspect “Taste of Soup” in Experiment 2)

words for “price.” For Experiment 2, we collected 1,200 review articles for Ra-men restaurants that include about 3,000 sentences. As the result of annotation, we obtained 800 evaluation words for “quality of noodles,” and 1,800 words for “taste of soup.”

As the process of the experiments, we asked all the participants to read all the review articles and to make a 10-grade score for each entity, where 10 is the best, and 1 is the worst grade of scores. 14 and 28 persons (subjects) participated to the Experiment 1 and 2, respectively, where all the participants were at the age of 20’s.

3.3 Evaluation Results

In Figs. 6-9, we show the average of human scores and the scores computed by the proposed method. The horizontal axis represents the restaurants in the order of scores. The left vertical axis represents the score of the proposed method, and the right one represents the average of human scores. The rank correlation coefficient in Experiment 1 is 0.94 for the aspect “taste” and 0.92 for “price,” which show that the proposed method predicts the human score with high precision in case of general aspects. In Experiment 2, the rank correlation coefficient is 0.74 for “quality of noodles” and 0.72 for “taste of soup,” which is not so high as Experiment 1, but relatively high value.

Figs. 10-13 shows the rank correlation coefficients between participants for each aspects in our experiments. Each alphabet represents a participant (14 and 28 persons participated in our Experiments 1 and 2, respectively), and for every pairs of the participants, we compute the rank correlation coefficient of the two rankings. In Experiment 1, the correlation coefficients takes high values as a whole, where the average value is 0.60 for “taste” and 0.67 for “price,” meaning that the ranking of participants are relatively similar to each other. On the other hand, in Experiment 2, they take low values where the average value is 0.22 for “quality of noodles” and 0.52 for “taste of soup,” meaning that the ranking differs significantly according to individuals.

3.4 Discussion

We obtained the result that the rank correlation coefficients take relatively high values in case of general aspects, while they take low values in case of specific aspects. In this section, we discuss the reason of this point.

First of all, we focus on the rank correlation coefficient between participants shown in Figs. 10-13, where they take quite low values in case of specific aspects. Especially, with the aspect “quality of noodles,” it takes very low value 0.22. It means that the rankings of participants are basically similar to each other for the general aspects in Experiment 1, whereas they are quite different for the specific aspects in Experiment 2. The reason of this is quite simple; It is due to likes and dislikes among people. In fact, from the hearing from participants after the Experiment 2, it is clarified that several participants were strongly affected by the expression words such as “hard” or “soft” on noodles, or “rich” or “plain” on soup. It would be natural that someone likes "hard" noodle or "rich" soup, while others would like "soft" or "plain" ones. On this point, we also examined the polarities of those words in the dictionary and found that the polarities of them are mostly neutral (i.e., near 0). It is considered that the person who annotated to the learning data set seemed to select *neutral* if a sentence includes the words that depends on like and dislikes of people. As a result, the proposed method also gave neutral polarities for this kind of words.

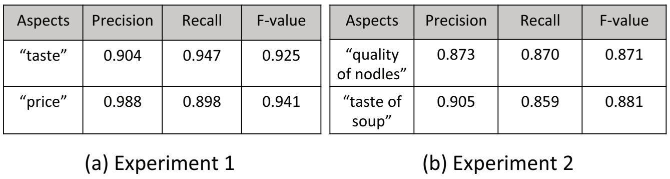
As another reason on this point, the precision of the dictionary possibly affects the performance in Experiment 2. Note that the number of words in Experiment 1 is 4,600 words for "taste" and 1,500 words for "price," while that in Experiment 2 is 800 words for "quality of noodle" and 1,800 words for "taste of soup." The number of words is smaller in Experiment 2, which may affects the performance. (Note that the performance for "price" is good although the number of words is relatively small. This may be because most of the words that evaluates "price" is clear to understand; the polarity of words "expensive" or "cheap" would be clear for most of people.)

Therefore, to examine the effects of the number of words in the dictionary, we conducted another experiment. Using the four dictionaries constructed for each evaluation aspects as used in Experiments 1 and 2, we evaluated the precision of evaluation-sentence judgments and polarity judgments for sentences described in Sections 2.3.1 and 2.3.2, respectively. For evaluation-sentence evaluation, we prepared 1,200 sentences that are related with the aspect and another 1,200 sentences not related with the aspect, and examined the precision of the proposed judging algorithm described in Section 2.3.1. For polarity evaluation, we prepared 1,200 sentences for each of *positive*, *neutral*, and *negative* polarities, and examined the precision of the proposed method described in Section 2.3.2. Results are shown in Figs. 14 and 15, respectively. Both results show that the proposed method marks about 90% of precisions regardless of aspects, which indicates that the precision of the dictionary is not related to the number of words in the dictionary. Thus, the cause that the rank correlation coefficient is relatively low in Experiment 2 would not be the number of words in the dictionary, but would be the effect of like and dislike of people for several specific evaluation words.

4 Conclusion

In this paper, we proposed a method that computes the evaluation scores for entities in review sites with respect to a given aspect of arbitrary text description. As a result of our evaluation, the proposed method computes evaluation scores that have high rank correlation coefficients with human scores. That is to say, the proposed method predicts the human scores with high precision. Also, from the evaluation result, we found there are aspects that include likes and dislikes of people, and that the correlation coefficients degrade for such aspects. The main reason of this degradation is the existence of the evaluation words for which people may give wide-range of polarities depending on persons.

As future work, it would be practical to develop a method to judge the evaluation words with polarity variation, which may lead more precise prediction of human evaluation scores. As another problem, our method requires a learning data set with human annotations. To develop a method that require less human labors would be more practical and desirable.



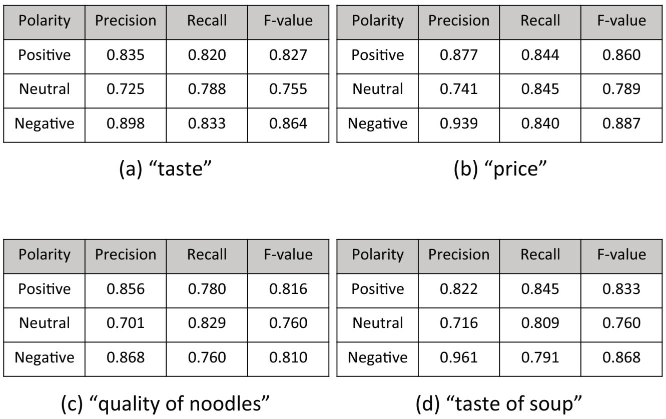
Aspects	Precision	Recall	F-value
"taste"	0.904	0.947	0.925
"price"	0.988	0.898	0.941

(a) Experiment 1

Aspects	Precision	Recall	F-value
"quality of nodles"	0.873	0.870	0.871
"taste of soup"	0.905	0.859	0.881

(b) Experiment 2

Figure 14: Accuracy of Aspect Judgments



Polarity	Precision	Recall	F-value
Positive	0.835	0.820	0.827
Neutral	0.725	0.788	0.755
Negative	0.898	0.833	0.864

(a) "taste"

Polarity	Precision	Recall	F-value
Positive	0.877	0.844	0.860
Neutral	0.741	0.845	0.789
Negative	0.939	0.840	0.887

(b) "price"

Polarity	Precision	Recall	F-value
Positive	0.856	0.780	0.816
Neutral	0.701	0.829	0.760
Negative	0.868	0.760	0.810

(c) "quality of noodles"

Polarity	Precision	Recall	F-value
Positive	0.822	0.845	0.833
Neutral	0.716	0.809	0.760
Negative	0.961	0.791	0.868

(d) "taste of soup"

Figure 15: Accuracy of Polarity Judgments

REFERENCES

[1] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL), pp. 417-424, 2002.

[2] M. Koppel and J. Schler, "The Importance of Neutral Examples in Learning Sentiment." Computational Intelligence, Vol.22, No.2, pp.100-109, 2006.

[3] B. Pang, and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respace to Rating Scales," In Proceeding of the 43rd Meeting of the Association for Computational Linguistics (ACL), pp. 115-124, 2005.

[4] D. Okanohara and J. Tsujii, "Assigning Polarity Scores to Reviews Using Machine Learning Techniques," Natural Language Processing, Vol.14, No.3, 2007.

[5] G. Carenini, R. Ng, and A. Pauls, "Multi-Document Summarization of Evaluative Text," In Prof. of the conference of the Eutopean chapter of the association for computational linguistics, 2006.

[6] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," In Proc. of Nineteenth National Conference on Artificial Intelligence, 2004.

[7] N. Jindal and B. Liu, "Identifying Comparative Sen-

tences in Text Documents," In Proc. of the 29th annual international ACM conference on Research and development in information retrieval (SIGIR), pp.244-251, 2006.

[8] G. Carenini, R. Ng, and E. Zwart, "Extracting Knowledge from Evaluative Text," In Proc. of the 3rd international conference on knowledge capture, pp.11-18, 2005.

[9] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," In Proc. of the 2004 ACM international conference on knowledge discovery and data mining (SIGKDD), pp.168–177, 2004.

[10] I. Titov, R. McDonald, "Modeling Online Reviews with Multi-grain Topic Models," In Proc. of the 17th international conference on World Wide Web (WWW), Pages 111-120, 2008.

[11] Tabelog, http://tabelog.com/ . (In Japanese)

Damage-resilient Network Services
- Trusted Cloud Computing -

Norio Shiratori [*1], Storu Izumi [*2], Takuo Suganuma [*2,*3],
Shinji Kitagami [*4], Yotaro Miyanishi[*5], and Yoshitaka Shibata [*6]

[*1]GITS, Waseda University, Japan / RIEC, Tohoku University, Japan
[*2]Cyberscience Center, Tohoku University, Japan
[*3]Graduate School of Information Sciences, Tohoku University, Japan
[*4]GITS, Waseda University, Japan [*5]ISEM, Inc., Japan
[*6]Faculty of Software and Information Science, Iwate Prefectural University, Japan
norio@shiratori.riec.tohoku.ac.jp

Abstract –
This is a vision paper on network computing towards a sustainable symbiotic information society. This research is partially supported by JSPS Kakenhi-Kiban (A) - 26240012 (2014-2016). In this paper, we propose and develop a basic methodology for damage-resilient network services by effectively amalgamating the two technologies, namely (1) Never Die Networks and (2) Trusted Cloud Computing.

Regarding (1), the basic concept of Never Die Networks originally proposed in 2003, is based on the following three features: 1) Real-time services, 2) Ability to support massively large number of simultaneous users, and 3) Offer such service for a short duration (at the very least 20 seconds, 30 seconds).

We propose the architecture for Green-oriented Never Die Networks as shown in Fig. 4. We also aim the Green-MIB (Management Information Base) to be standardized in IETF.

Regarding (2), we propose a new security method from the practical point of view to increase user's sense of safety in cloud services. The basic idea of the proposed method of security for users' information such as data, programs, etc., is based on distributed computing by using hybrid cloud systems. We are now promoting an implementation of a prototype system by amalgamating the above technologies.

Keywords: Natural disaster, Never Die Networks, Interplanetary internet, Cloud computing, and Security.

1 TOWARDS SUSTAINABLE SYMBIOTIC INFORMATION SOCIETY

This is a vision paper on network computing towards a sustainable symbiotic information society. This research is partially supported by JSPS Kakenhi-Kiban (A) - 26240012 (2014-2016).

1.1 Global Warming, Natural Disaster and Aging Society

In science and technology of the 21st century, how to face global environmental change such as (1) global warming, (2) natural disaster and (3) aging society has come into question. Regarding (1), we first explain problems of modern information societies based on the rationality which consists of high efficiency, cost effective and high function. And then, we discuss solutions to overcome these problems such as environmental contamination, global warming, and natural disaster towards the creation of a sustainable information society.

1.2 Problems of Modern (Industrial) Society

Regarding information technology, in the initial stage the main concern was Quantity, whether it is data or the production of materials. But now we are looking for the Quality of the things. The same trend can be seen in these three fields as well (comp, network, society) as in Fig.1. When we will reach the Post Modern age, that is year 2020, from the recent Information society, in our words it will become a Symbiotic Society.

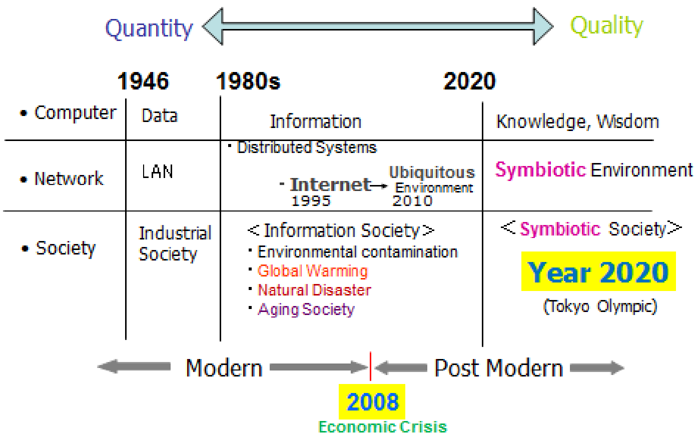
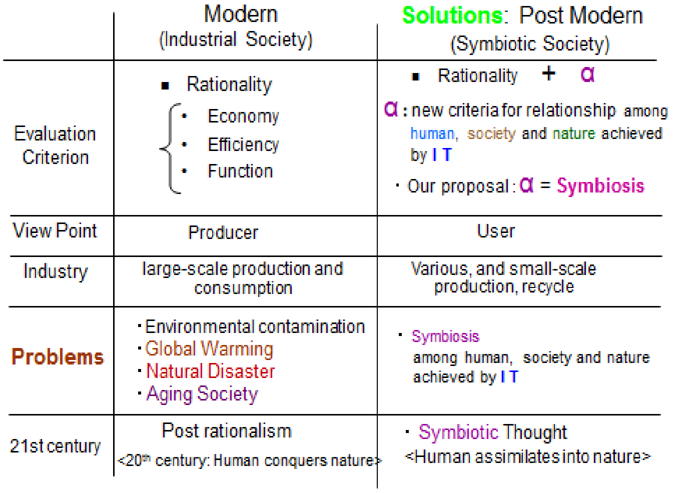


Fig. 1: Towards Sustainable Symbiotic Information Society (1).

Modern society is termed as Industrial Society, whereas Post Modern society can be termed as Knowledge / Wisdom society. In modern society, the evaluation criterion was Rationality, which consists of Economy, Efficiency and Function. These three elements are beneficial to our society in many ways: we have absolutely succeeded in obtaining wealth and affluence. However, at the same time, it caused the following problems such as environmental contamination, global warming, natural disaster and aging society as in Fig.2.

1.3 Solutions Towards Post Modern Society

To solve these problems, the research and development on green computing (energy-saving systems), disaster resilient-network and support systems for elderly people are actively promoted in the world.



	Modern (Industrial Society)	Solutions: Post Modern (Symbiotic Society)
Evaluation Criterion	▪ Rationality • Economy • Efficiency • Function	▪ Rationality + α α : new criteria for relationship among human, society and nature achieved by I T · Our proposal : α = Symbiosis
View Point	Producer	User
Industry	large-scale production and consumption	Various, and small-scale production, recycle
Problems	· Environmental contamination · Global Warming · Natural Disaster · Aging Society	· Symbiosis among human, society and nature achieved by I T
21st century	Post rationalism <20th century: Human conquers nature>	· Symbiotic Thought <Human assimilates into nature>

Fig. 2: Towards Sustainable Symbiotic Information Society (2).

We also need another evaluation criterion in addition to Rationality to overcome the problems. We denote it as 'alpha'. 'Alpha' is a new criterion for a relationship among human, society and nature achieved by IT. We propose a "symbiosis" as 'alpha' as in Fig.2. Symbiosis consists of cooperation, conflict and tension.

2 NEVER DIE NETWORKS

2.1 Personal Miserable Experiences of Two Big Earthquakes

The author has experienced the following two big earthquakes:

(A) June 12, 1978 Miyagi-oki Earthquake
(M7.4, Maximum Seismic Intensity 5)

(B) March 11, 2011 The Great East Japan Earthquake
(M 9.0, Maximum Seismic Intensity 7)

When Miyagi-oki Earthquake mentioned in above (A) occurred, I was in my laboratory at Tohoku University, the center of Sendai City. At that time, files, documents, books and other items flew off shelves and scattered throughout the room. The hard strike by the earthquake caused blackouts and then people could not establish telephone communications. On roads, Traffic lights were broken down and therefore just about everywhere, roads were jammed with people and cars. For a while after the earthquake, I could not confirm the safety of my gravid wife who supposed to stay at home 12 km away from my laboratory at Tohoku University. I realized that if there was a way to hear her voice only a few seconds, then I could know her safety. When people are rescued within the first 72 hours after the earthquake, their chance of survival will be increased: the sooner will be better to increase surviving rate of victims.

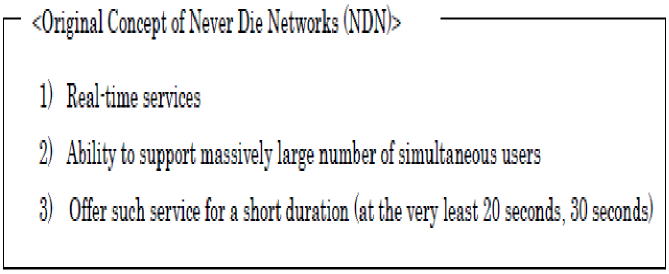
2.2 Year 1978 : The origin of NDN

From the miserable experience of Miyagi-oki Earthquake mentioned in 2.1, the author has keenly realized the necessity to develop networks named Never Die Networks (NDN) which never break down during the time of disaster and is based on the original concepts as shown in Fig.3. It was on June, 1978.

2.3 Proposal of Basic Concepts of NDN

(1) Original Basic Concept of NDN and Presentation of NDN Research

The ``original concept'' (Fig. 3) of NDN was conceived and developed from the author's frustrating experience of Miyagi-oki Earthquake in 1978, and in 2003, we have lunched research and development of NDN to realize our proposed 'original concept' as one function of flexible networks [32]. Specifically, as a concept of Never Die Networks, we aimed for the realization of the network which maximizes its system availability using resources effectively without halting the whole system during the malfunction of the information network system caused by disasters such as the concentration of user access, traffic congestion, and malfunctions of transmission paths and nodes. As shown in reference [7], functions of NDN were configured as functional modules of the Flexible Network Layer (FNL). The concept of NDN was realized through the combination of the acquisition function of QoS requests centered on NDN functions in QoS control unit, and the collection function of network information [8].



<Original Concept of Never Die Networks (NDN)>

1) Real-time services

2) Ability to support massively large number of simultaneous users

3) Offer such service for a short duration (at the very least 20 seconds, 30 seconds)

Fig. 3: The definition of Never Die Networks

(2) NDN Project Adopted to JSPS Grants-in-Aid for Scientific Research on Challenging Exploratory Research [30]

In 2007, our project, "Never Die Networks: Towards Networks Resilient to Worsened Network Environment" (Project Leader: Norio Shiratori, 3 years from 2007-2009) was adopted as the Grants-in-Aid for Scientific Research on Challenging Exploratory Research sponsored by the Japan Society for the Promotion of Science (JSPS). The proposed model and architecture of NDN are shown in [30]. Remarkable 2 points of the research are listed below:

<P1> The original characteristic of NDN research is an **invention of a new network control method to keep the system working without halting its whole system** in case of emergency. It will be achieved through the development and an effective fusion of technologies to collect and analyze network information in **"real time"** by utilizing the research achievements of the applicants in April 2006, the success in standardizing one of the NDN basic technologies in IETF.

<P2> The innovative point of NDN research is that in a system composed of multiple parts, if a part of the system is damaged in some reason, the NDN still continues working and maintaining the provision of its services without halting its whole system by **"creating the mechanisms of the autonomous cooperation and coordination" within its undamaged parts.**

2.4 Adoption of Grant-in-Aid for Scientific Research Kiban (A) [31]

Our research project "Green-oriented Never Die Networks to Adapt to Disaster Situation and Maximize Satisfaction Levels of Connectivity by Massively Large Number of Users" (2014~2017) is adopted as Grant-in-Aid for Scientific Research Kiban (A). The proposed network architecture is shown as Fig. 4. The overview of the project is as follows:

- Goal: Based on our painful experiences of the Great East Japan Earthquake, we learned if we can establish voice communication, and hear the voices of our family, friends, etc. in disaster areas only a few seconds, we are able to confirm their safety.

Our goal is making a contribution to solve the two big problems, disaster and global warming. Hence, we propose the "Green-oriented Never Die Networks Adaptive to Disaster Situation" in order to maximize satisfaction levels of connectivity for over a million users. We develop the following three technologies to confirm safety for 1,000 - 500,000 users per unit time at disaster with keeping minimize the power consumption of the network system.

(1) Spatiotemporal Segmentation Communication Protocol Adaptive to Disaster Situation

This protocol assigns the available access line to users flexibly according to the disaster situation for maximization of satisfaction levels of connectivity.

(2) Multi-layered Communication Network using Software Defined Network Adaptive to Disaster Situation

The network reconfigures the commination layer according to the disaster situation.

(3) Passive Green ICT Technology

This technology estimates power consumptions of the network system without using a smart meter and controls the system to reduce its power consumptions.

We are now promoting an implementation of a prototype system by combining the above technologies to confirm its effectiveness. Regarding the above (3), we also aim the Green-MIB (Management Information Base) to be standardized in IETF.

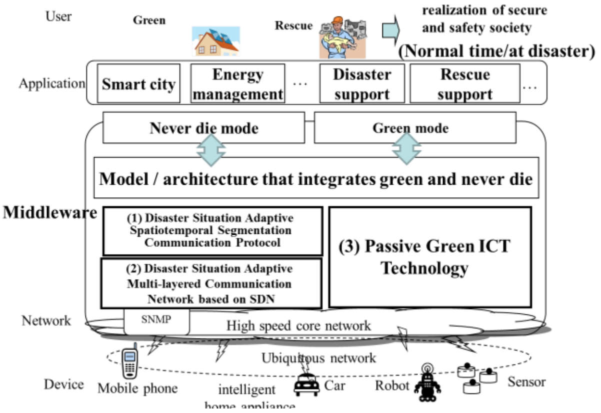
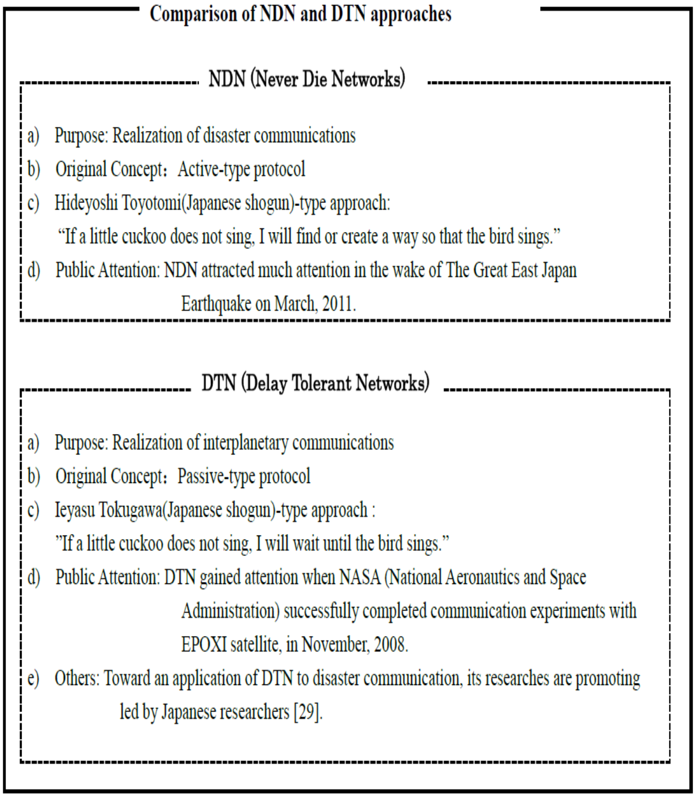

User
Green
Rescue
realization of secure and safety society
(Normal time/at disaster)
Application
Smart city
Energy management
…
Disaster support
Rescue support
…
Never die mode
Green mode
Model / architecture that integrates green and never die
Middleware
(1) Disaster Situation Adaptive Spatiotemporal Segmentation Communication Protocol
(2) Disaster Situation Adaptive Multi-layered Communication Network based on SDN
(3) Passive Green ICT Technology
Network
SNMP
High speed core network
Ubiquitous network
Device
Mobile phone
intelligent home appliance
Car
Robot
Sensor

Fig.4: The proposed architecture for realizing green and never die technologies

2.5 Comparison of NDN with DTN

In 2003, NDN was proposed as a network resilient to disaster. In the same year, DTN (Delay Tolerant Network) was also introduced towards the realization of an interplanetary internet. Fig.5 shows a comparison of the NDN with DTN.

As a basic idea of DTN, DTN temporarily store messages which are failed their transmissions caused by errors occurred in communication channels, and retransmit the messages after the recovery of the communication channels (Japanese shogun in Edo Period, Ieyasu Tokugawa-type approach: If a little cuckoo does not sing, I will wait until the bird sings.)



Comparison of NDN and DTN approaches

NDN (Never Die Networks)

a) Purpose: Realization of disaster communications
b) Original Concept: Active-type protocol
c) Hideyoshi Toyotomi(Japanese shogun)-type approach:
"If a little cuckoo does not sing, I will find or create a way so that the bird sings."
d) Public Attention: NDN attracted much attention in the wake of The Great East Japan Earthquake on March, 2011.

DTN (Delay Tolerant Networks)

a) Purpose: Realization of interplanetary communications
b) Original Concept: Passive-type protocol
c) Ieyasu Tokugawa(Japanese shogun)-type approach :
"If a little cuckoo does not sing, I will wait until the bird sings."
d) Public Attention: DTN gained attention when NASA (National Aeronautics and Space Administration) successfully completed communication experiments with EPOXI satellite, in November, 2008.
e) Others: Toward an application of DTN to disaster communication, its researches are promoting led by Japanese researchers [29].

Fig. 5: A comparison of NDN with DTN.

On the other hand, NDN aims to provide uninterrupted real-time services through routing, line allocation controls, and switching and multiplexing of transmission media (wired and wireless) (Japanese shogun in Azuchi-Momoyama Period, Hideyoshi Toyotomi-type approach: If a little cuckoo does not sing, I will let the bird sing.) Application research of DTN to disaster communication is advancing led by Japanese researchers [29]. Toward the realization of an interplanetary internet, noise interference and long propagation delay in outer space will be a major problem; therefore, an application of research achievements of NDN and DTN to an interplanetary internet, a provision of basic technology, will be expected.

3 TRUSTED CLOUD COMPUTING

3.1 Towards Security to Increase User's Sense of Safety in Cloud Services

Users of cloud computing could not wipe away the anxiety about the data and programs may be abused or leaked because users submit their almost whole data and programs to a cloud provider. Those data and programs usually embody the company's stored knowledge. Then they shall be key elements of core competences of the user's company. We study countermeasures against the abuse or leakage of data and programs caused by cloud providers' careless or intentional crime. As one of those countermeasures, "secret sharing and secret computation by secure multi parties method" is studied and use practically in some applications. This method can calculate addition or subtraction easily, but to do multiplication it needs rather complex processes. We propose an idea to do multiplication and division easily as well as addition and subtraction.

3.2 Proposal of A New Secret Computation Method

We propose a new idea to do arithmetic multiplication and division easily as well as addition and subtraction, without complex computational processes. This method divides the data for adding or subtraction so as summation of the divided data is the original and the data for multiplication or division so as production of the divided data is the original. The mechanism of the new secret computation method is shown in Fig 6.

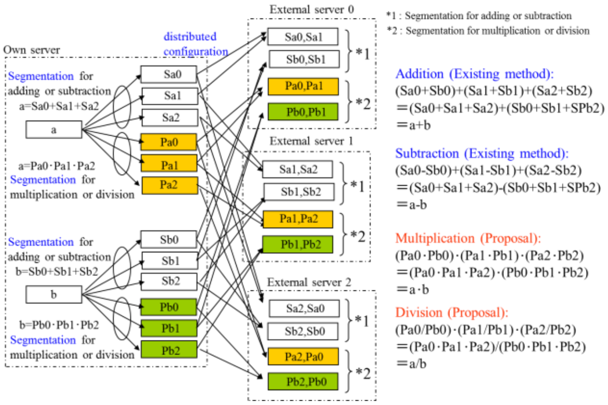

External server 0
*1 : Segmentation for adding or subtraction
*2 : Segmentation for multiplication or division
distributed configuration
Own server
Segmentation for adding or subtraction
a=Sa0+Sa1+Sa2
a
a=Pa0·Pa1·Pa2
Segmentation for multiplication or division
Sa0
Sa1
Sa2
Pa0
Pa1
Pa2
Segmentation for adding or subtraction
b=Sb0+Sb1+Sb2
b
b=Pb0·Pb1·Pb2
Segmentation for multiplication or division
Sb0
Sb1
Sb2
Pb0
Pb1
Pb2
Sa0,Sa1
Sb0,Sb1
Pa0,Pa1
Pb0,Pb1
*1
*2
External server 1
Sa1,Sa2
Sb1,Sb2
Pa1,Pa2
Pb1,Pb2
*1
*2
External server 2
Sa2,Sa0
Sb2,Sb0
Pa2,Pa0
Pb2,Pb0
*1
*2
Addition (Existing method):
(Sa0+Sb0)+(Sa1+Sb1)+(Sa2+Sb2)
=(Sa0+Sa1+Sa2)+(Sb0+Sb1+SPb2)
=a+b
Subtraction (Existing method):
(Sa0-Sb0)+(Sa1-Sb1)+(Sa2-Sb2)
=(Sa0+Sa1+Sa2)-(Sb0+Sb1+SPb2)
=a-b
Multiplication (Proposal):
(Pa0·Pb0)·(Pa1·Pb1)·(Pa2·Pb2)
=(Pa0·Pa1·Pa2)·(Pb0·Pb1·Pb2)
=a·b
Division (Proposal):
(Pa0/Pb0)·(Pa1/Pb1)·(Pa2/Pb2)
=(Pa0·Pa1·Pa2)/(Pb0·Pb1·Pb2)
=a/b

Fig. 6: New secret computation method.

3.3 Application of Proposed Method

In this subsection, an application example of the proposed method is explained. As statistical processing, this method calculates the average score for each subject at test for students by using 3 servers (own server, cloud 1, cloud 2) as shown in Fig 7. In this case, each server calculates summation of longitudinal direction.

- Own server: sets subject g (g=1 or 2) and requests a program to calculate the average score to cloud 1 and cloud 2.
- Could 1 (k=1): calculates summation of longitudinal direction of xg1s for subject g and returns the result to the own server.
- Could 2 (k=2): calculates summation of longitudinal direction of xg2s for subject g and returns the result to the own server.
- Own server: add the result of the cloud 1 to the result of the cloud 2, and then divides it by number of students. The result is average score E(xg).

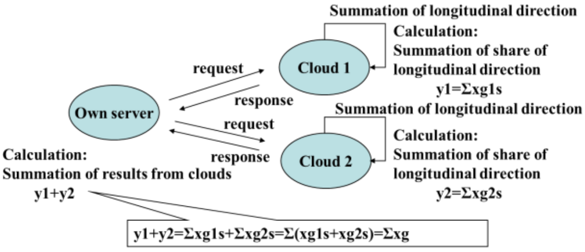

Summation of longitudinal direction
Calculation:
Summation of share of longitudinal direction
y1=Σxg1s
Cloud 1
request
response
Own server
request
response
Summation of longitudinal direction
Calculation:
Summation of share of longitudinal direction
y2=Σxg2s
Cloud 2
Calculation:
Summation of results from clouds
y1+y2
y1+y2=Σxg1s+Σxg2s=Σ(xg1s+xg2s)=Σxg

Fig. 7: An application of the proposed method.

4 CONCLUSION

This is a vision paper on network computing towards a sustainable symbiotic information society. We are now promoting an implementation of a prototype system by amalgamating the above technologies. As the future works, it is expected to promote research towards the realization of interplanetary internet based on contributions of NDN and DTN.

This research is partially supported by JSPS Kakenhi-Kiban (A) - 26240012 (2014-2016).

REFERENCES

[1] N. Shiratori, N. Uchida, Y. Shibata, and S. Izumi, Never Die Network towards Disaster-resistant Information Communication Systems, ASEAN Engineering Journal Part D, Vol.1, No.2, pp.1-22, (2013)[Invited Paper].

[2] T. Aoyama, N. Shiratori, K. Hagimoto, H Gambe, and Y. Mochida, Lessons of the Great East Japan Earthquake [Guest Editorial], IEEE Communications Magazine, Vol.52, No.3, pp.21-22 (2014).

[3] Y. Shibata, N. Uchida, and N. Shiratori, Analysis of and Proposal of Disaster Information Network from

Experience of the Great East Japan Earthquake, IEEE Comm. Mag., Vol.52, No.3, pp.44-48 (2014).

[4] The Asahi Shinbun, http://ajw.asahi.com/article/0311disaster/analysis/AJ201208300060S

[5] J. Sullivan, Network Fault tolerance system, A Thesis to the Faculty of the WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the Degree of Master of Science (2000).

[6] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, Resilient Overlay Network, Proc. of the 18th ACM SOSP2001, pp.131-145 (2001).

[7] T. Suganuma, G. Kitagata, T. Katoh, and N. Shiratori, Configuration of Never Die Network Service Function for Wireless Network, Proc. of the IEICE General Conference, p.376 (2003).

[8] N. Shiratori, Never Die Network: Towards Network Operating under Worsening Environment, http://kaken.nii.ac.jp/d/p/19650007.en.html

[9] G. Keeni, RFC4498: The Managed Object Aggregation MIB (2006).

[10] W. Kuji, K. Koide, and N. Shiratori, Development of Network Management Technology towards Never Die Network, IPSJ Tohoku-branch, 07-5-A-23 (2008).

[11] W. Kuji, G. Satou, K. Koide, Y. Shibata, and N. Shiratori, Never-Die Network and Disaster-Control System, IPSJ SIG Notes Vol.54, pp.131-135 (2008).

[12] T. Inaba, T. Ogasawara, N. Kita, N. Nakamura, T. Suganuma, and N. Shiratori, Green-oriented Never Die Network Management: The Concept and Design, Proc. of ICSAI 2012, pp.529-535 (2012).

[13] N. Shiratori, T. Inaba, N. Nakamura, and Takuo Suganuma, Disaster-resistant Green-oriented Never Die Network, IPSJ Journal, Vol.53, No.7, pp.1821-1831 (2012).

[14] Japan Police Department, The Great East Japan Disaster, http://www.npa.go.jp/archive/keibi/biki/index.htm (August 29, 2012)

[15] Japan Meteorological Agency, Past Reports of Earthquake and Tsunami, http://www.seisvol.kishou.go.jp/eq/higai/higai-1995.html

[16] N. Uchida, K. Takahata, and Y. Shibata, Disaster Information System from Communication Traffic Analysis and Connectivity (Quick Report from Japan Earthquake and Tsunami on March 11th, 2011), Proc. of NBIS2011, pp.279-285 (2011).

[17] Y. Shibata, N. Uchida, and Y. Ohashi, Problem Analysis and Solutions of Information Network Systems on East Japan Great Earthquake, Proc. of IWDENS2012, pp.1054-1059 (2012).

[18] Ministry of Internal Affairs and Communication, About How to Secure Communication Systems on Emergent Affair such as A Large Scale Disaster, http://www.soumu.go.jp/main_sosiki/kenkyu/saigai/index.html

[19] T. Watanabe, T. Oishi, et al., Research and Development of Disaster People/Local Government Support Information System, The Second Convention of Japan Society for Disaster Information Studies, pp.163-172 (2000).

[20] D. Nakamura, N. Uchida, H. Asahi, K. Takahata, K. Hashimoto, and Y. Shibata, Wide Area Disaster Information Network and Its Resource Management System, Proc. of AINA2003, pp.146-149 (2003).

[21] N. Uchida, K. Takahata, Y. Shibata, and N. Shiratori, Proposal of Never Die Network with the Combination of Cognitive Wireless Network and Satellite System, Proc. of NBIS2010, pp.365-370 (2010).

[22] N. Uchida, K. Takahata, Y. Shibata, and N. Shiratori, A Large Scale Robust Disaster information System based on Never Die Network, Proc. of AINA2012, pp.89-96 (2012).

[23] N. Shiratori and I. Noda, Special Issue on ICT which encourages society, IPSJ Journal, Vol.53, No.7, pp.1663-1664 (2012).

[24] http://www.ieice.org/cs/jpn/cs-edit/CFP/cfp_JB_2013.6.pdf

[25] http://www.nict.go.jp/en/index.html

[26] http://www.kyushu-u.ac.jp/pressrelease/2012/2012_01_23.pdf

[27] http://www.softbankmobile.co.jp/ja/news/press/2012/20120510_01/

[28] K. Fall, A Delay-Tolerant Network Architecture for Challenged Internets, Proc. of ACM SIGCOMM2003, pp.27-34 (2003).

[29] M. Tsuru, M. Uchida, T. Takine, A. Nagata, T. Matsuda, H. Miwa, and S. Yamamura, Delay Tolerant Networking Technology - The Latest Trends and Prospects, IEICE Communications Society Magazine, Vol.16, pp.57-68 (2011).

[30] JSPS Kakenhi, Grants-in-Aid for Scientific Research on Challenging Exploratory Research, Never Die Networks: Towards Networks Resilient to Worsened Network Environment (2007-2009).

[31] JSPS Kakenhi, Grants-in-Aid for Scientific Research (A), Green-oriented Never Die Networks which Adapt to Disaster Situation and Maximizes Satisfaction Levels of Connectivity by Massively Large Number of Users (2014-2016).

[32] N. Shiratori, T. Suganuma, S. Sugiura, G. Chakraborty, K. Sugawara, T. Kinoshita, and E.S. Lee, Framework of a flexible computer communication network, Computer Communications, Vol.19, No.14, pp.1268-1275, 1996.

Keynote Speech 3:
Dr. Masashi Saitoh
(Information Technology R&D Center,
Mitsubishi Electric Corp.)

Future Informatics Environment

- a Perspective from 30 years experience -

Mitsubishi Electric Corp.
Information Technology R&D Center
Masashi SAITO

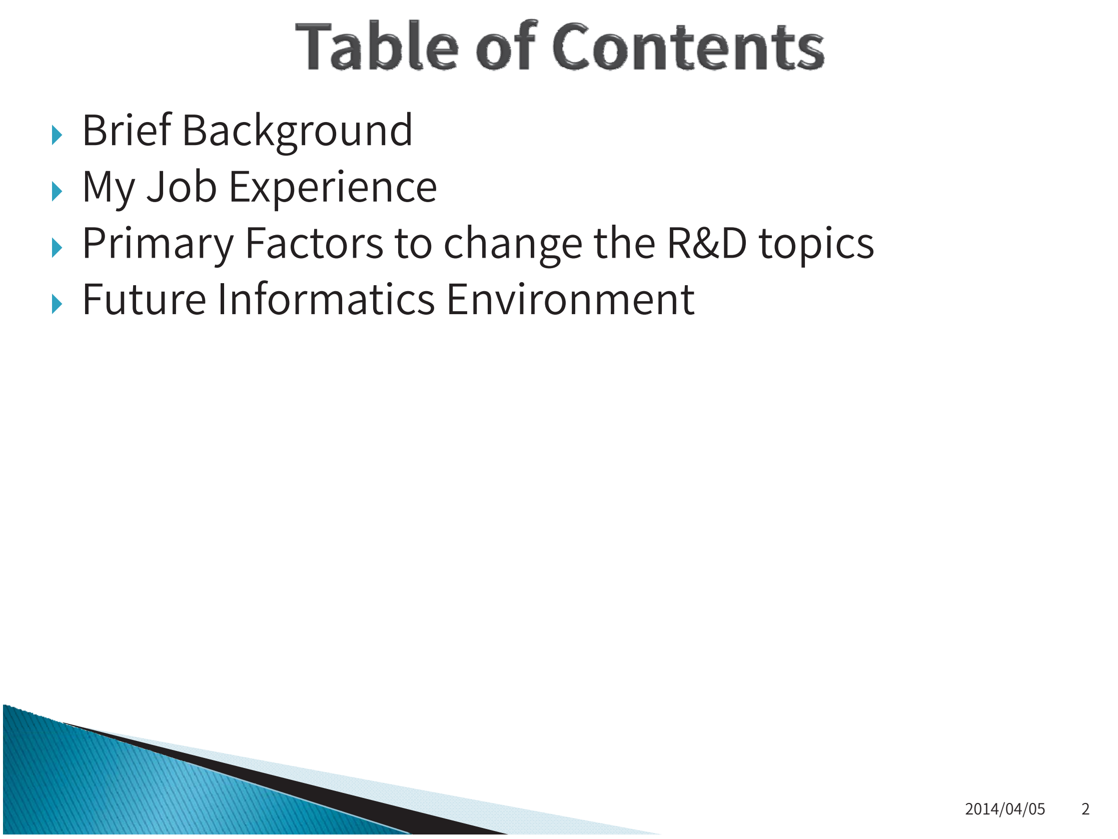
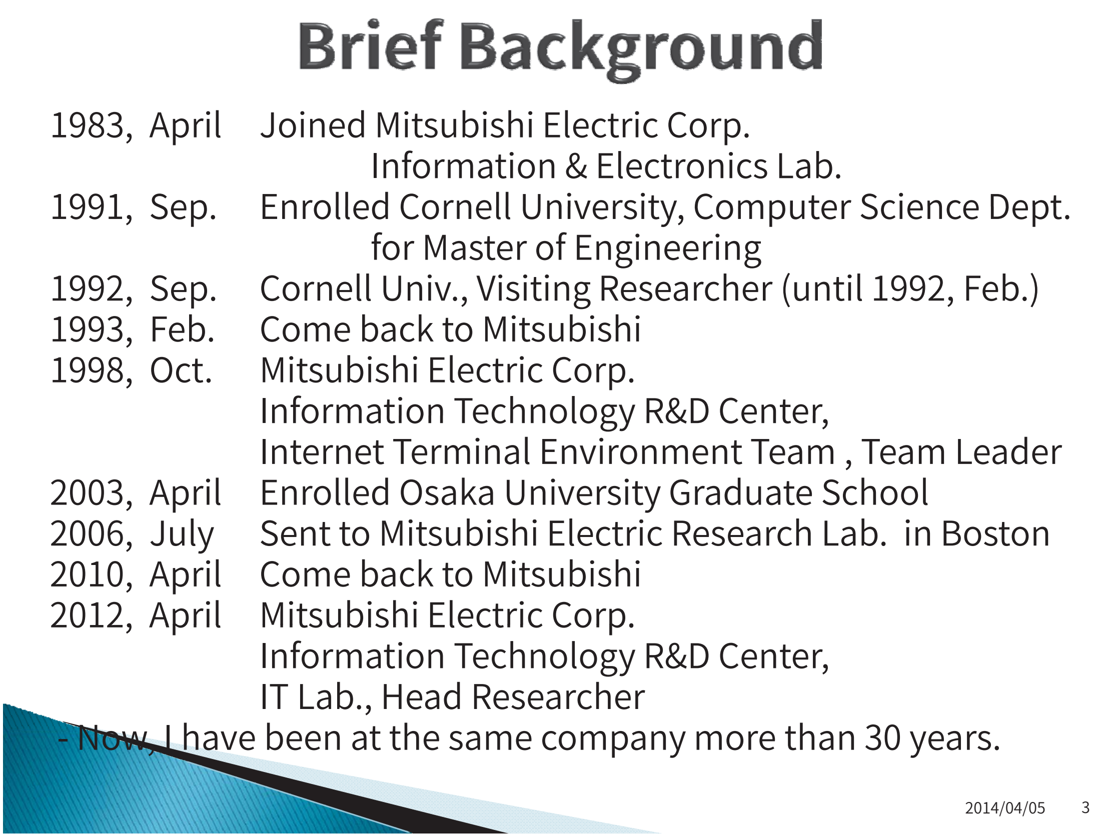


Table of Contents

- Brief Background
- My Job Experience
- Primary Factors to change the R&D topics
- Future Informatics Environment

2014/04/05 2

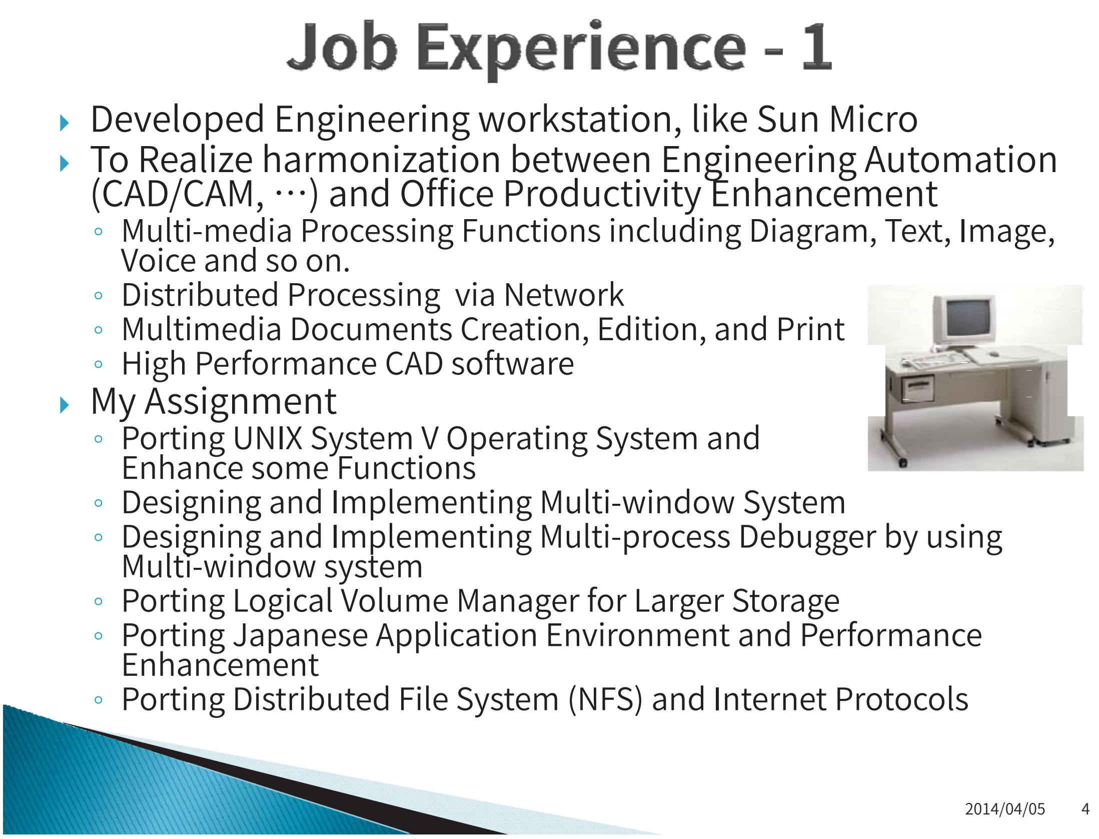


Brief Background

1983, April Joined Mitsubishi Electric Corp.
Information & Electronics Lab.
1991, Sep. Enrolled Cornell University, Computer Science Dept.
for Master of Engineering
1992, Sep. Cornell Univ., Visiting Researcher (until 1992, Feb.)
1993, Feb. Come back to Mitsubishi
1998, Oct. Mitsubishi Electric Corp.
Information Technology R&D Center,
Internet Terminal Environment Team , Team Leader
2003, April Enrolled Osaka University Graduate School
2006, July Sent to Mitsubishi Electric Research Lab. in Boston
2010, April Come back to Mitsubishi
2012, April Mitsubishi Electric Corp.
Information Technology R&D Center,
IT Lab., Head Researcher

- Now, I have been at the same company more than 30 years.

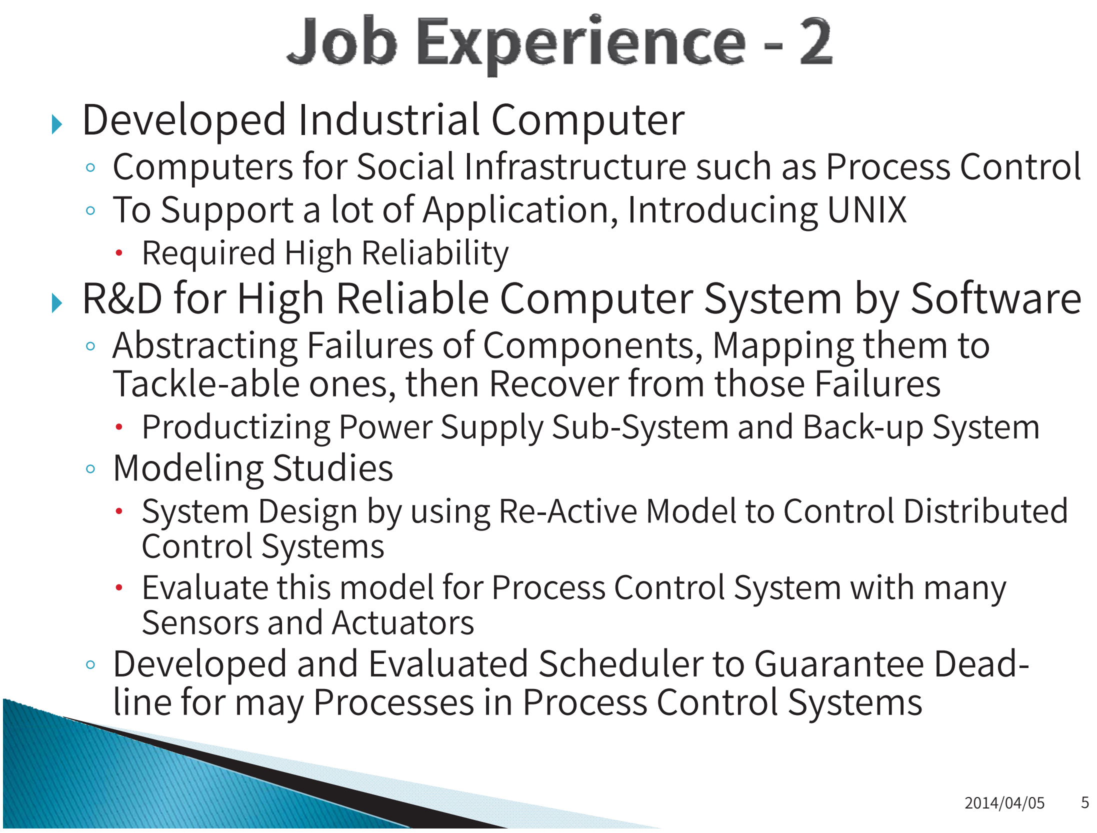
2014/04/05 3



Job Experience - 1

- Developed Engineering workstation, like Sun Micro
- To Realize harmonization between Engineering Automation (CAD/CAM, …) and Office Productivity Enhancement
  - Multi-media Processing Functions including Diagram, Text, Image, Voice and so on.
  - Distributed Processing via Network
  - Multimedia Documents Creation, Edition, and Print
  - High Performance CAD software
- My Assignment
  - Porting UNIX System V Operating System and Enhance some Functions
  - Designing and Implementing Multi-window System
  - Designing and Implementing Multi-process Debugger by using Multi-window system
  - Porting Logical Volume Manager for Larger Storage
  - Porting Japanese Application Environment and Performance Enhancement
  - Porting Distributed File System (NFS) and Internet Protocols

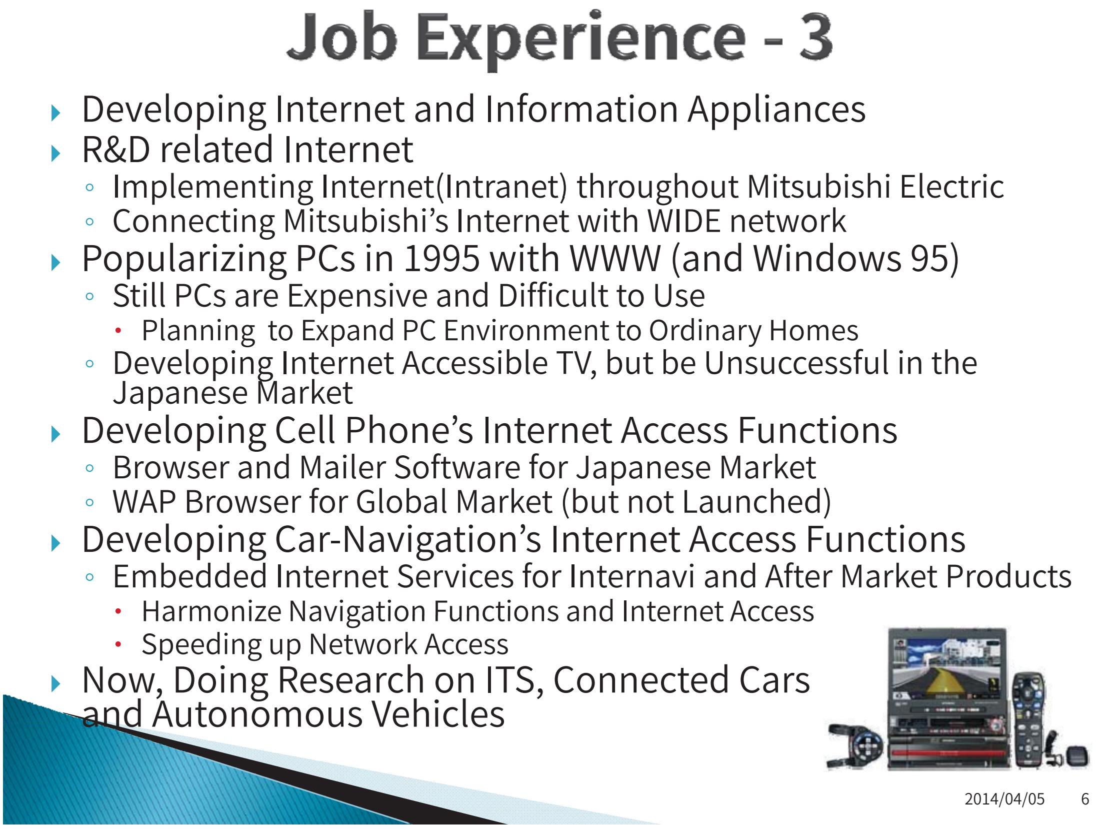
2014/04/05 4



Job Experience - 2

- Developed Industrial Computer
  - Computers for Social Infrastructure such as Process Control
  - To Support a lot of Application, Introducing UNIX
    - Required High Reliability
- R&D for High Reliable Computer System by Software
  - Abstracting Failures of Components, Mapping them to Tackle-able ones, then Recover from those Failures
    - Productizing Power Supply Sub-System and Back-up System
  - Modeling Studies
    - System Design by using Re-Active Model to Control Distributed Control Systems
    - Evaluate this model for Process Control System with many Sensors and Actuators
  - Developed and Evaluated Scheduler to Guarantee Dead-line for may Processes in Process Control Systems

2014/04/05 5

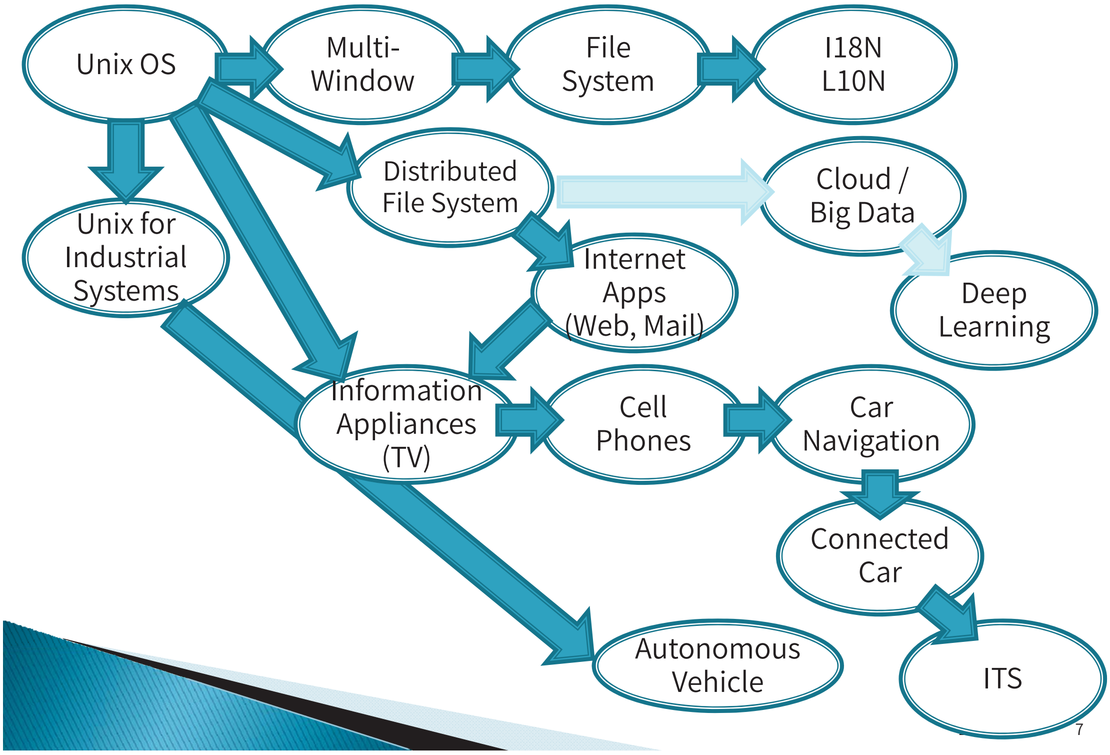


Job Experience - 3

- Developing Internet and Information Appliances
- R&D related Internet
  - Implementing Internet(Intranet) throughout Mitsubishi Electric
  - Connecting Mitsubishi's Internet with WIDE network
- Popularizing PCs in 1995 with WWW (and Windows 95)
  - Still PCs are Expensive and Difficult to Use
    - Planning to Expand PC Environment to Ordinary Homes
  - Developing Internet Accessible TV, but be Unsuccessful in the Japanese Market
- Developing Cell Phone's Internet Access Functions
  - Browser and Mailer Software for Japanese Market
  - WAP Browser for Global Market (but not Launched)
- Developing Car-Navigation's Internet Access Functions
  - Embedded Internet Services for Internavi and After Market Products
    - Harmonize Navigation Functions and Internet Access
    - Speeding up Network Access
- Now, Doing Research on ITS, Connected Cars and Autonomous Vehicles

2014/04/05 6

Transitions of Jobs



Unix OS

Multi-Window

File System

I18N L10N

Unix for Industrial Systems

Distributed File System

Cloud / Big Data

Internet Apps (Web, Mail)

Deep Learning

Information Appliances (TV)

Cell Phones

Car Navigation

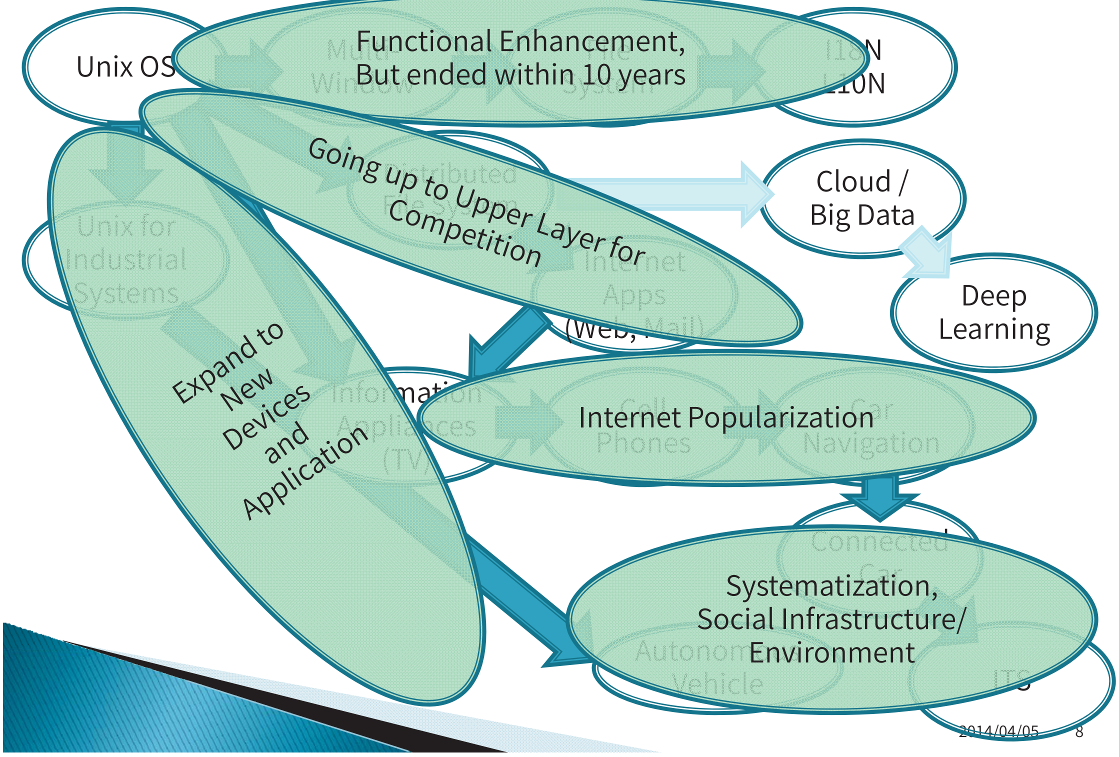
Connected Car

Autonomous Vehicle

ITS

7

Primary Factors to change the R&D topics



Functional Enhancement, But ended within 10 years

Going up to Upper Layer for Competition

Expand to New Devices and Application

Internet Popularization

Systematization, Social Infrastructure/ Environment

Unix OS

Multi-Window

File System

I18N L10N

Distributed File System

Cloud / Big Data

Deep Learning

Unix for Industrial Systems

Internet Apps (Web, Mail)

Information Appliances (TV)

Cell Phones

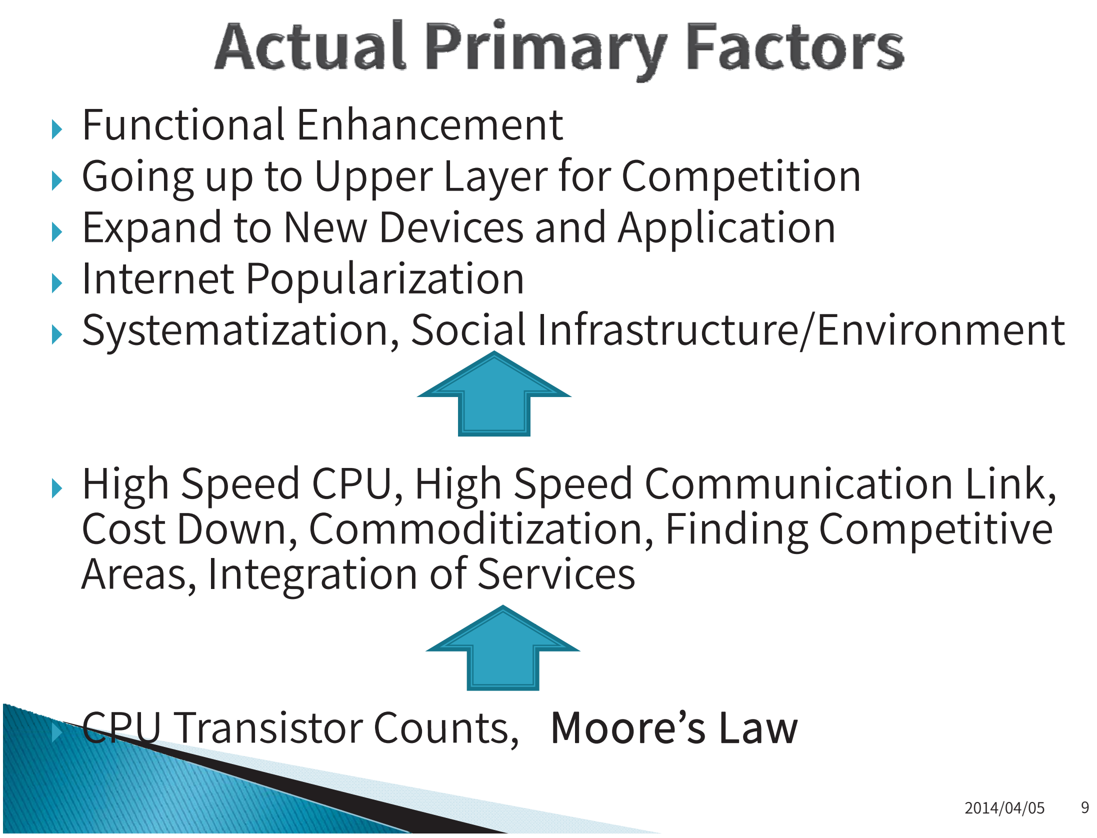
Car Navigation

Connected Car

Autonomous Vehicle

ITS

2014/04/05 8



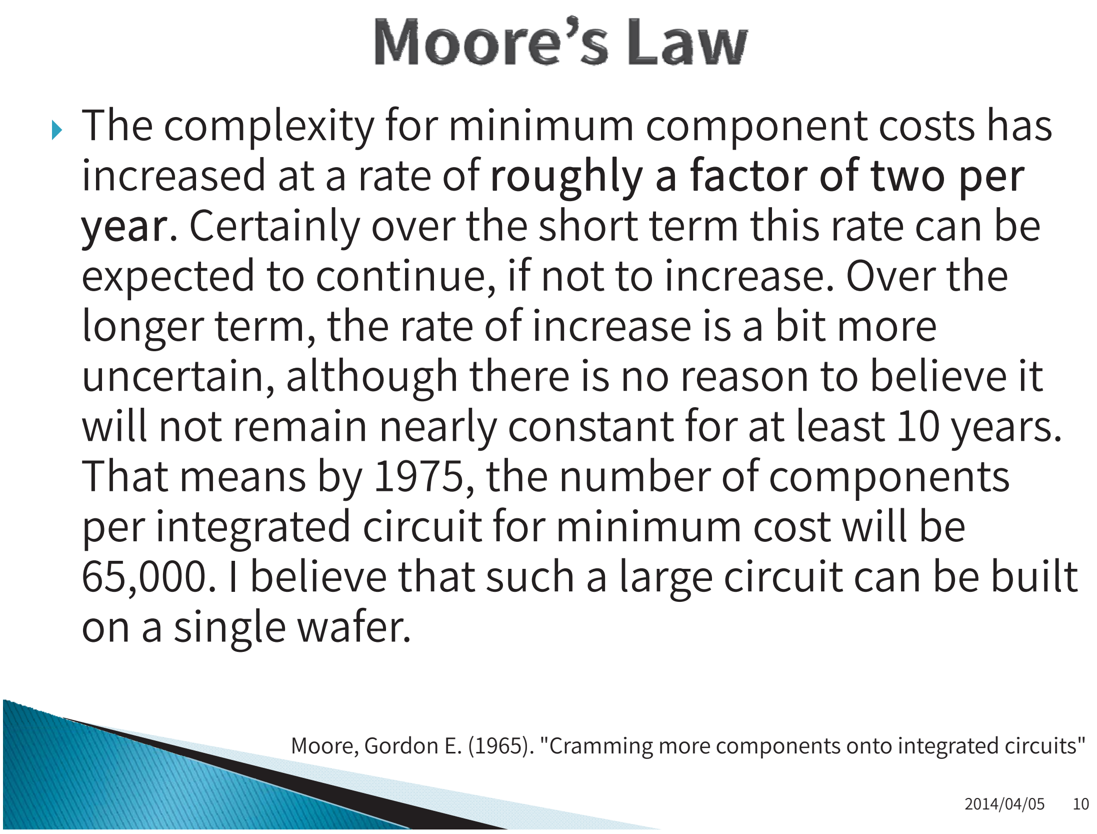
Actual Primary Factors

- Functional Enhancement
- Going up to Upper Layer for Competition
- Expand to New Devices and Application
- Internet Popularization
- Systematization, Social Infrastructure/Environment

- High Speed CPU, High Speed Communication Link, Cost Down, Commoditization, Finding Competitive Areas, Integration of Services

- CPU Transistor Counts, **Moore’s Law**

2014/04/05 9



Moore’s Law

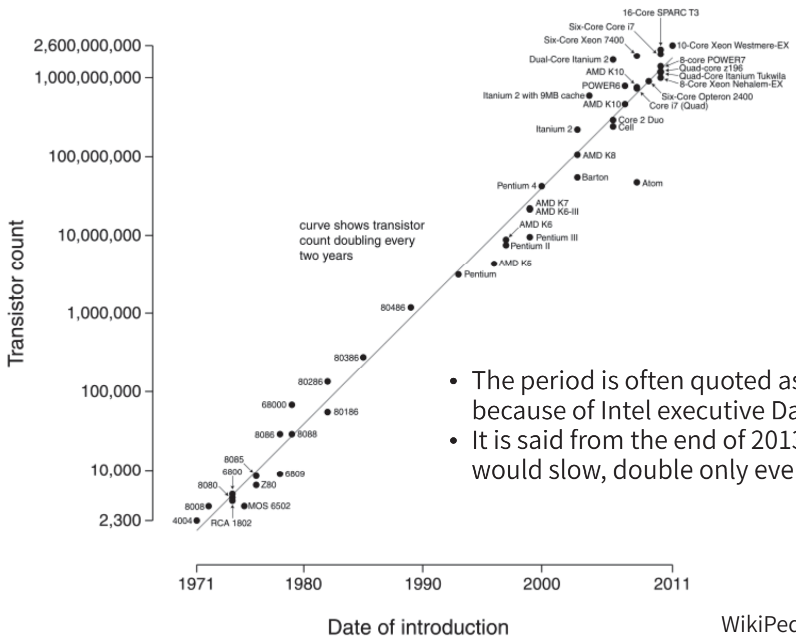
- The complexity for minimum component costs has increased at a rate of **roughly a factor of two per year**. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer.

Moore, Gordon E. (1965). "Cramming more components onto integrated circuits"

2014/04/05 10

Microprocessor Transistor Counts

Microprocessor Transistor Counts 1971-2011 & Moore’s Law

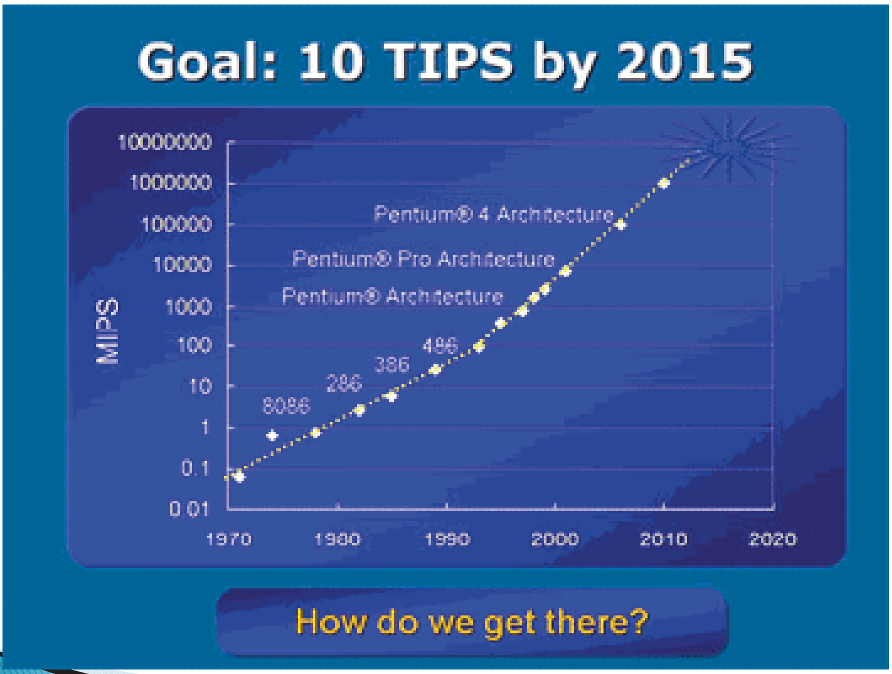


- The period is often quoted as 18 months because of Intel executive David House.
- It is said from the end of 2013, growth would slow, double only every 3 years.

WikiPedia, “Moore’s Law’

2014/04/05 11

Converted to MIPS



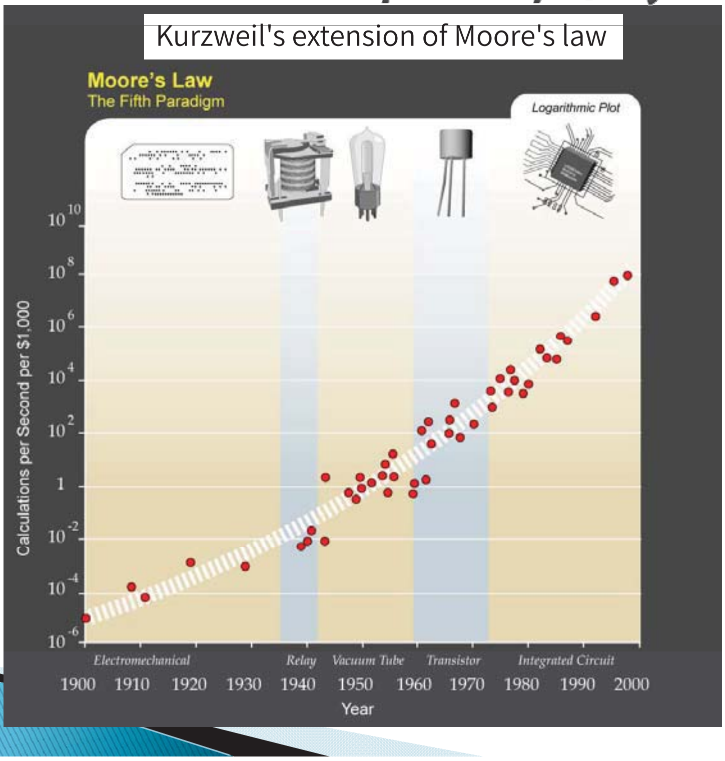
Goal: 10 TIPS by 2015

How do we get there?

https://software.intel.com/en-us/articles/the-new-era-of-tera-scale-computing

2014/04/05 12

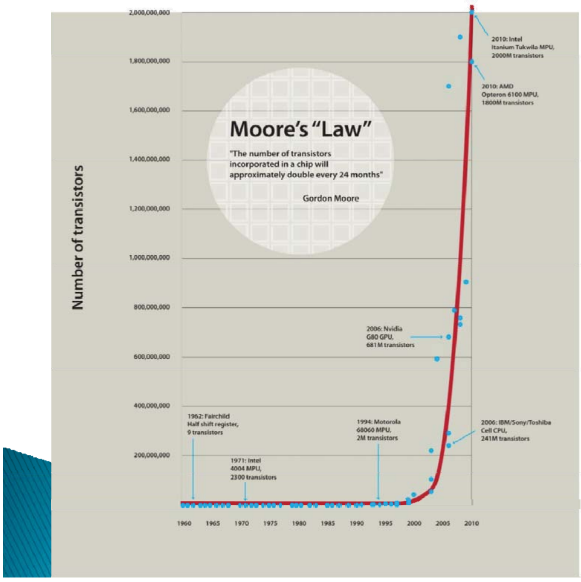
Calculations / sec / $1,000



WikiPedia
“Moore’s Law’

2014/04/05 13

In Reality, we are at Vertical Line

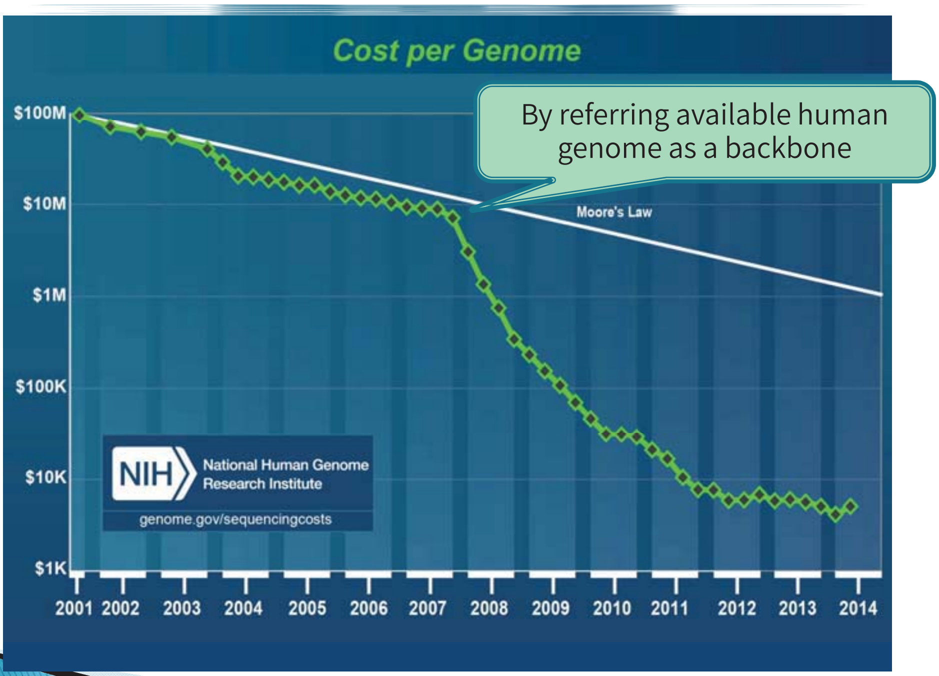


- Everything being Digital
  - Text, Still Image, Voice, Video, …
- Use More IT Tend to Have Higher Levels of Productivity and Faster Productivity growth than competitors
- Less Focused on a Diversified Set of Applications and Process Innovation

Erik Brynjolfsson, Andrew McAfee (2014).
“The Second Machine Age”

2014/04/05 14

Impact to Applications – Human Genome Sequencing”

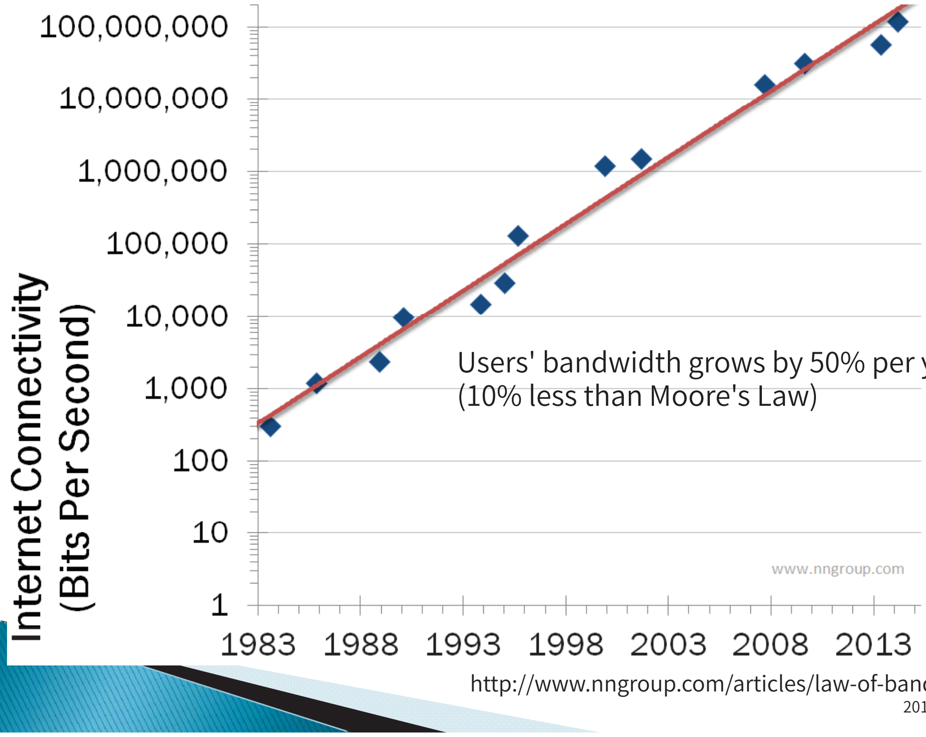


By referring available human genome as a backbone

http://www.genome.gov/sequencingcosts/

2014/04/05 15

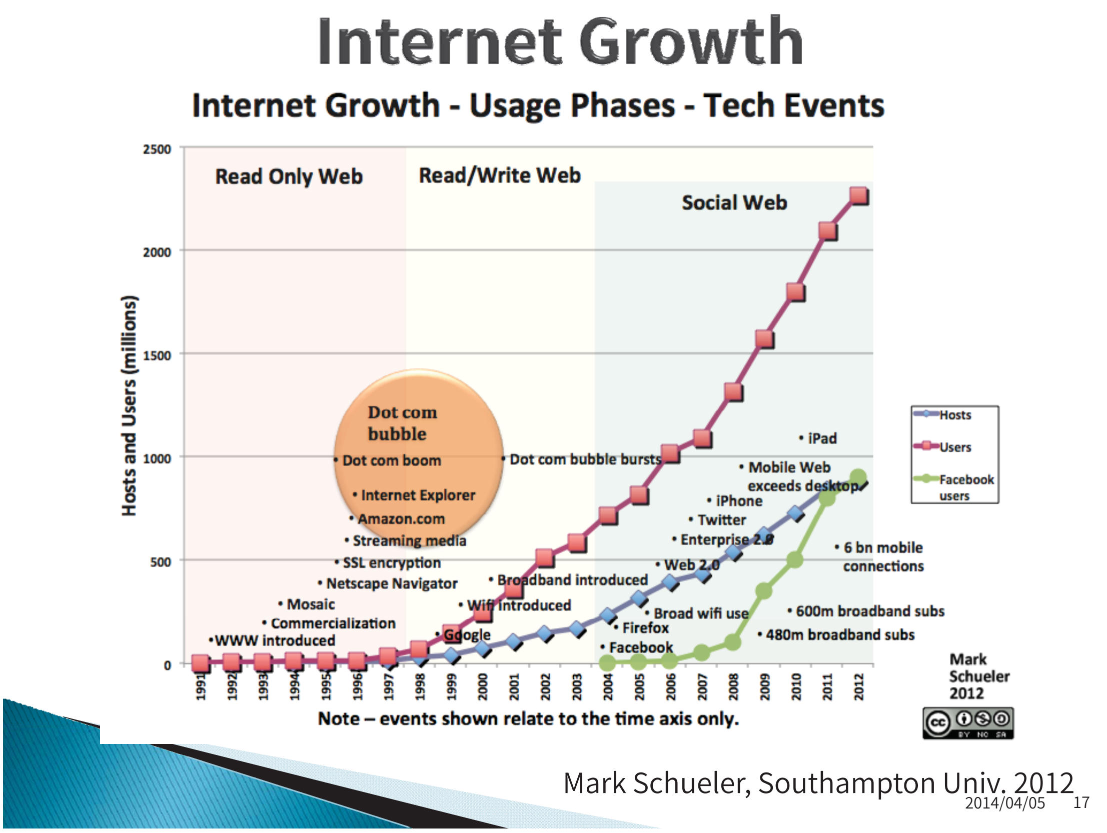
Internet Connectivity, Nielsen’s Law



Users' bandwidth grows by 50% per year (10% less than Moore's Law)

http://www.nngroup.com/articles/law-of-bandwidth/

2014/04/05 16



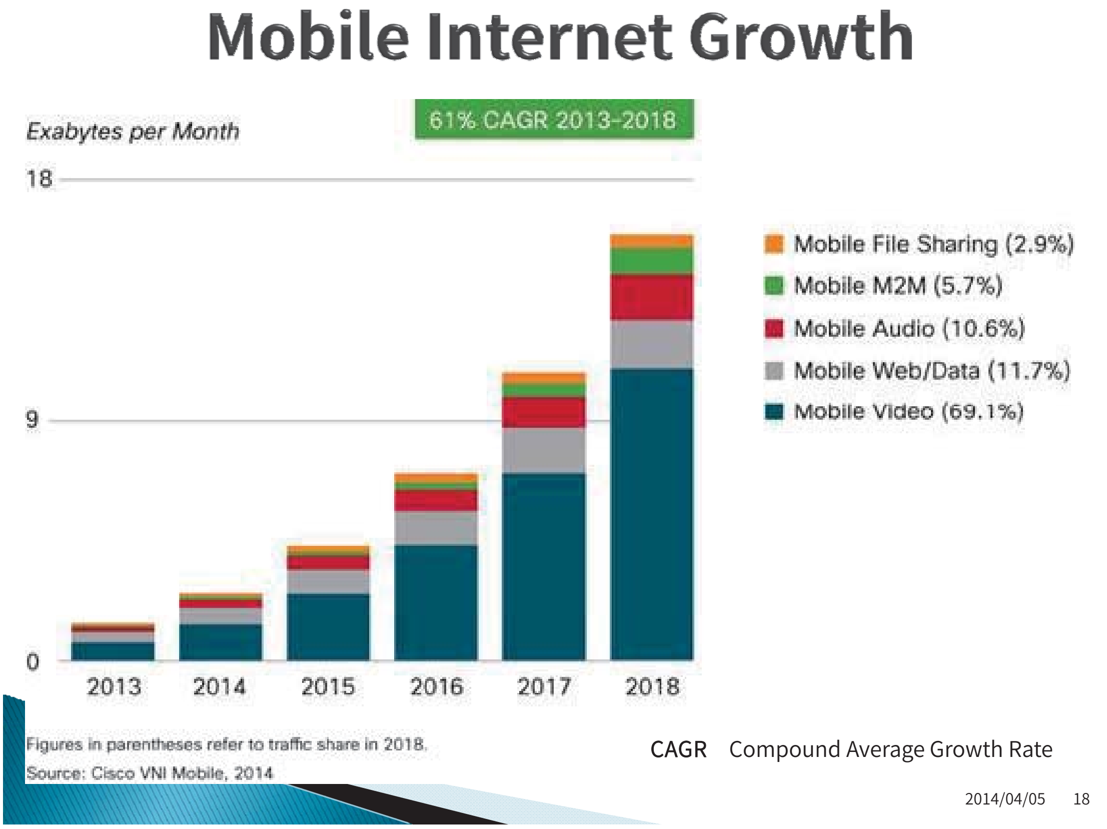
Internet Growth

Internet Growth - Usage Phases - Tech Events

Note – events shown relate to the time axis only.

Mark Schueler, Southampton Univ. 2012

2014/04/05 17



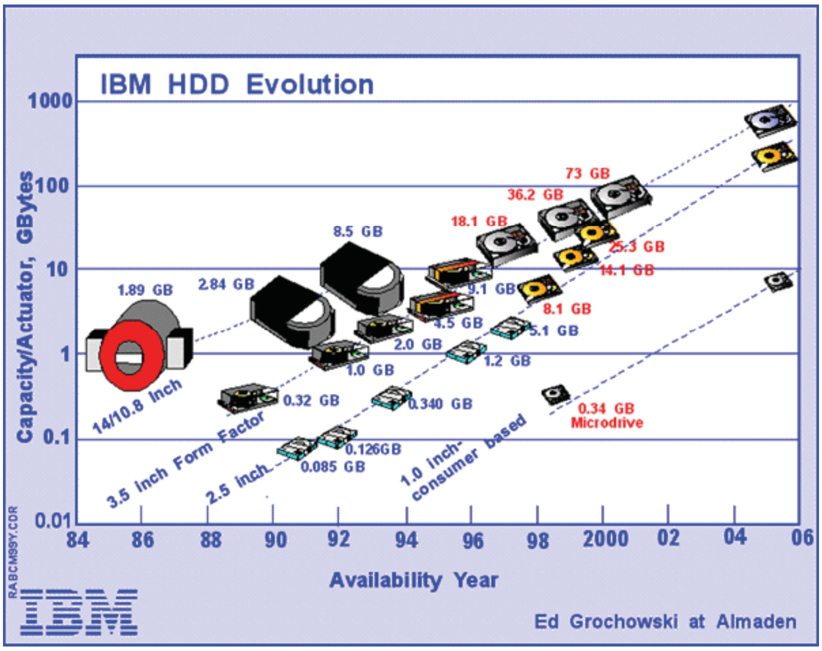
Mobile Internet Growth

Figures in parentheses refer to traffic share in 2018.
Source: Cisco VNI Mobile, 2014

CAGR　Compound Average Growth Rate

2014/04/05 18

Storage Growth



- Seagate Announced 3.5inch 8TB HDD on Aug. 2014.

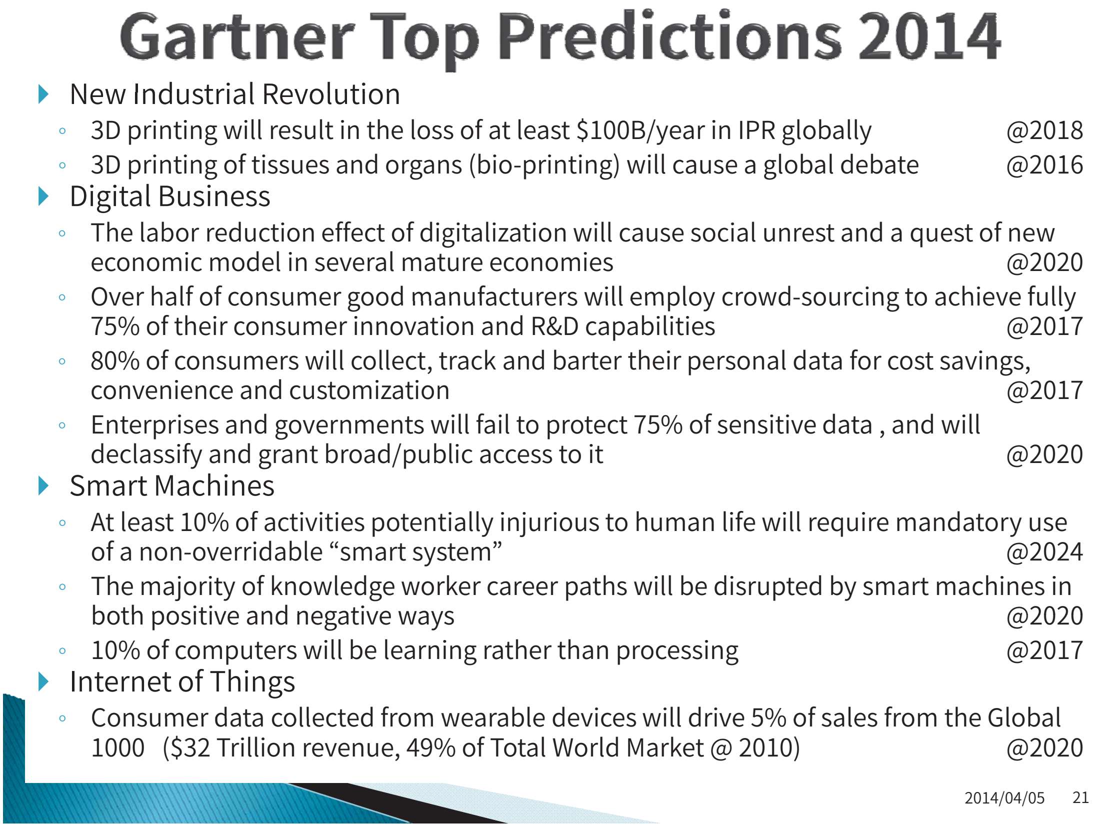
2014/04/05 19

Exponential Growth

- CPU, Internet, Storage, Cost, Commoditization
- Continuous Progress of IT and Creation of New Industries
  - MilkyWay-2(China, 33PFLOPS) is Faster than Human Brain (10PFLOPS)
  - Bonkras won Yonenaga-san in 2012 and Watson won the Jeopardy! Game in 2011
    - In case of Jeopardy, to Support Real Time Communication, it was held at Watson Lab., NY.
- Expanding Digital Data
  - 50B Machines and Devices will be Connected @ 2020 (CISCO), World Population is 7.2B
  - Expand Stored data from 130 Exa-Bytes @ 2005 to 40 Zetta-Bytes @ 2020 (IDC)
  - Those Huge Data would be used in Business Area and Decreasing Social Life Cost
- Fusing Physical World and Cyber Word
  - New Services which Integrate Components in Physical World and Cyber one (CPS)
  - New Applications are not only Embedded Equipment but Transportation, Defense, Energy, Industry, Health Care, Bio-tech, Agriculture, Social Infrastructure
- IT is Penetrating into Social Life (Lights and Shadows)
  - Social Infrastructure would NOT Work without IT, Depending our Life on IT, …
  - Privacy, Copyrights, Compliance, Security, Dependability are Required from the point of view our Safe and Secure Life
  - Continuous IT Support Becomes Essential (365 Days, 24 Hours)

JST/CRDS (2013), “Panoramic View Point (2013)”

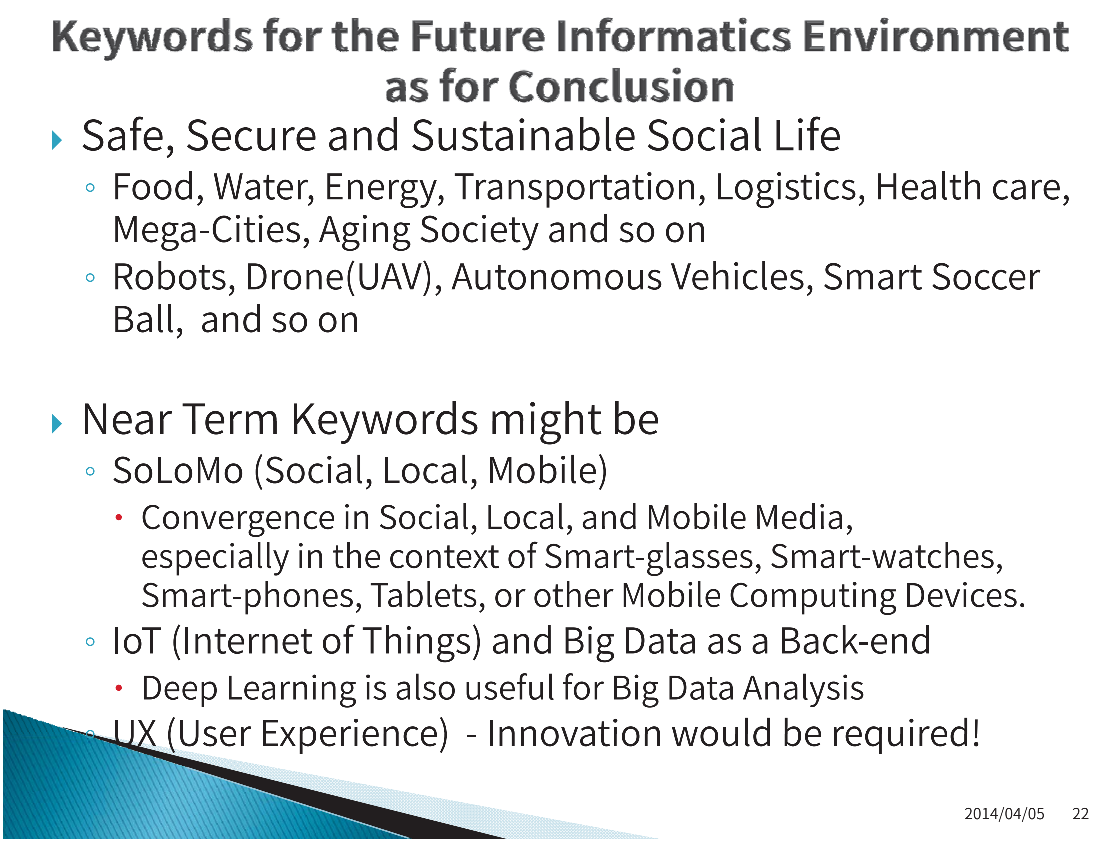
2014/04/05 20



Gartner Top Predictions 2014

- New Industrial Revolution
  - 3D printing will result in the loss of at least $100B/year in IPR globally @2018
  - 3D printing of tissues and organs (bio-printing) will cause a global debate @2016
- Digital Business
  - The labor reduction effect of digitalization will cause social unrest and a quest of new economic model in several mature economies @2020
  - Over half of consumer good manufacturers will employ crowd-sourcing to achieve fully 75% of their consumer innovation and R&D capabilities @2017
  - 80% of consumers will collect, track and barter their personal data for cost savings, convenience and customization @2017
  - Enterprises and governments will fail to protect 75% of sensitive data , and will declassify and grant broad/public access to it @2020
- Smart Machines
  - At least 10% of activities potentially injurious to human life will require mandatory use of a non-overridable “smart system” @2024
  - The majority of knowledge worker career paths will be disrupted by smart machines in both positive and negative ways @2020
  - 10% of computers will be learning rather than processing @2017
- Internet of Things
  - Consumer data collected from wearable devices will drive 5% of sales from the Global 1000 ($32 Trillion revenue, 49% of Total World Market @ 2010) @2020

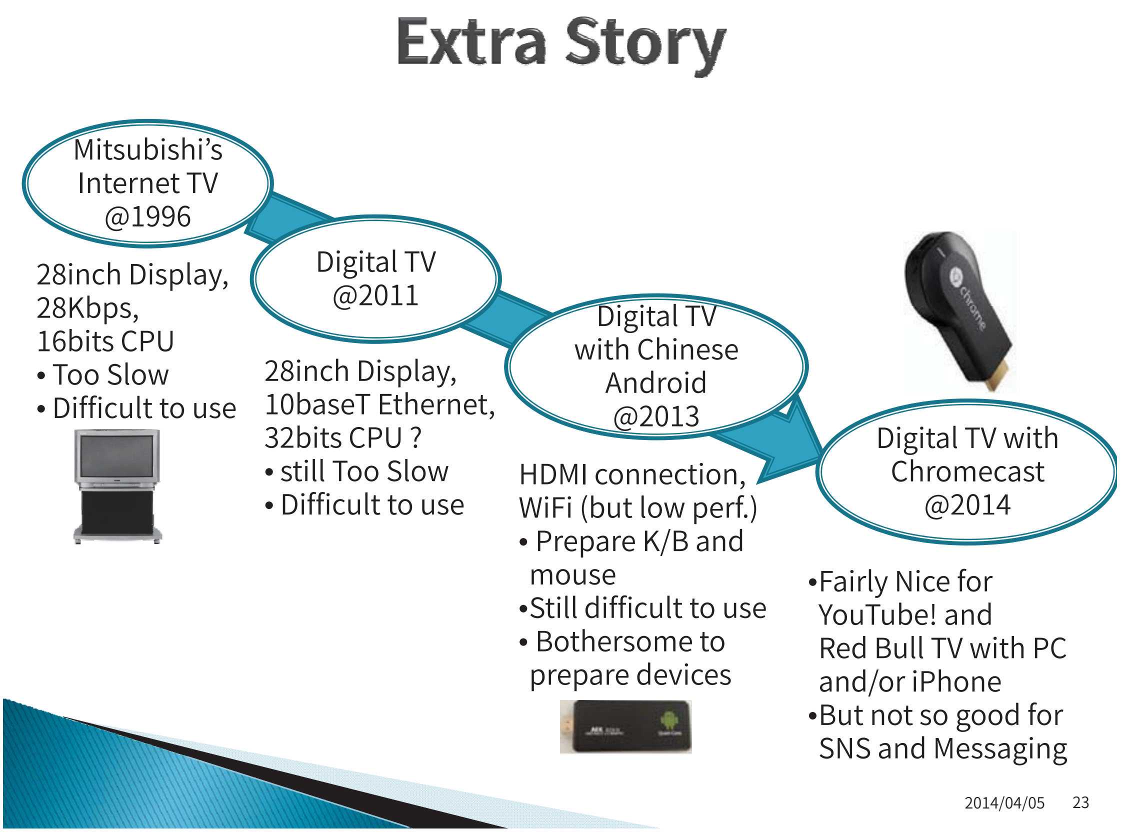
2014/04/05 21



Keywords for the Future Informatics Environment as for Conclusion

- Safe, Secure and Sustainable Social Life
  - Food, Water, Energy, Transportation, Logistics, Health care, Mega-Cities, Aging Society and so on
  - Robots, Drone(UAV), Autonomous Vehicles, Smart Soccer Ball, and so on
- Near Term Keywords might be
  - SoLoMo (Social, Local, Mobile)
    - Convergence in Social, Local, and Mobile Media, especially in the context of Smart-glasses, Smart-watches, Smart-phones, Tablets, or other Mobile Computing Devices.
  - IoT (Internet of Things) and Big Data as a Back-end
    - Deep Learning is also useful for Big Data Analysis
  - UX (User Experience) - Innovation would be required!

2014/04/05 22



Extra Story

Mitsubishi's Internet TV @1996

28inch Display, 28Kbps, 16bits CPU
- Too Slow
- Difficult to use

Digital TV @2011

28inch Display, 10baseT Ethernet, 32bits CPU ?
- still Too Slow
- Difficult to use

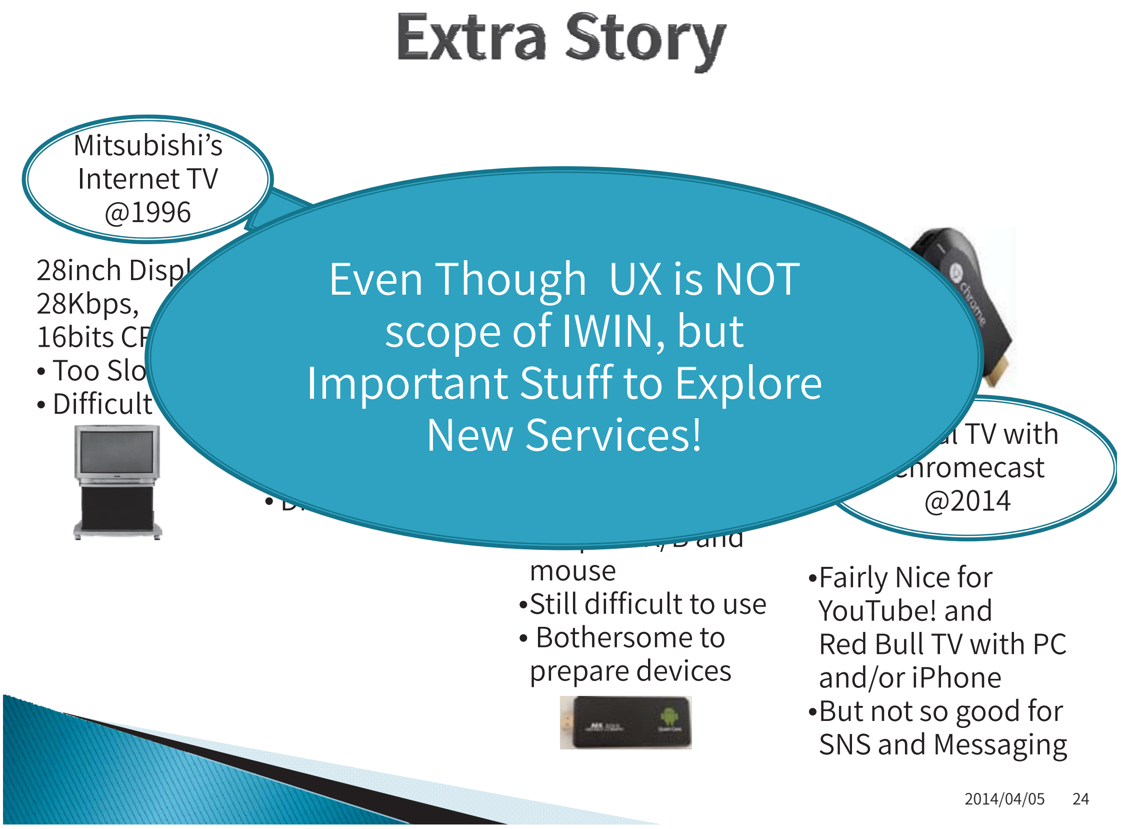
Digital TV with Chinese Android @2013

HDMI connection, WiFi (but low perf.)
- Prepare K/B and mouse
- Still difficult to use
- Bothersome to prepare devices

Digital TV with Chromecast @2014

- Fairly Nice for YouTube! and Red Bull TV with PC and/or iPhone
- But not so good for SNS and Messaging

2014/04/05 23



Extra Story

Mitsubishi's Internet TV @1996

28inch Displ

28Kbps,

16bits CP

• Too Slo

• Difficult

• D

Even Though UX is NOT scope of IWIN, but Important Stuff to Explore New Services!

al TV with

hromecast

@2014

/B and

mouse

- Still difficult to use
- Bothersome to prepare devices

- Fairly Nice for YouTube! and Red Bull TV with PC and/or iPhone
- But not so good for SNS and Messaging

2014/04/05 24

Panel Discussion

(Chair : Norio Shiratori)

＜Panel Session＞

1) Title

- **Toward Future ICT Systems**

  - Requirements and Solutions –

2) Panelists

- Dr. Masashi Saito, Mitsubishi Electric Corporation, Japan
- Prof. Yoh Shiraishi, Future University Hakodate, Japan
- Prof. Tomoya Kitani, Shizuoka University, Japan

3) Chair

- Prof. Norio Shiratori, Waseda University, Japan

Damage - resilient Network Services

- Trusted Cloud Computing -

Norio SHIRATORI

Waseda University / Tohoku University

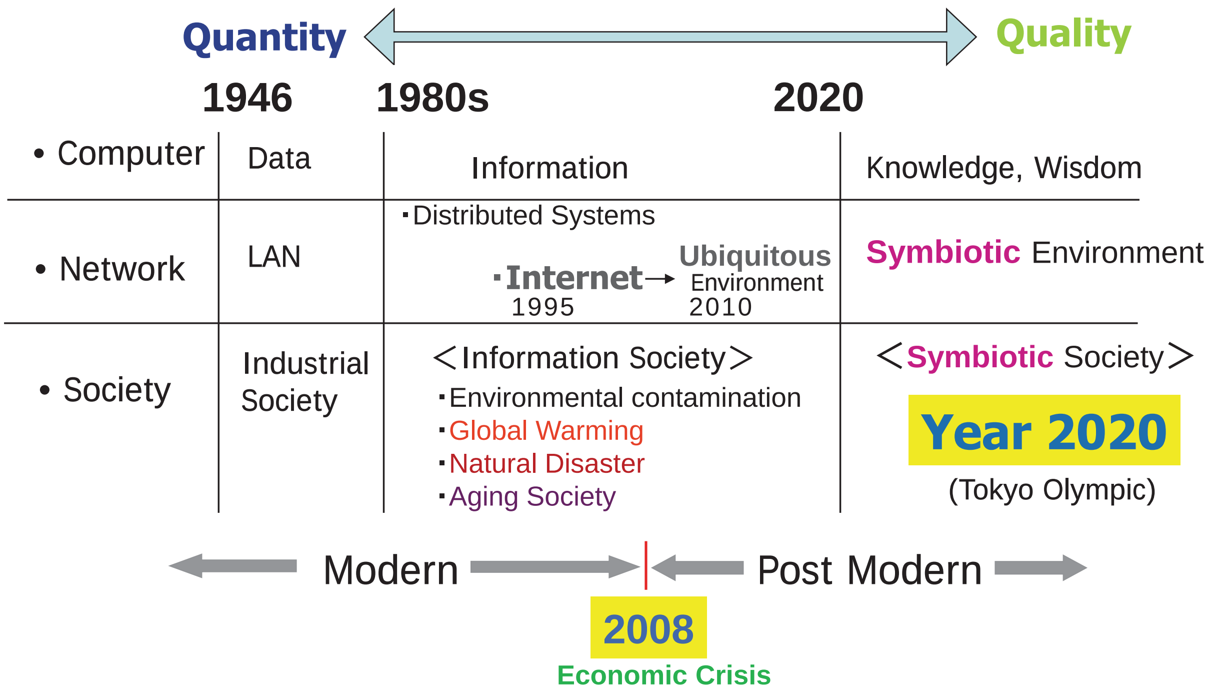
IEEE Fellow

Table of Contents

1. **Towards Sustainable Symbiotic Information Society**
2. **Never Die Networks**
3. **Trusted Cloud Computing**
4. **Conclusion**
   - Towards Sustainable Information Society -

1. Towards Sustainable Symbiotic Information Society

1.1 Problems



Quantity ↔ **Quality**

	1946	1980s	2020
• Computer	Data	Information	Knowledge, Wisdom
• Network	LAN	・Distributed Systems ・Internet → Ubiquitous Environment 1995 2010	**Symbiotic** Environment
• Society	Industrial Society	<Information Society> ・Environmental contamination ・Global Warming ・Natural Disaster ・Aging Society	<**Symbiotic** Society> **Year 2020** (Tokyo Olympic)

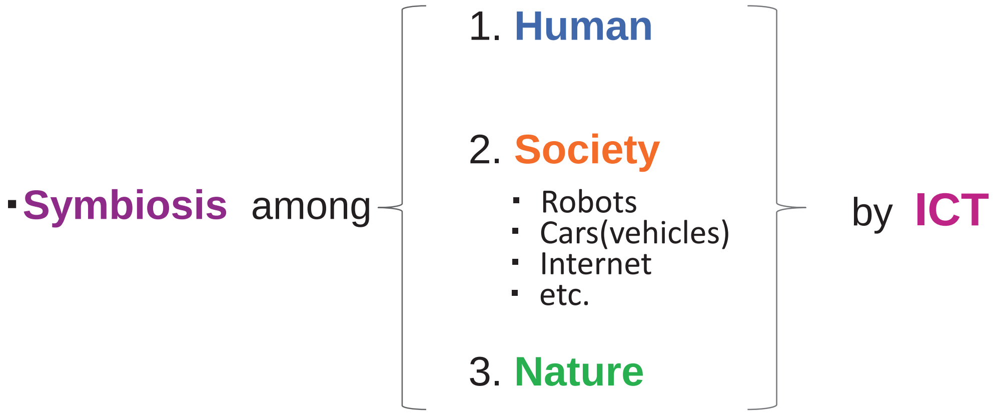
2008

Economic Crisis

1.2 Problems & Solutions

	Modern (Industrial Society)	Solutions: Post Modern (Symbiotic Society)
Evaluation Criterion	■ Rationality • Economy • Efficiency • Function	■ Rationality + α α : new criteria for relationship amon human, society and nature achieved by I T ・Our proposal : α = **Symbiosis**
View Point	Producer	User
Industry	large-scale production and consumption	Various, and small-scale production, recycle
Problems	･Environmental contamination ･Global Warming ･Natural Disaster ･Aging Society	・Symbiosis among human, society and nature achieved by **I T**
21st century	Post rationalism <20th century: Human conquers nature>	・Symbiotic Thought <Human assimilates into nature>

1.3 Symbiosis by ICT among Human, Society and Nature



▪**Symbiosis** among

1. **Human**
2. **Society**
   - Robots
   - Cars(vehicles)
   - Internet
   - etc.
3. **Nature**

by **ICT**

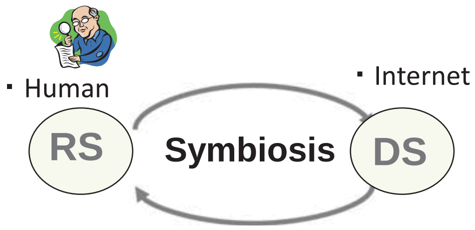
＜Definition of Symbiotic Computing＞

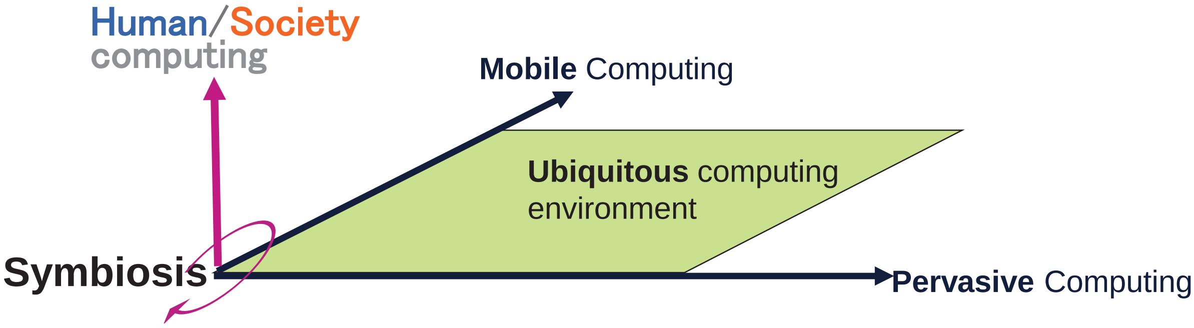
(1) Symbiosis by ICT

Symbiosis is defined as Integration and Fusion of both **RS** (Real Space) and **DS** (Digital Space)

▪**Symbiosis** by **ICT** among

1. Human
2. Society
   - ▪Robots
   - ▪Cars(vehicles)
   - ▪Internet, etc.
3. Nature





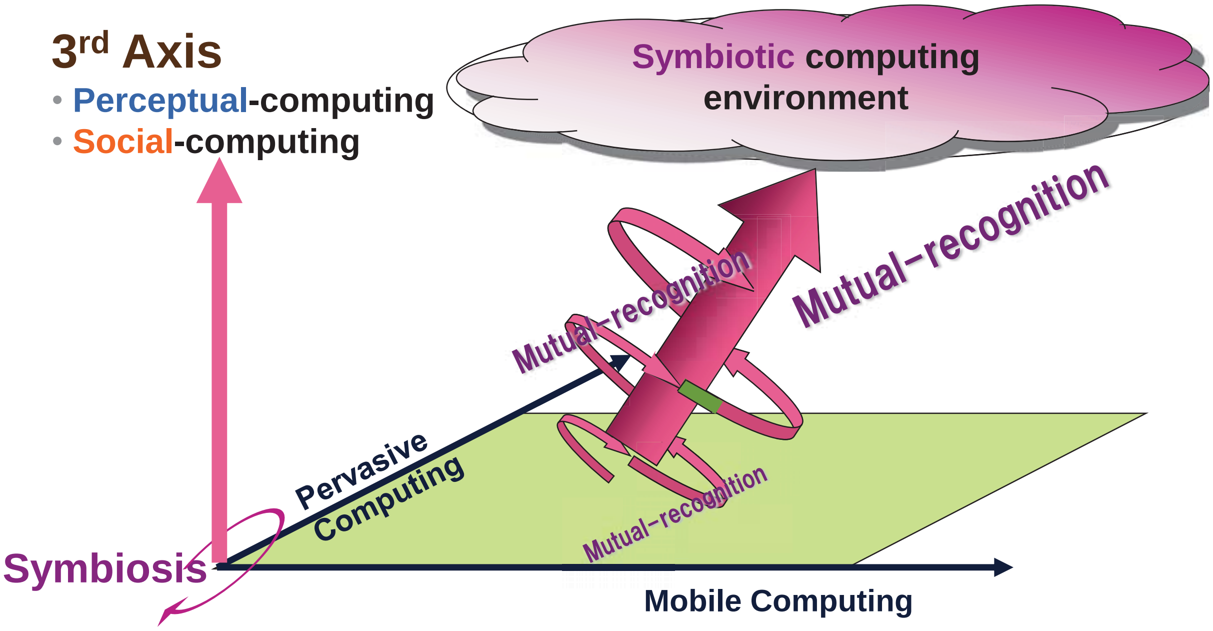
(2) Symbiotic Computing

- Symbiotic Computing is defined as a new computing to realize Symbiosis.
- Symbiotic Computing (Technologies) will be realized by mutual recognition between RS and DS based on Perceptual computing and Social computing.

(3) Symbiotic Society

- Symbiotic information environment will be constructed based on Symbiotic computing.
- Symbiotic Service (application) will be created by using Symbiotic Information Environment.
- Finally, everyone will be supported, whenever necessary only when needed, in an oblivious, gently and unintrusive sympathetic way.
- In addition, Human and IT environment will cooperatively coexist.

<Symbiotic Computing environment>



1.4 Symbiosis between Human's Life and Nature via ICT (Info. Comm. Tech.)

Part-1: Green-oriented Network and Its Applications

- Kurihara Green Project (2010-2011) –

A National Project Sponsored by *MIC, Japan

*MIC: Ministry of Internal Affairs and Communications

“Kurihara Green Project”

Green: Towards Symbiosis between Human and Nature

1. R&D Program (MIC project)
An ICT based experiment towards *Green* on a region of wide-area distributed community

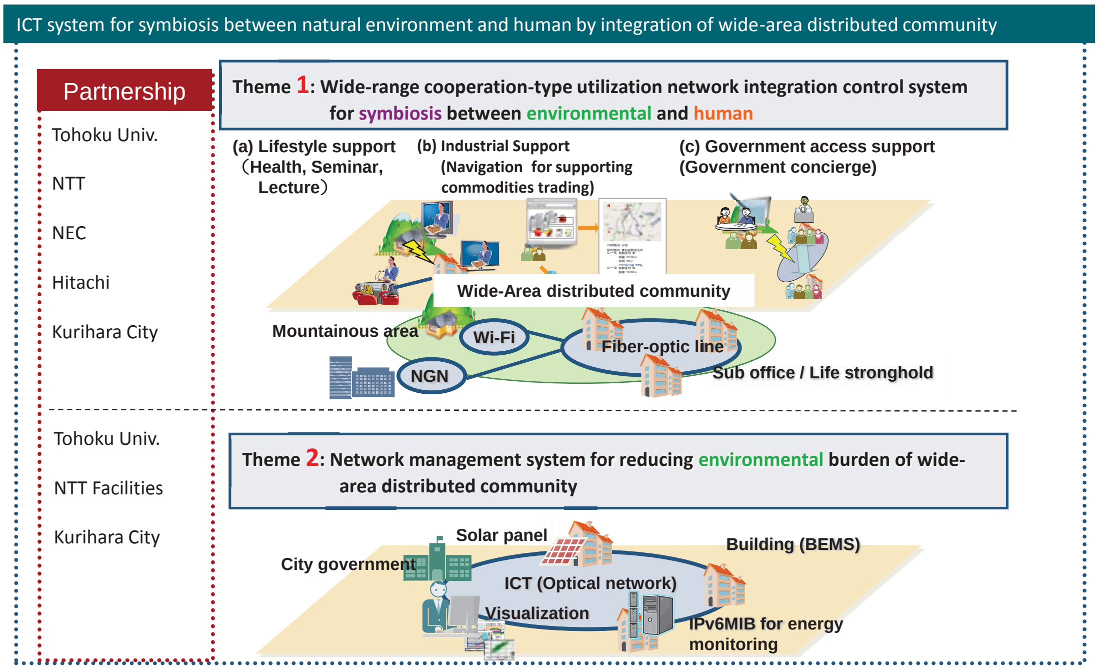
2. Summary of the project demonstration plan
Experiment on the region to construct ICT system for symbiosis between human and natural environment by integration of Life stronghold/City function decentralizes wide-area (wide-area distributed community) in Kurihara city of Miyagi prefecture, Japan.

3. Project Members
・Leader: Prof. Norio Shiratori (Waseda University / Tohoku University)
・Collaborative organizations：
NTT East, NTT Facilities, Hitachi East Japan Solutions, Mitsubishi Research Institute, Miyagi prefecture’s Kurihara city

4. Budget
・257 million yen (2,570 thousand US dollar) for 2010

Application of ICT to Kurihara project

Goal: Demonstration experiment in community-based field to test technical specifications about necessary communication protocol etc. to realize network integration control system aimed at community development to reduce environmental burden in the wide-area distributed community



ICT system for symbiosis between natural environment and human by integration of wide-area distributed community

Partnership

Tohoku Univ.
NTT
NEC
Hitachi
Kurihara City

Theme 1: Wide-range cooperation-type utilization network integration control system for symbiosis between environmental and human

(a) Lifestyle support （Health, Seminar, Lecture）

(b) Industrial Support (Navigation for supporting commodities trading)

(c) Government access support (Government concierge)

Tohoku Univ.
NTT Facilities
Kurihara City

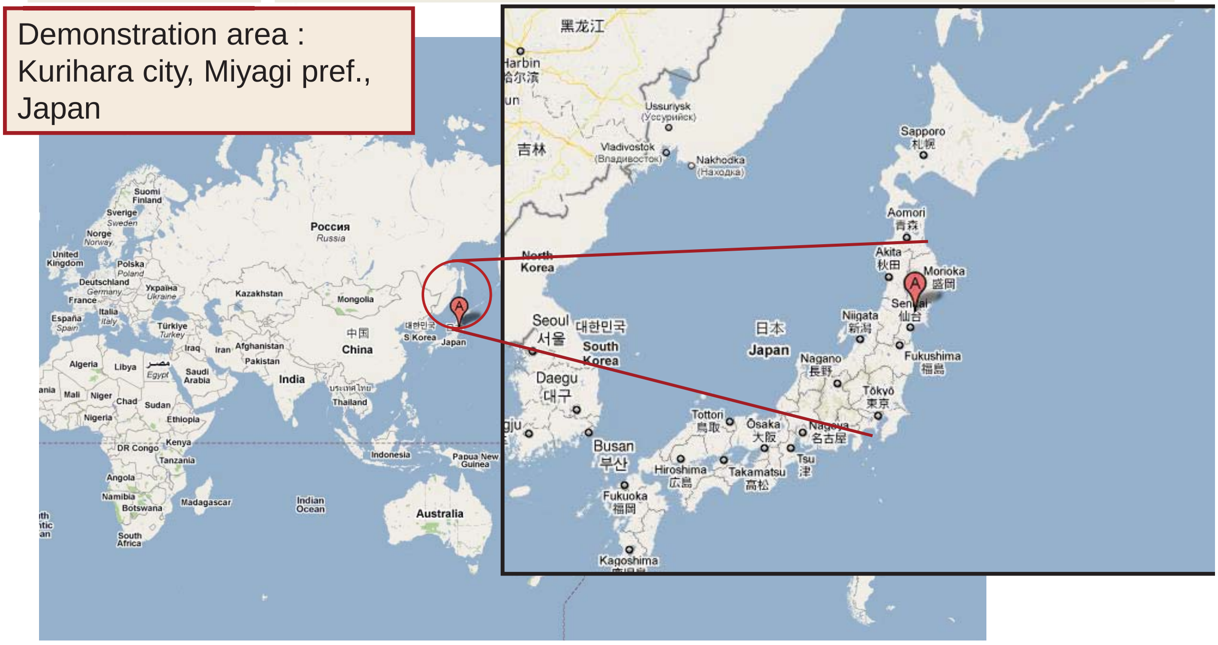
Theme 2: Network management system for reducing environmental burden of wide-area distributed community

(1) Demonstration area

Characteristics of the region

Life stronghold and City function decentralize in a wide-area
(Wide-Area distributed community)

Demonstration area :
Kurihara city, Miyagi pref., Japan

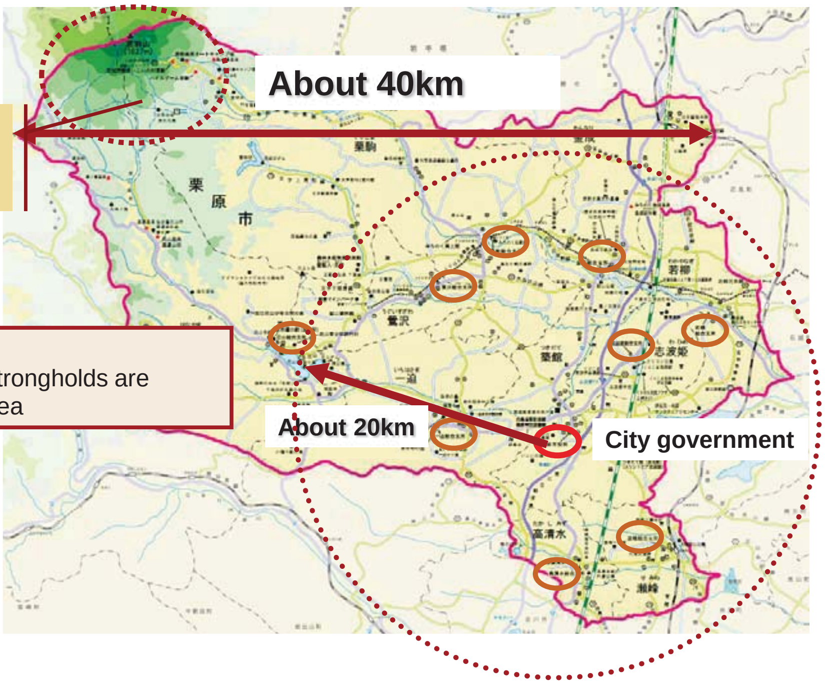


(2) Feature of demonstration area

- Population of Kurihara city: **73** thousands

City givernment

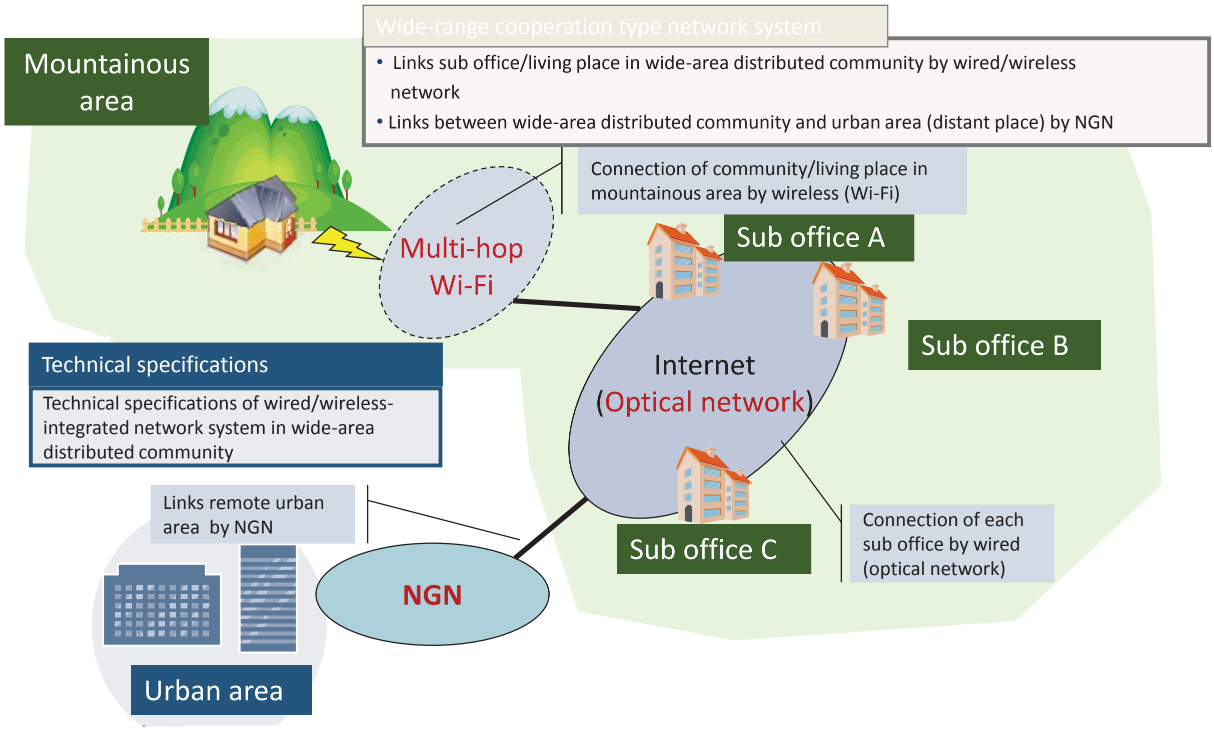
Sub-offices



Characteristics 1:
■Life stronghold in Mountainous area is far from city center and sub offices

Characteristics 2:
■ City offices and life strongholds are distributed in wide area

(3) Wide-range cooperation-type utilization network for symbiosis between environment and humans



Wide-range cooperation type network system

- Links sub office/living place in wide-area distributed community by wired/wireless network
- Links between wide-area distributed community and urban area (distant place) by NGN

Connection of community/living place in mountainous area by wireless (Wi-Fi)

Technical specifications

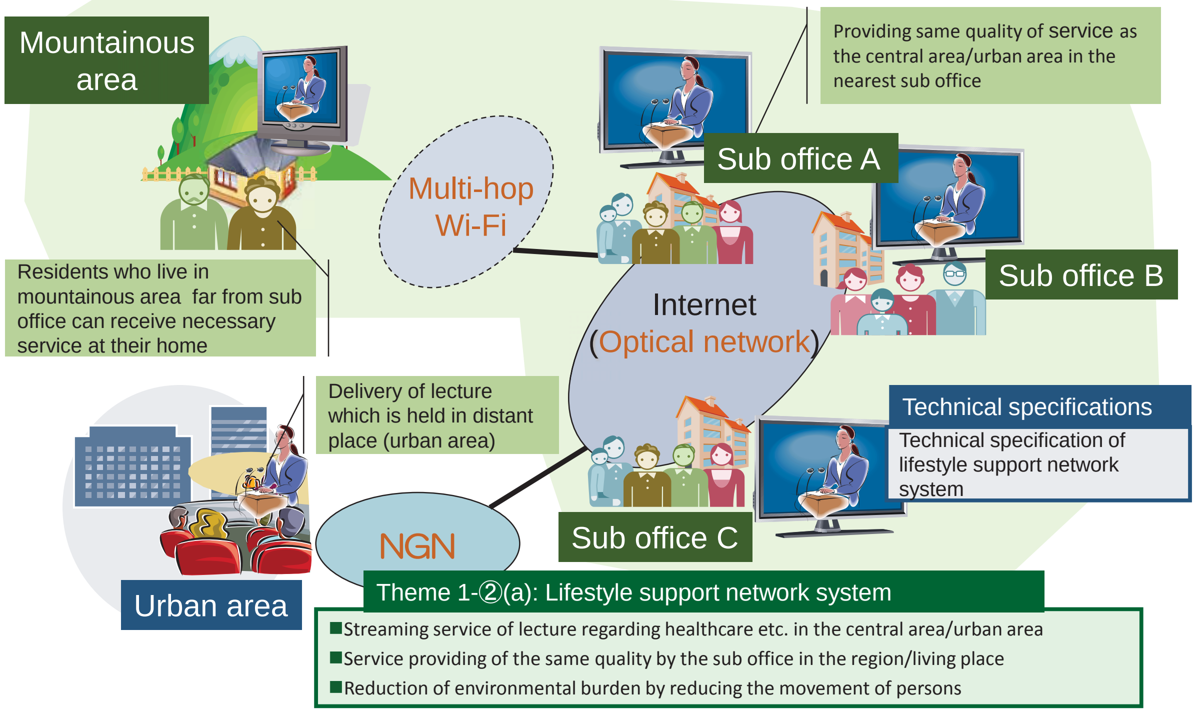
Technical specifications of wired/wireless-integrated network system in wide-area distributed community

Links remote urban area by NGN

Connection of each sub office by wired (optical network)

(4) Utilization network integration control system to integrate wide-range cooperation area

a. Life-supporting network system （healthcare guidance, seminar, lecture)



Providing same quality of service as the central area/urban area in the nearest sub office

Residents who live in mountainous area far from sub office can receive necessary service at their home

Delivery of lecture which is held in distant place (urban area)

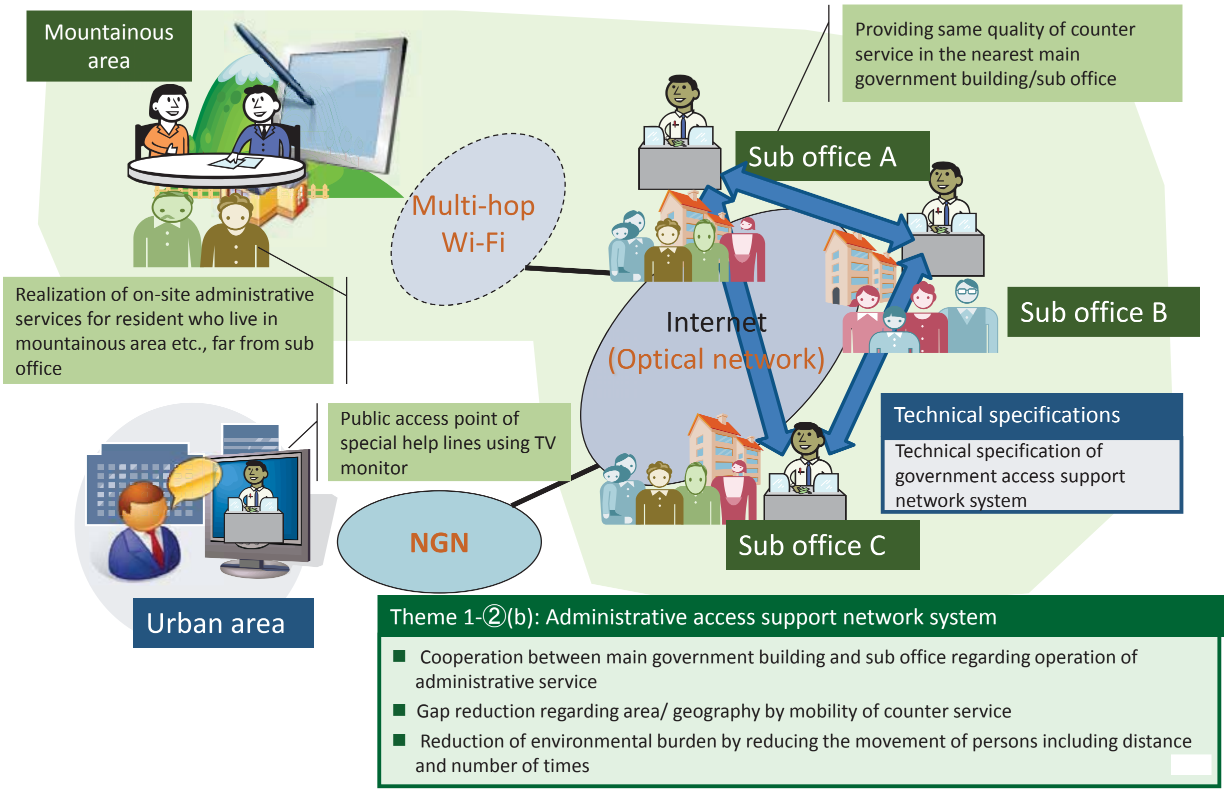
Technical specifications

Technical specification of lifestyle support network system

Theme 1-②(a): Lifestyle support network system

■Streaming service of lecture regarding healthcare etc. in the central area/urban area
■Service providing of the same quality by the sub office in the region/living place
■Reduction of environmental burden by reducing the movement of persons

b. Administrative access support network system (Administrative concierge)



Providing same quality of counter service in the nearest main government building/sub office

Realization of on-site administrative services for resident who live in mountainous area etc., far from sub office

Public access point of special help lines using TV monitor

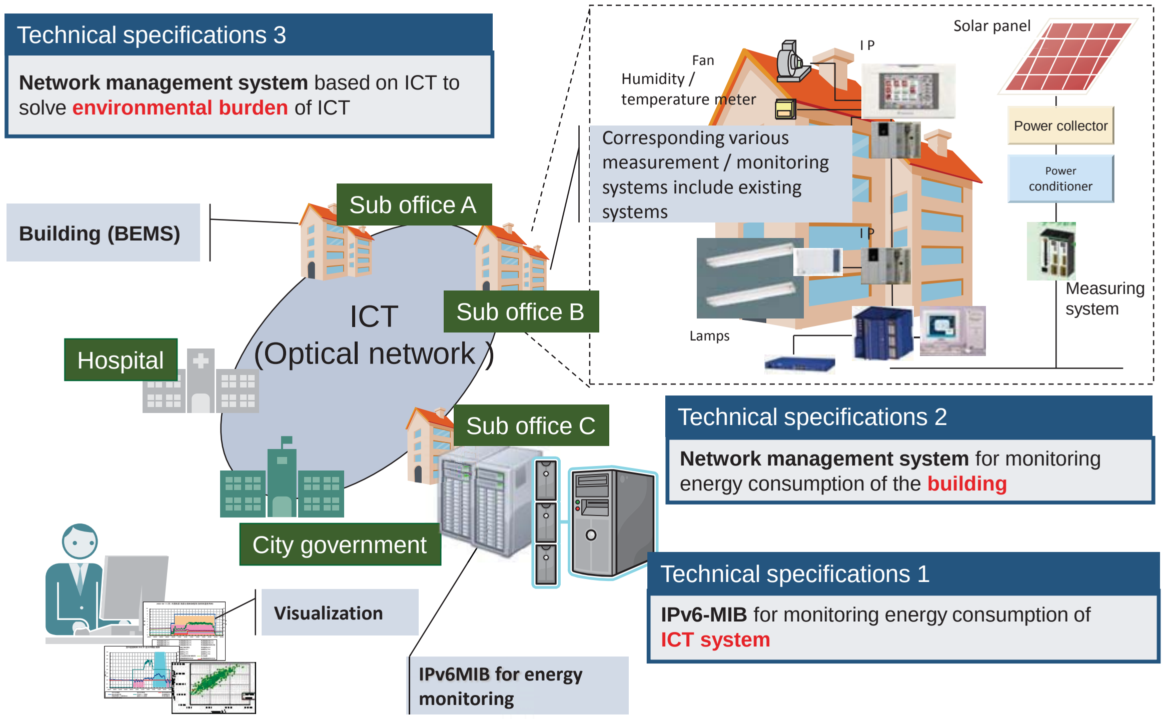
Technical specifications

Technical specification of government access support network system

Theme 1-②(b): Administrative access support network system

- Cooperation between main government building and sub office regarding operation of administrative service
- Gap reduction regarding area/ geography by mobility of counter service
- Reduction of environmental burden by reducing the movement of persons including distance and number of times

(5) Network management system for reducing environmental load of the distributed widespread area



Technical specifications 3

Network management system based on ICT to solve **environmental burden** of ICT

Corresponding various measurement / monitoring systems include existing systems

Technical specifications 2

Network management system for monitoring energy consumption of the **building**

Technical specifications 1

IPv6-MIB for monitoring energy consumption of **ICT system**

IPv6MIB for energy monitoring

Part-**2:** **Green Koban** **Project** (2011-2013)

Project Title:

"Green-oriented Network Management Technology for Saving Power Consumption of Information System"

— Make ICT system Eco-friendly—

- A national project supported by MIC

(1) Leader: Norio SHIRATORI,
(Waseda University / Tohoku University)

(2) Research Partners: NTT EAST corp.
Cyber Solutions Inc.
Tohoku Institute of Technology

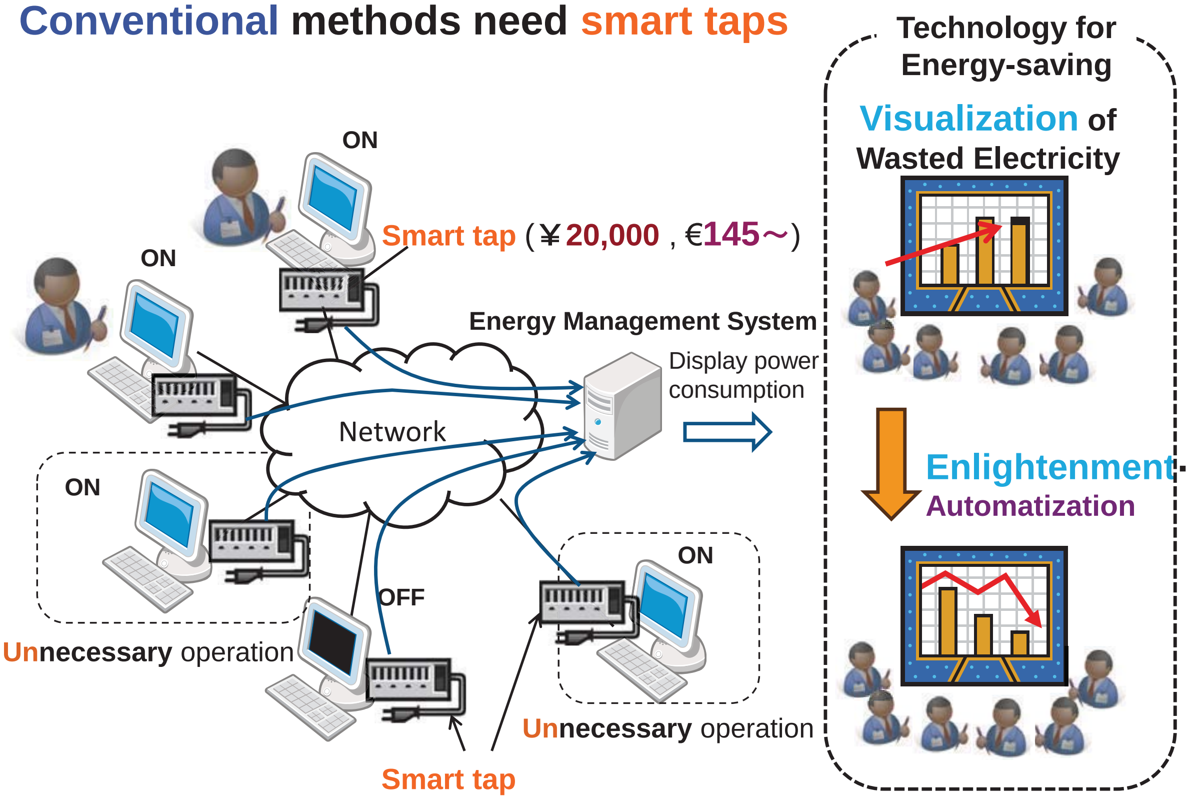
(3) Budget: Approx. 150 million yen (1.5 million US dollar)

(4) Sponsor: MIC (Ministry of Internal Affairs and Communications)

(5) Period: 2011-2013 (for 3 yrs.)

< Energy Saving Technology >

- **Conventional methods need smart taps**


Technology for Energy-saving
Visualization of Wasted Electricity
Enlightenment·Automatization
ON
ON
Smart tap (￥20,000 , €145～)
Energy Management System
Display power consumption
Network
ON
OFF
ON
Unnecessary operation
Unnecessary operation
Smart tap

10 thousand terminals : No. of Terminal Devices at Tohoku Univ. Seiryo Campus

→ **300 million yen for Smart taps**

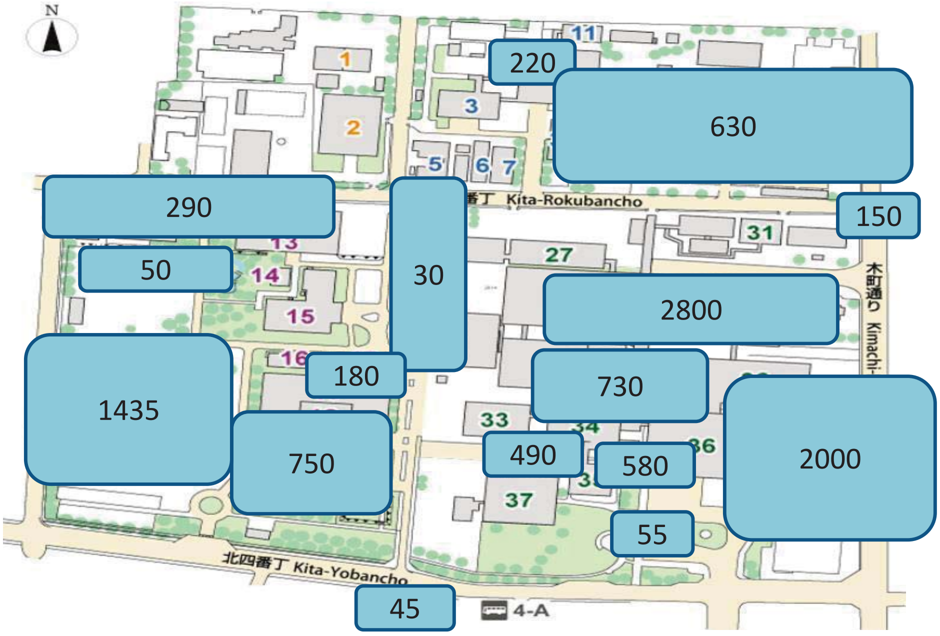
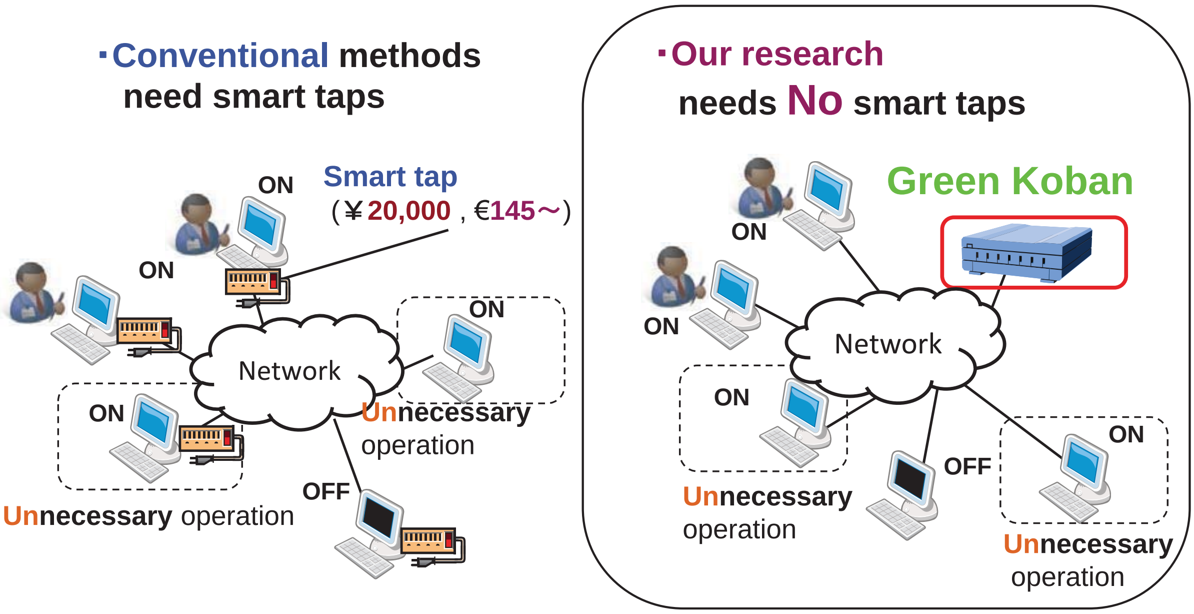


Figure. Seiryo Campus Network

Green Koban needs No smart taps


・Conventional methods need smart taps

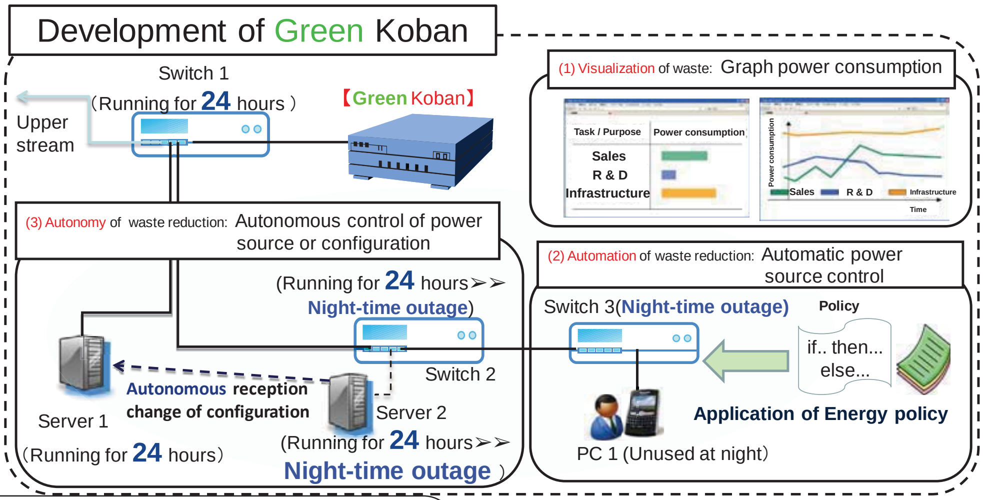
Smart tap (￥20,000 , €145～)

・Our research needs No smart taps

Outline of Green Koban

Purpose

Development of management technology to reduce power consumption of the entire network ⇒ Reducing the CO2 emissions to **10 ~ 35**%



Development of Green Koban

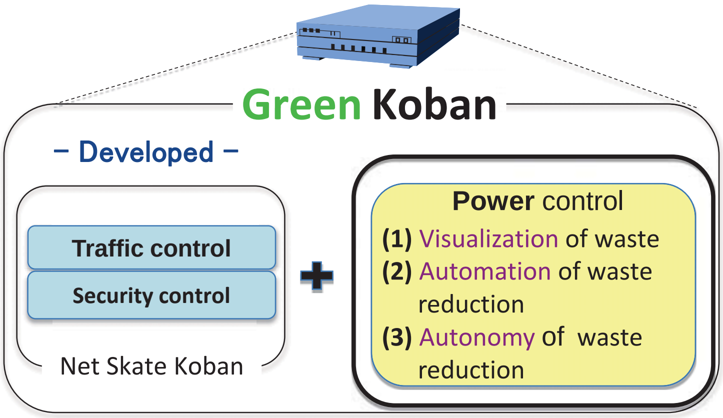
R & D of management technology toward power-saving related network
(1) Visualization of waste
(2) Automation of waste reduction
(3) Autonomy of waste reduction

Result → Management technology toward power-saving related network system (By 2013, reducing the CO2 emissions to 35%.)

Significance → It will be possible to reduce not only power consumption of a device, but also the entire network system.

Design of Green Koban

- Three major characteristic for G-K



Power control
(1) Visualization of waste
(2) Automation of waste reduction
(3) Autonomy of waste reduction

(1) Visualization

a) Acquisition/monitoring of environmental burden of information system

b) Analysis of environmental burden

c) Visualization of environmental burden, display of reception change information

(3) Autonomy

g) Reduction by autonomous system's plan

h) Optimization of network configuration autonomously

(2) Automation

d) Reduction by manager's plan

e) Introduction of energy policy

f) Automatic reduction based on energy policy

(1) Development of Green Koban which reduces Introduction costs under $\frac{1}{10}$

- Calculated under the following condition:
  Smart Tap : ￥20,000 (€145) per unit , Labor Charges : ￥ 10,000 (€ 72) per unit

Comparison Table for Introduction Costs

Size of Network	No. of Devices	**Conventional** Method （**Needs** Smart Taps）	**Our Researcn** (**Green Koban**)	Costs which can be reduced
Small (1 Foor of small & medium sized companies)	10	¥300,000	¥50,000	under $\frac{1}{10}$
Middle (Whole bdg. of small & medium sized companies)	100	¥3,000,000	¥100,000	
Large (Whole bdg. of Univ. Hospital)	5,000	¥150,000,000	¥12,000,000	

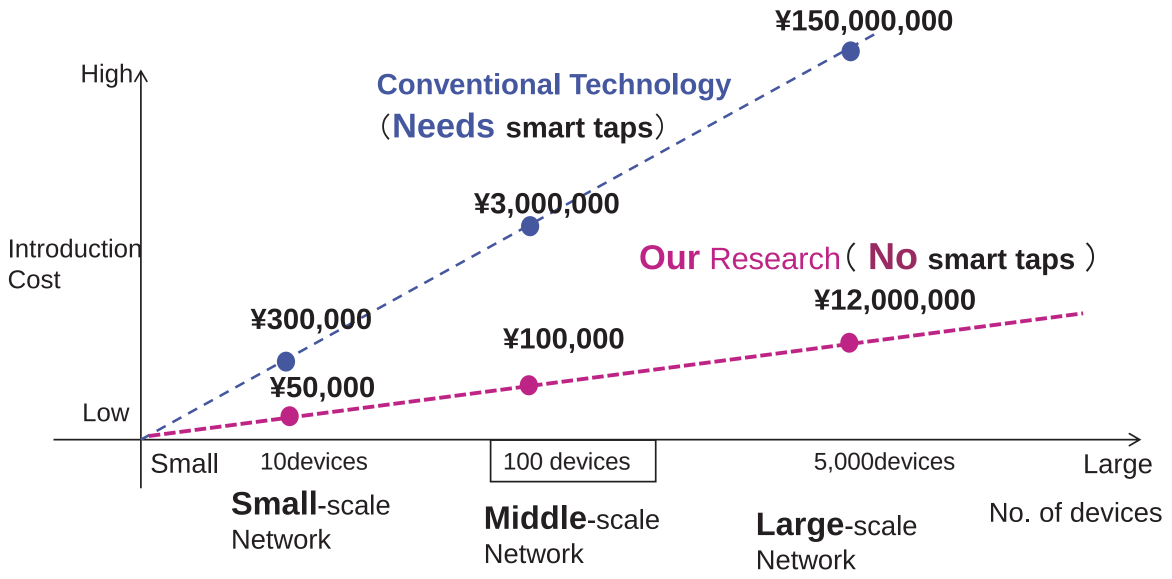
- Our research achievement : Energy-saving technology which does not need expensive smart taps
  - Even under a large-scale network environment, energy-saving can be realized easily with **low cost**.

Propagation effects on smart city-related fields such as HEMS and BEMS are quite big

HEMS: Home Energy Management System, Energy management system for homes
BEMS : Building Energy Management System, Energy management system for whole buildings

Points of Our Research Achievements

- **Conventional technology** : Needs to introduce expensive devices, **smart taps**.
  → The larger-scale network environments need to pay more expensive introduction costs.
- **Our Research** : **Green Koban** (**No** smart taps)
  → Even under a large-scale network environment, energy-saving can be realized easily with **low cost.**

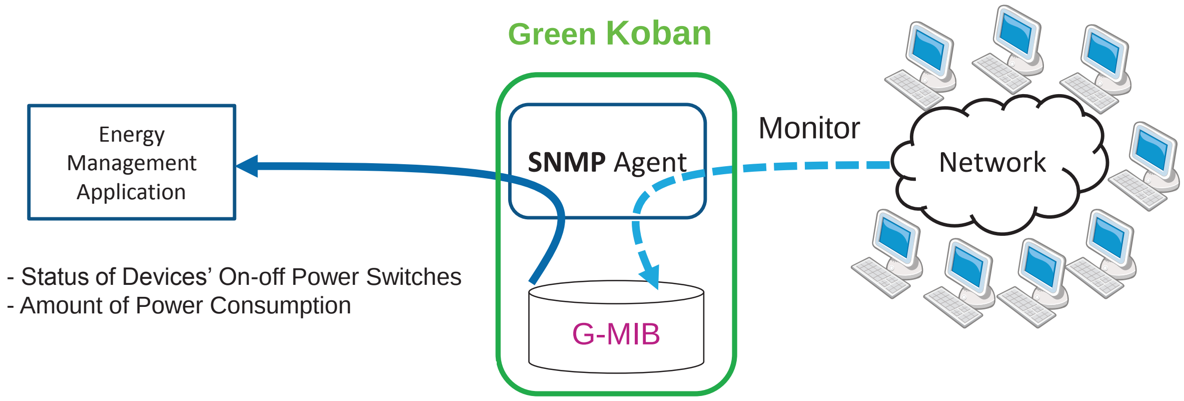


(2) Mechanism of Green Koban applying for international/ national patent

- Patent-pending Technology
  - A technology to build a green architecture which efficiently reduces power consumption based on network access records in the network.

- Monitors devices’ connection status to networks such as the starting and ending time of connection
- Calculates power consumption and status of devices’ on-off power switches based on connection records of G-MIB*, then manages energy by G-MIB

* G-MIB (Green-Management Information Base)



- Status of Devices’ On-off Power Switches
- Amount of Power Consumption

Energy monitoring

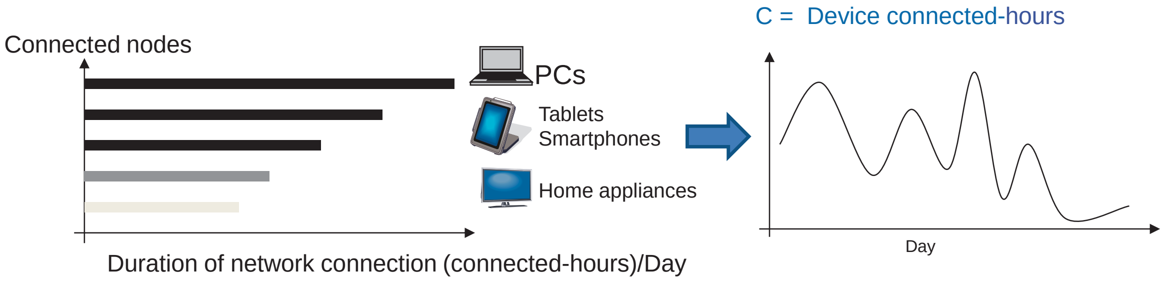
1. Direct Monitoring: Smart devices

devices equipped for energy monitoring

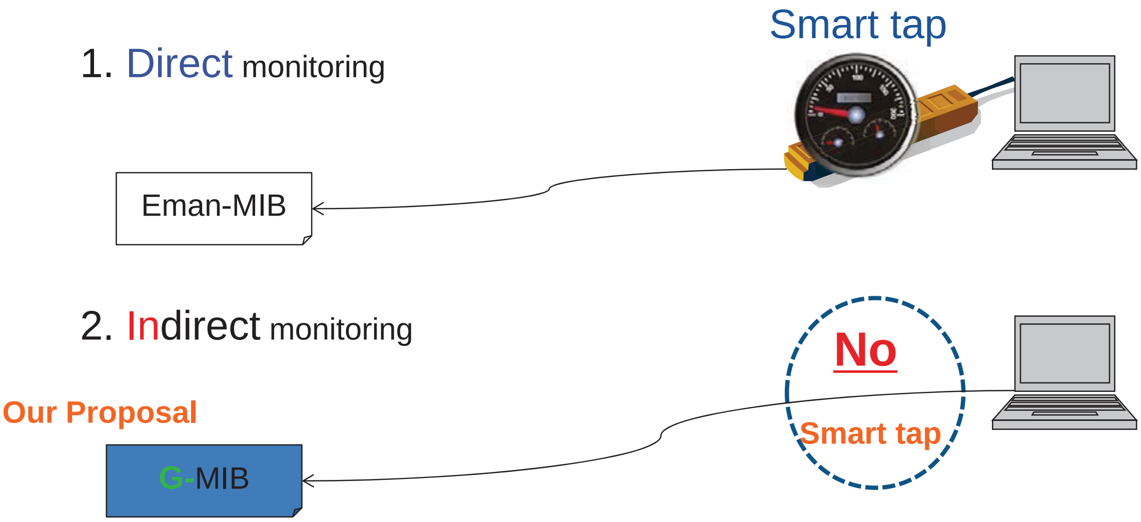
W = Kilo Watt hours
Too expensive !!
Smart Tap
High-end network devices

2. Indirect Monitoring: Non smart devices

・For devices NOT equipped energy monitoring function



Energy management framework
(EMAN WG @ IETF)



1. Direct monitoring

Smart tap

Eman-MIB

2. Indirect monitoring

Our Proposal

G-MIB

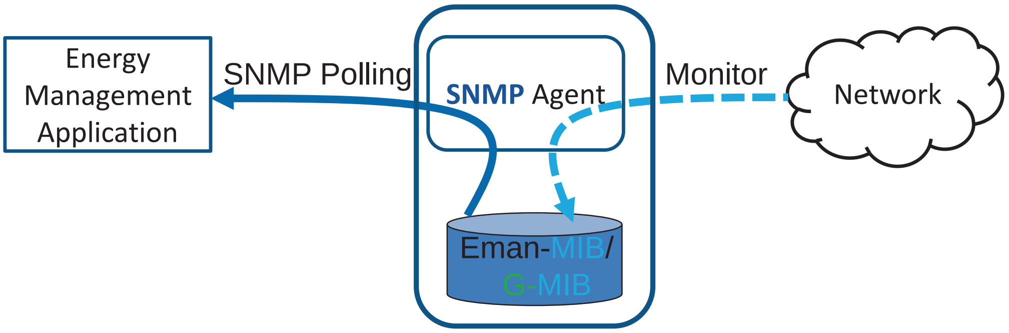
No Smart tap

Provide energy consumption
via
well-standardized and deployed protocol
SNMP

Energy management framework
(EMAN WG @ IETF)

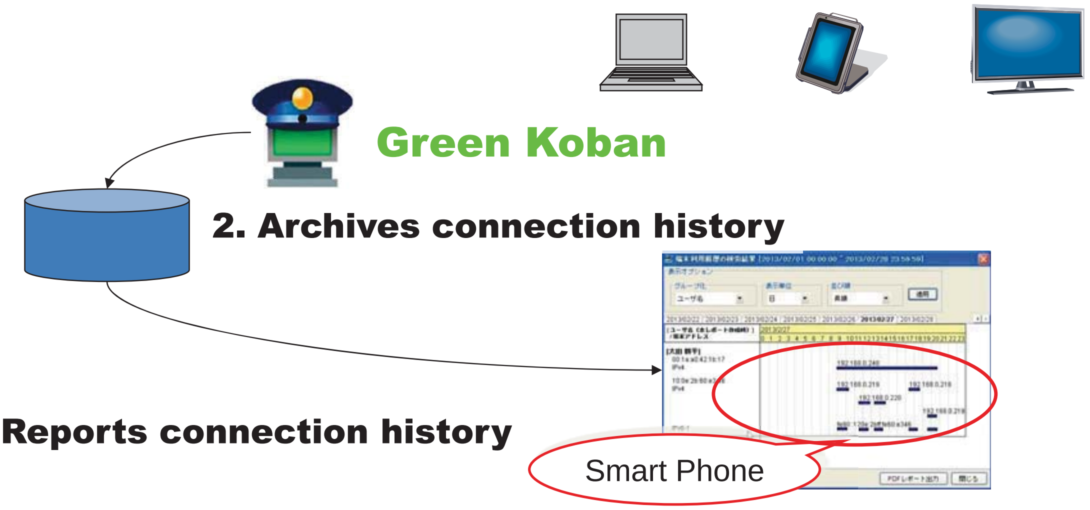
- **Uses the SNMP management protocol.**
- **Management Information Bases (MIB) is defined:**
  - **Eman-MIB** (draft-ietf-eman-energy-monitoring-mib-06 .txt): **(currently) uses direct monitoring**

Our proposal : G-MIB (draft-suganuma-greenmib-02.txt):
uses indirect monitoring



Applications: Geen Koban

1. **Monitors connectivity of ALL networked devices**



Green Koban

2. **Archives connection history**

3. **Reports connection history**

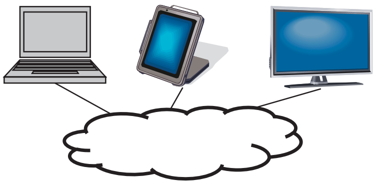
Connected-hours is a measure of energy consumption

Green Koban

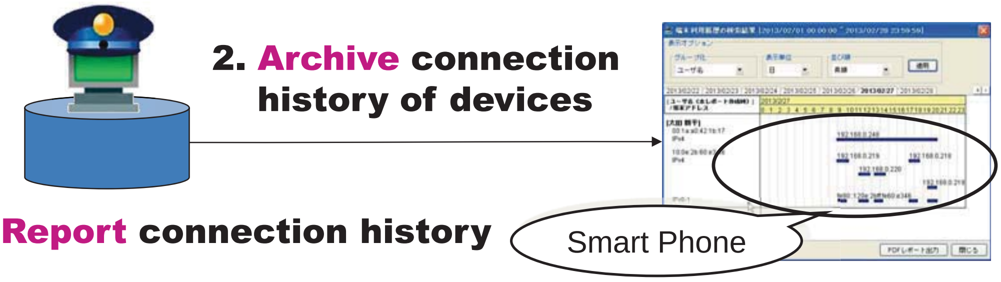
— How to monitor connectivity of devices —

1. **Monitor the connectivity of ALL the networked devices using ARP packets**

ARP packets :
- Basic protocol for IP communication
- Any devices broadcast ARP packets to get address of target devices when the devices are connected to network
- The device keeps the addresses in ARP table.
- If the address was not used during a certain period of time, the address is deleted from the ARP table.



Green Koban



2. **Archive connection history of devices**

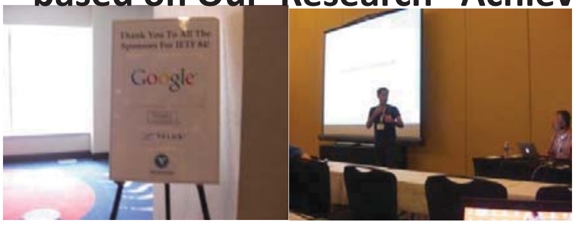
3. **Report connection history**

Green Koban can provide energy information via indirect monitoring

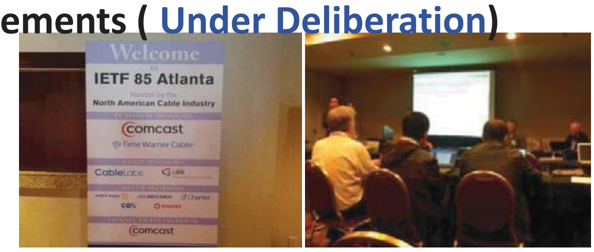
ARP packet

- ARP (Address Resolution Protocol)
  - Basic protocol for IP communication
  - Any devices broadcast ARP packets to get address of target devices when the devices are connected to network
  - The device keeps the addresses in ARP table.
  - If the address was not used during a certain period of time, the address is deleted from the ARP table.

(4) Proposal of International Standardization to IETF based on Our Research Achievements (Under Deliberation)



1). IETF 84 Meeting@Vancouver



2). IETF 85 Meeting@Atlanta

3). IETF 86 Meeting@Orlando

4). IETF 87 Meeting@Berlin

Thank You To All The Sponsors For IETF 88!
HUAWEI
cable
TELUS
Green Usage Monitoring Information Base

5). IETF 88 Meeting@Vancouver

6). IETF 89 Meeting@London

Picture: Attended to IETF Conferences for Intnl. Standardization (6 times/ 3 yrs.)

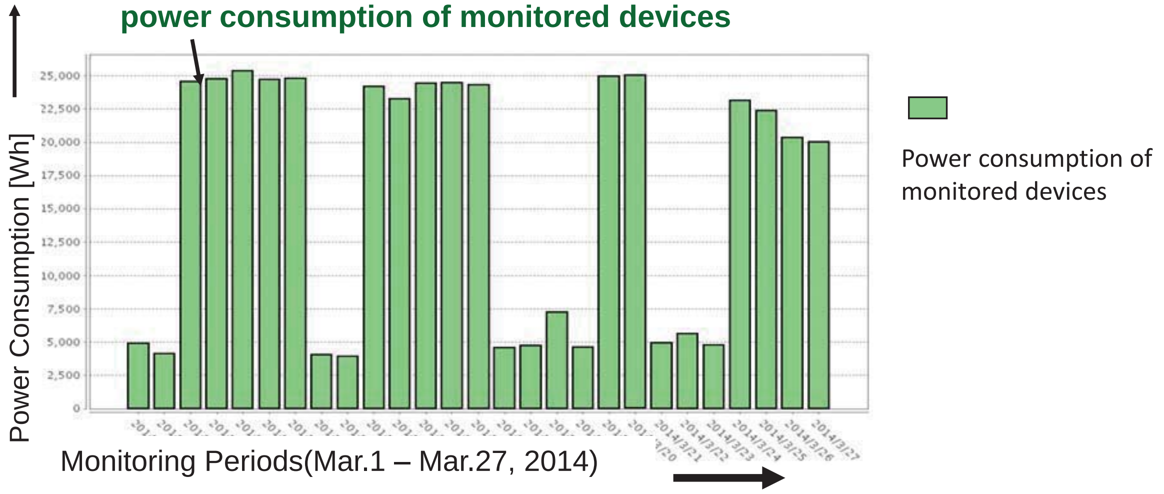
Visualization of Wasted Electricity with Green Koban

— Demonstration Experiments using a large-scale Tohoku Univ.
campus network which has about 5,000 terminal devices—

＜Descriptions of Demonstration Experiments for Visualization of **Wasted** electric power＞

1) Visualization of **reducible**(**wasted**) power consumption of Terminal Devices which are not used more than **1** hour

2) Visualization **of reducible**(**wasted**) power consumption of Terminal Devices which are not used more than **30** min.

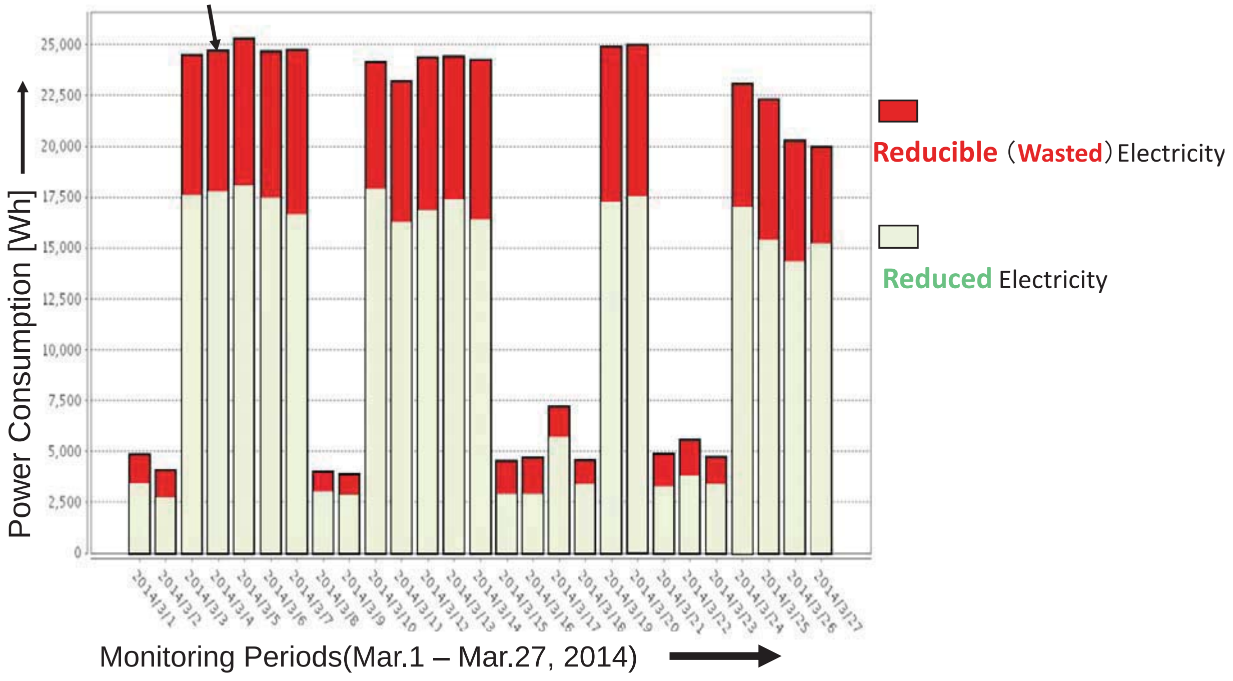


power consumption of monitored devices

Details of Wasted Electricity

1）Power Consumption of Terminal Devices which are not used more than 1 hour

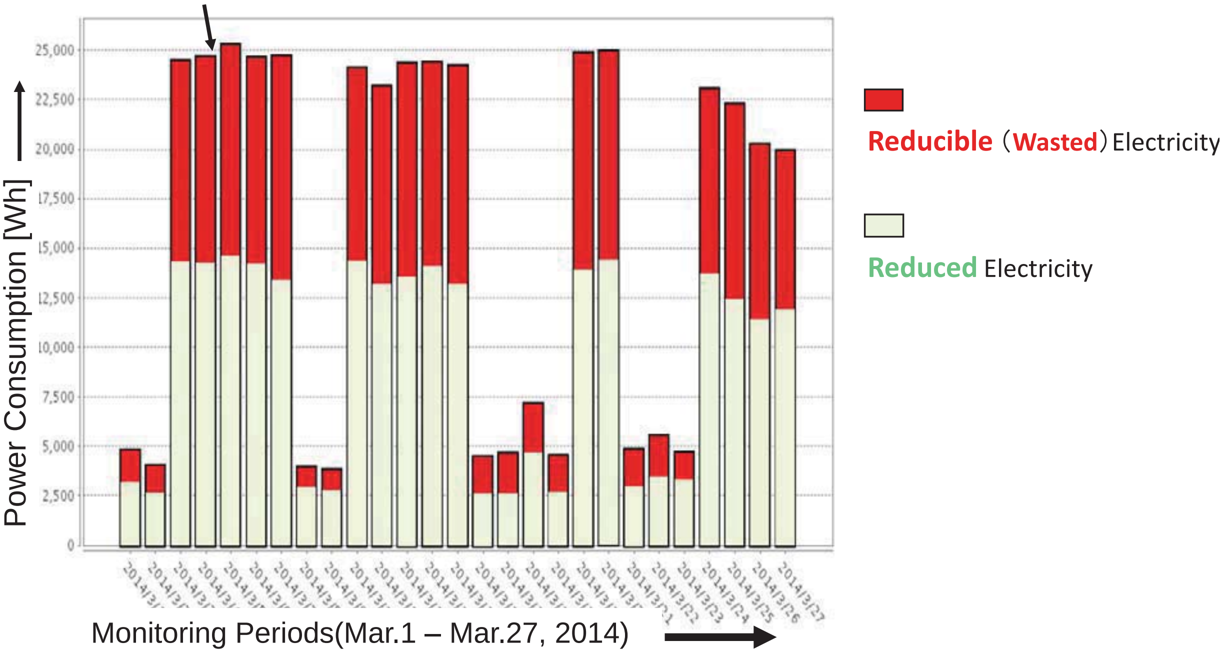
Wasted electricity which can be reduced



Details of Wasted Electricity

2) Power Consumption of Terminal Devices which are not used more than 30 min.

Wasted electricity which can be reduced



(7) Our Developed Visualization Technology

ー Evaluation of Green Koban ー

Comparison with Conventional Technology

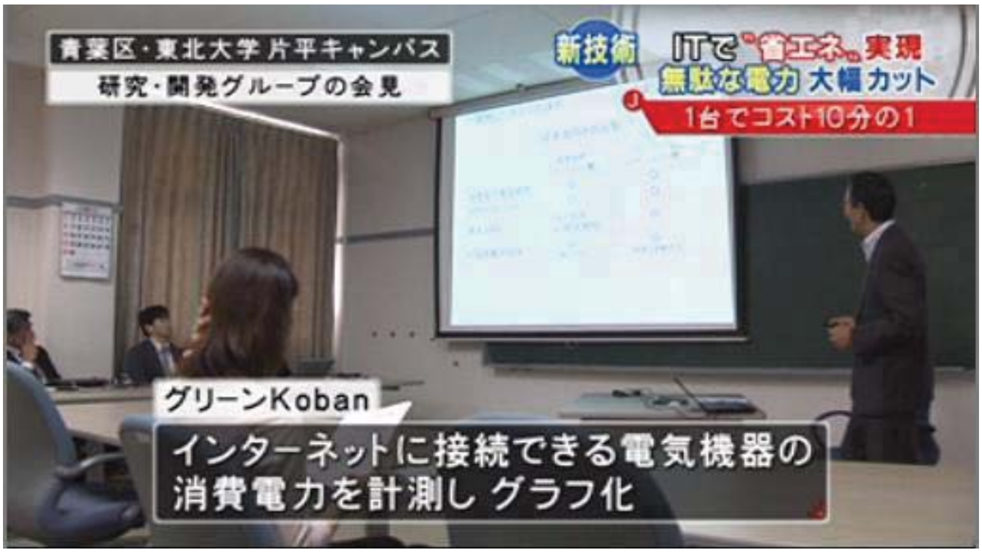
	Conventional Technology	Green Koban
	Needs smart taps	**No** smart taps
Estimate Accuracy of Power Consumption	◎	○
Scalability	○	○
Introduction Cost	△(**Small**-scale) × (**Medium**- & **Large**-scale)	◎
Acquisition of Additional Info. - Prohibit unauthorized info. acquisition	△ (only with electric power)	◎ (Status of devices' on-off power switches)

スーパーJチャンネル
みやぎ

東北発の新技術
ネットワークの消費電力↓

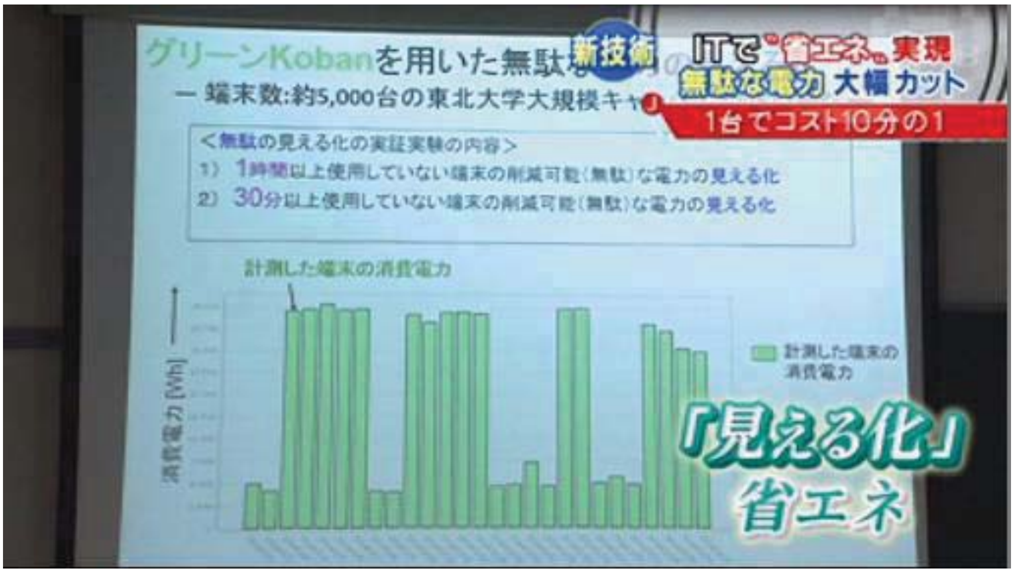
新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1

新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
グリーンKoban
▶東北大学で4機関が共同開発
▶電気の消費量を測定


青葉区・東北大学 片平キャンパス
研究・開発グループの会見
新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
グリーンKoban
インターネットに接続できる電気機器の
消費電力を計測し グラフ化

青葉区・東北大学 片平キャンパス
研究・開発グループの会見
新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
グリーンKoban
インターネットに接続できる電気機器の
消費電力を計測し グラフ化

新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
グリーンKoban
1時間以上未使用 ➡ "無駄な電力"と表示



グリーンKobanを用いた無駄

ー 端末数:約5,000台の東北大学大規模キャ

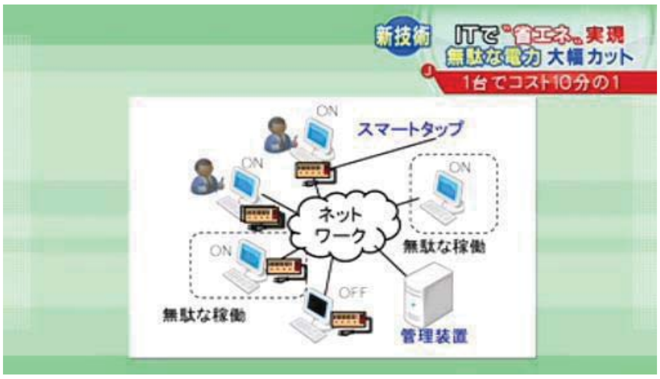
＜無駄の見える化の実証実験の内容＞

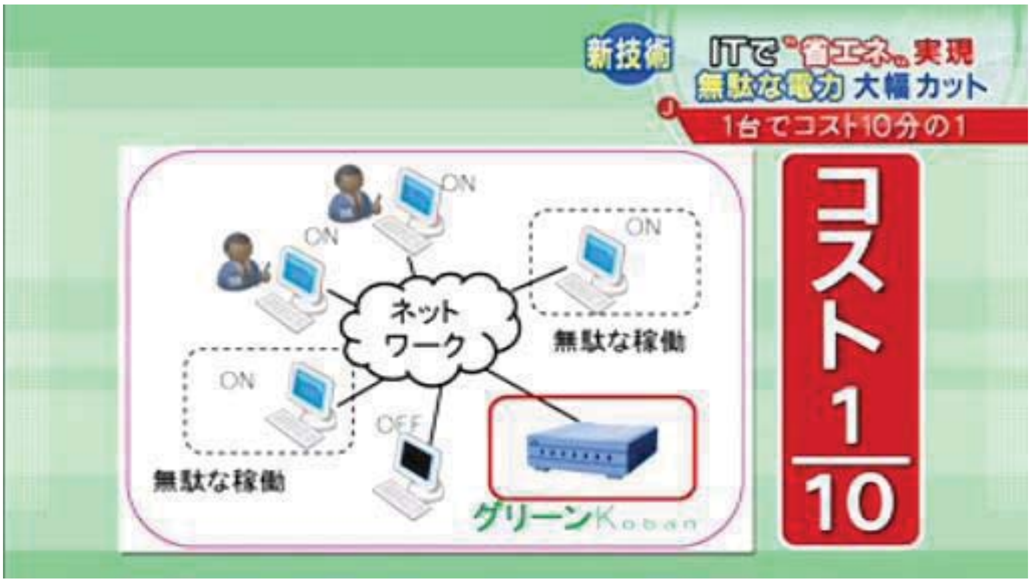
1） 1時間以上使用していない端末の削減可能（無駄）な電力の見える化

2） 30分以上使用していない端末の削減可能（無駄）な電力の見える化

スマートタップ

●個々電気製品の消費電力を測定




新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
ON
ON
ON
ネット
ワーク
無駄な稼働
ON
OFF
無駄な稼働
グリーンKoban
コスト1/10

新技術
ITで"省エネ"実現
無駄な電力 大幅カット
1台でコスト10分の1
東北大学
白鳥 則郎 名誉教授
人間と同様 自立的にコンピューターが判断
電源のオンオフ制御が最終的なゴール

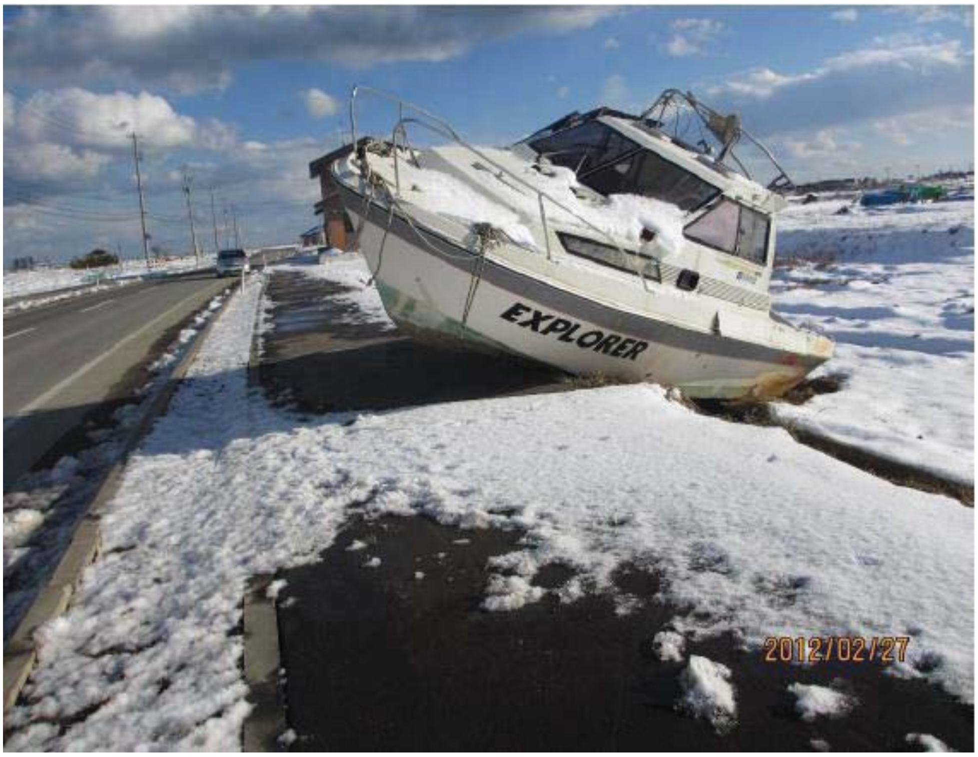
2. NDN : Never Die Networks

< Miserable Experiences of 2 Big Earthquakes >

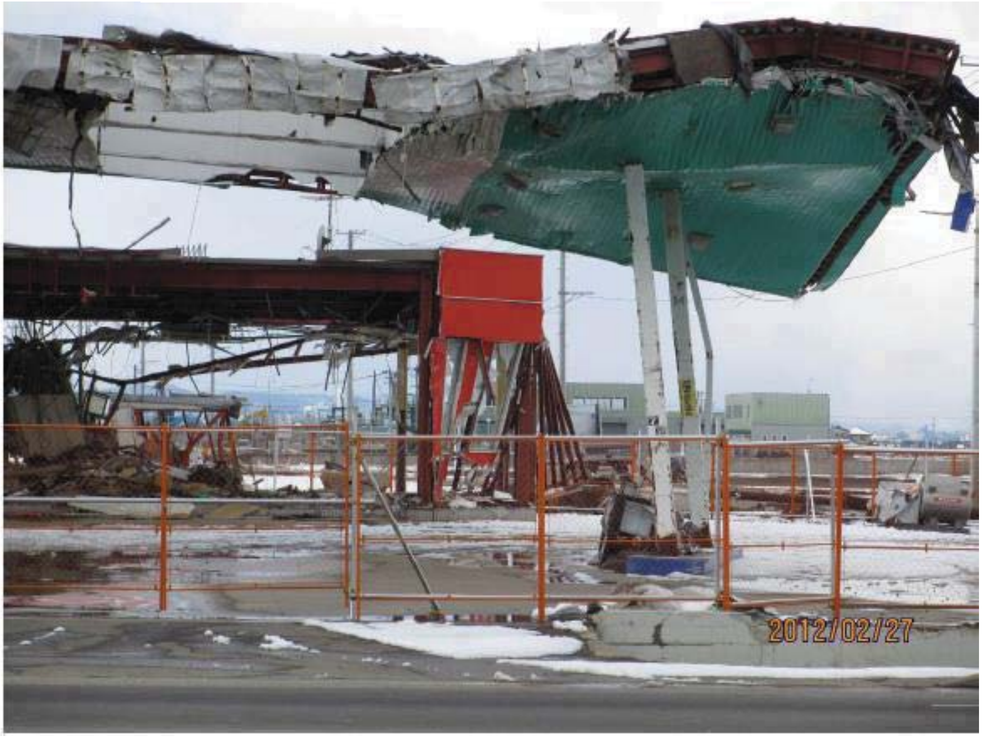
1) “Miyagi-oki Earthquake” (June, 1978)

2) “Great East Japan Earthquake”(March, 2011)

Prefectural road recovered by removing tsunami rubble in Miyagi Prefecture



Boat swept by tsunami about 2 km away from the sea in Miyagi Prefecture



Damaged gas station by tsunami located about 1 km away from coast in Miyagi Prefecture

2.1 June 12, 1978 : Miyagi-oki Earthquake

- Home: Tomiya town, Miyagi prefecture
- Office (My Lab.) : Tohoku University, Center of Sendai City

 —**12 km** from the office to my home, **40 min**. by car

Unable to make a contact from office to my home

⇒ **Could not confirm the safety** of my gravid wife who supposed to stay at home

If there was a way to hear her voice only a few seconds,
I could know the safety of my wife

Confirmation of Safety

Within **72** hours
(Survival Potential)

<Original Concept of Never Die Networks (NDN)>

1) Real-time services
2) Ability to support massively large number of simultaneous users
3) Offer such services for a short duration (at the very least 20 seconds, 30seconds)

■ **2003** : Proposal of "**Never Die Networks**"

■ **2007** : Adopted to Grants-in-Aid for Scientific Research on Challenging Exploratory Research("JSPS "Kakenhi")

2.2 March 11, 2011 ：Great East Japan Earthquake

Started to compete with others
in conducting the researches on
“Disaster-resilient Network”

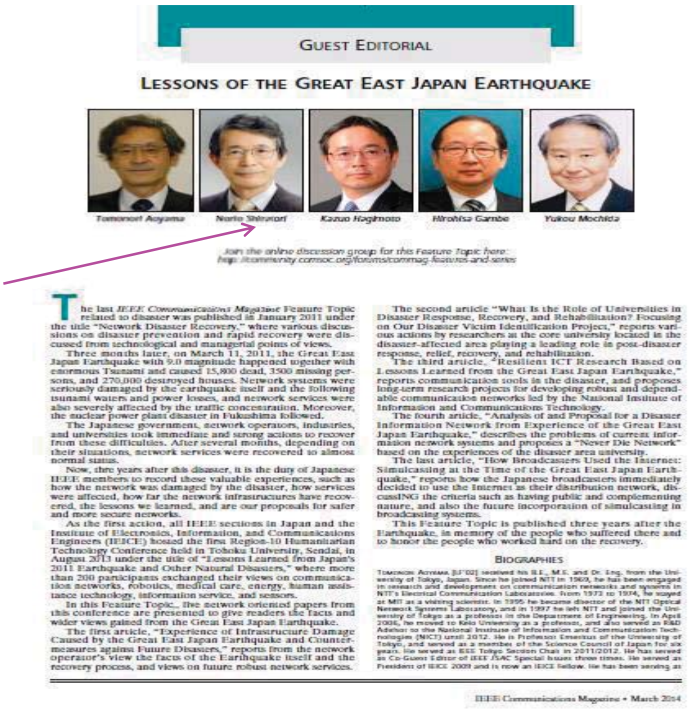
▪Industry−Government−Academia
started the researches in chorus

- IEEE HTC 2013 in Sendai（August, 2013）
— International Conference themed on “Disaster”—

3 Presidents of IEEE attended
−Former, Present and Next Presidents−

IEEE Communications Magazine • March 2014

LESSONS OF THE GREAT EAST JAPAN EARTHQUAKE



GUEST EDITORIAL

LESSONS OF THE GREAT EAST JAPAN EARTHQUAKE

Tomonori Aoyama Norio Shiratori Kazuo Hagimoto Hirohisa Gambe Yukou Mochida

Join the online discussion group for this Feature Topic here: http://community.comsoc.org/forums/commag-features-and-series

The last *IEEE Communications Magazine* Feature Topic related to disaster was published in January 2011 under the title “Network Disaster Recovery,” where various discussions on disaster prevention and rapid recovery were discussed from technological and managerial points of views.

Three months later, on March 11, 2011, the Great East Japan Earthquake with 9.0 magnitude happened together with enormous Tsunami and caused 15,800 dead, 3500 missing persons, and 270,000 destroyed houses. Network systems were seriously damaged by the earthquake itself and the following tsunami waters and power losses, and network services were also severely affected by the traffic concentration. Moreover, the nuclear power plant disaster in Fukushima followed.

The Japanese government, network operators, industries, and universities took immediate and strong actions to recover from these difficulties. After several months, depending on their situations, network services were recovered to almost normal status.

Now, thre years after this disaster, it is the duty of Japanese IEEE members to record these valuable experiences, such as how the network was damaged by the disaster, how services were affected, how far the network infrastructures have recovered, the lessons we learned, and are our proposals for safer and more secure networks.

As the first action, all IEEE sections in Japan and the Institute of Electronics, Information, and Communications Engineers (IEICE) hosted the first Region-10 Humanitarian Technology Conference held in Tohoku University, Sendai, in August 2013 under the title of “Lessons Learned from Japan's 2011 Earthquake and Other Natural Disasters,” where more than 200 participants exchanged their views on communication networks, robotics, medical care, energy, human assistance technology, information service, and sensors.

In this Feature Topic,, five network oriented papers from this conference are presented to give readers the facts and wider views gained from the Great East Japan Earthquake.

The first article, “Experience of Infrastructure Damage Caused by the Great East Japan Earthquake and Countermeasures against Future Disasters,” reports from the network operator's view the facts of the Earthquake itself and the recovery process, and views on future robust network services.

The second article “What Is the Role of Universities in Disaster Response, Recovery, and Rehabilitation? Focusing on Our Disaster Victim Identification Project,” reports various actions by researchers at the core university located in the disaster-affected area playing a leading role in post-disaster response, relief, recovery, and rehabilitation.

The third article, “Resilient ICT Research Based on Lessons Learned from the Great East Japan Earthquake,” reports communication tools in the disaster, and proposes long-term research projects for developing robust and dependable communication networks led by the National Institute of Information and Communications Technology.

The fourth article, “Analysis of and Proposal for a Disaster Information Network from Experience of the Great East Japan Earthquake,” describes the problems of current information network systems and proposes a “Never Die Network” based on the experiences of the disaster area university.

The last article, “How Broadcasters Used the Internet: Simulcasting at the Time of the Great East Japan Earthquake,” reports how the Japanese broadcasters immediately decided to use the Internet as their distribution network, discussING the criteria such as having public and complementing nature, and also the future incorporation of simulcasting in broadcasting systems.

This Feature Topic is published three years after the Earthquake, in memory of the people who suffered there and to honor the people who worked hard on the recovery.

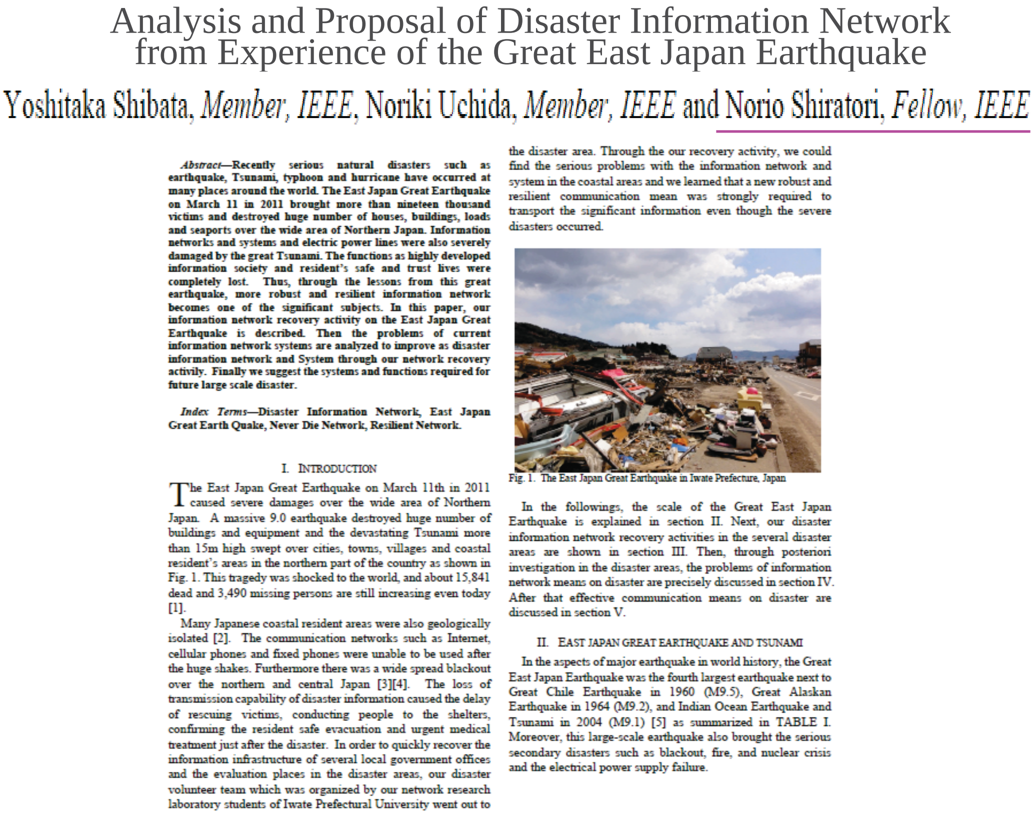
BIOGRAPHIES

TOMONORI AOYAMA [LF'02] received his B.E., M.E. and Dr. Eng. from the University of Tokyo, Japan. Since he joined NTT in 1969, he has been engaged in research and development on communication networks and systems in NTT's Electrical Communication Laboratories. From 1973 to 1974, he stayed at MIT as a visiting scientist. In 1995 he became director of the NTT Optical Network Systems Laboratory, and in 1997 he left NTT and joined the University of Tokyo as a professor in the Department of Engineering. In April 2006, he moved to Keio University as a professor, and also served as R&D Advisor to the National Institute of Information and Communication Technologies (NICT) until 2012. He is Professor Emeritus of the University of Tokyo, and served as a member of the Science Council of Japan for six years. He served as IEEE Tokyo Section Chair in 2011/2012. He has served as Co-Guest Editor of IEEE JSAC Special Issues three times. He served as President of IEICE 2009 and is now an IEICE Fellow. He has been serving as

IEEE Communications Magazine • March 2014

Norio
Shiratori

IEEE Communications Magazine, March 2014



Analysis and Proposal of Disaster Information Network from Experience of the Great East Japan Earthquake

Yoshitaka Shibata, *Member, IEEE*, Noriki Uchida, *Member, IEEE* and Norio Shiratori, *Fellow, IEEE*

***Abstract*—Recently serious natural disasters such as earthquake, Tsunami, typhoon and hurricane have occurred at many places around the world. The East Japan Great Earthquake on March 11 in 2011 brought more than nineteen thousand victims and destroyed huge number of houses, buildings, loads and seaports over the wide area of Northern Japan. Information networks and systems and electric power lines were also severely damaged by the great Tsunami. The functions as highly developed information society and resident's safe and trust lives were completely lost. Thus, through the lessons from this great earthquake, more robust and resilient information network becomes one of the significant subjects. In this paper, our information network recovery activity on the East Japan Great Earthquake is described. Then the problems of current information network systems are analyzed to improve as disaster information network and System through our network recovery activily. Finally we suggest the systems and functions required for future large scale disaster.**

***Index Terms*—Disaster Information Network, East Japan Great Earth Quake, Never Die Network, Resilient Network.**

I. Introduction

The East Japan Great Earthquake on March 11th in 2011 caused severe damages over the wide area of Northern Japan. A massive 9.0 earthquake destroyed huge number of buildings and equipment and the devastating Tsunami more than 15m high swept over cities, towns, villages and coastal resident's areas in the northern part of the country as shown in Fig. 1. This tragedy was shocked to the world, and about 15,841 dead and 3,490 missing persons are still increasing even today [1].

Many Japanese coastal resident areas were also geologically isolated [2]. The communication networks such as Internet, cellular phones and fixed phones were unable to be used after the huge shakes. Furthermore there was a wide spread blackout over the northern and central Japan [3][4]. The loss of transmission capability of disaster information caused the delay of rescuing victims, conducting people to the shelters, confirming the resident safe evacuation and urgent medical treatment just after the disaster. In order to quickly recover the information infrastructure of several local government offices and the evaluation places in the disaster areas, our disaster volunteer team which was organized by our network research laboratory students of Iwate Prefectural University went out to the disaster area. Through the our recovery activity, we could find the serious problems with the information network and system in the coastal areas and we learned that a new robust and resilient communication mean was strongly required to transport the significant information even though the severe disasters occured.

Fig. 1. The East Japan Great Earthquake in Iwate Prefecture, Japan

In the followings, the scale of the Great East Japan Earthquake is explained in section II. Next, our disaster information network recovery activities in the several disaster areas are shown in section III. Then, through posteriori investigation in the disaster areas, the problems of information network means on disaster are precisely discussed in section IV. After that effective communication means on disaster are discussed in section V.

II. East Japan Great Earthquake and Tsunami

In the aspects of major earthquake in world history, the Great East Japan Earthquake was the fourth largest earthquake next to Great Chile Earthquake in 1960 (M9.5), Great Alaskan Earthquake in 1964 (M9.2), and Indian Ocean Earthquake and Tsunami in 2004 (M9.1) [5] as summarized in TABLE I. Moreover, this large-scale earthquake also brought the serious secondary disasters such as blackout, fire, and nuclear crisis and the electrical power supply failure.

2.3. Performance of Never Die Networks

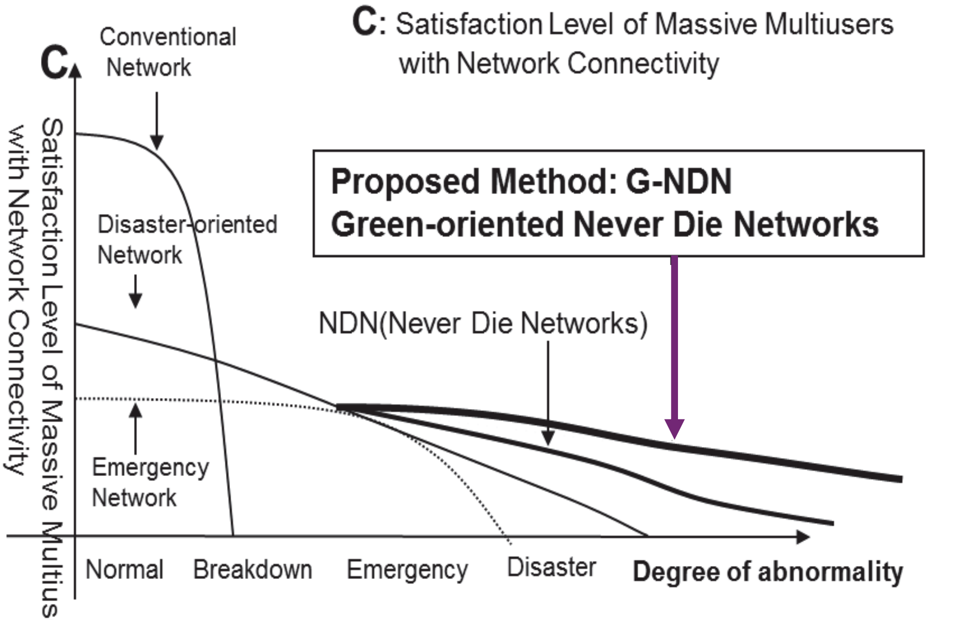


Figure 1 : Performance of Network corresponding to Degree of Abnormality: Satisfaction Level of Massive Multiusers with Network Connectivity

2.4 Architecture of Green-oriented Never Die NetworkS

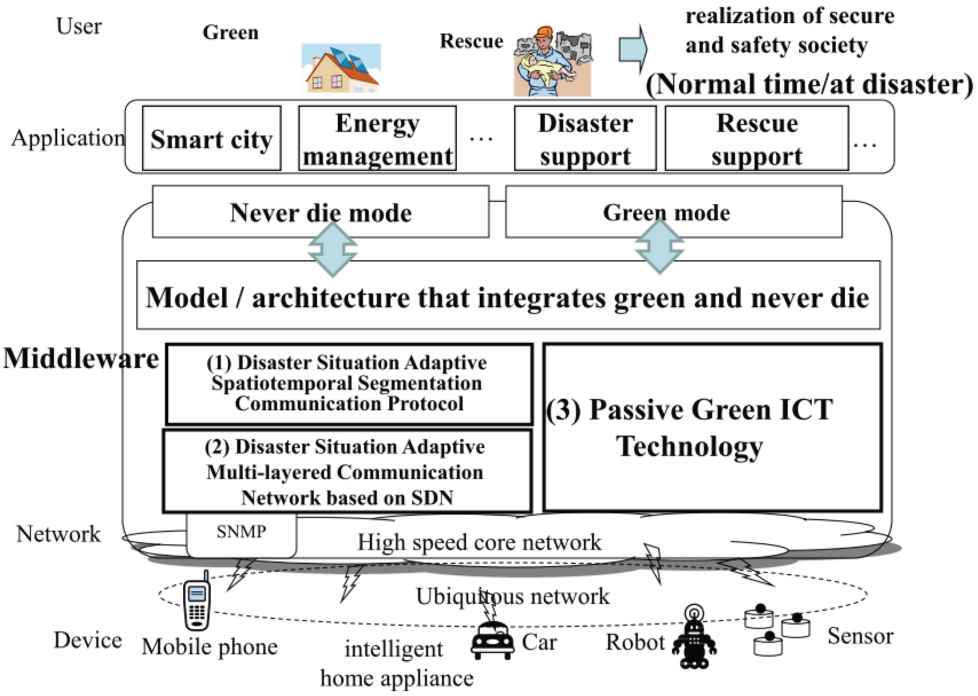

User
Green
Rescue
realization of secure and safety society
(Normal time/at disaster)
Application
Smart city
Energy management
…
Disaster support
Rescue support
…
Never die mode
Green mode
Model / architecture that integrates green and never die
Middleware
(1) Disaster Situation Adaptive Spatiotemporal Segmentation Communication Protocol
(2) Disaster Situation Adaptive Multi-layered Communication Network based on SDN
(3) Passive Green ICT Technology
Network
SNMP
High speed core network
Ubiquitous network
Device
Mobile phone
intelligent home appliance
Car
Robot
Sensor

Fig: The proposed architecture for realizing green and never die technologies

2.5. Comparison of NDN with DTN

Proposal of NDN (Never Die Networks) in 2003

a) Purpose: Realization of **disaster** communications
b) Original Concept: **Active**-type protocol
c) Hideyoshi Toyotomi (Japanese shogun)-type approach: (figuratively speaking)
“If a little **cuckoo** does not sing, I will **create** the way to **let** the bird sing.”
d) Public Attention: NDN attracted much attention in the wake of The Great East Japan Earthquake on March, 2011.

Proposal of DTN (Delay Tolerant Networks) in 2003

a) Purpose: Realization of **interplanetary** communications
b) Original Concept: **Passive**-type protocol
c) Ieyasu Tokugawa(Japanese shogun)-type approach : (figuratively speaking)
”If a little **cuckoo** does not sing, I will **wait** until the bird sings.”
d) Public Attention: DTN gained attention when NASA (National Aeronautics and Space Administration) successfully completed communication experiments with EPOXI satellite, in November, 2008.
e) Others: Toward an application of DTN to disaster communication, its researches are promoting led by Japanese researchers.

3. Trusted Cloud Computing

Technical Problems(技術的課題):

When people uses cloud services, they need to send and place their **own data & programs** to cloud servers.

クラウド利用者は、**自分のデータやプログラム**をクラウドサーバーに送信・配置する。

This leads to problems such as **unfair use of the data & programs, information leakage, etc.** by cloud service providers .

その結果、データやプログラムが**クラウド事業者**に不正使用されたり，漏洩する等 の問題が生じる.

Traditional(Existing) Security Measures in Cloud Services

クラウドサービスにおける従来のセキュリティ

＜Security by Data Encryption＞

Data Encryption is an effective method for data protection against external attacks such as unauthorized reading of data from outside of the clouds.
When encrypted data is processed in a cloud, the data are decrypted to plaintexts.
As a result, cloud providers can know the original data of the encrypted data.

＜暗号化 によるSecurity＞

データの暗号化は，クラウド外部から不正にデータを読み取ろうとする攻撃には有効である。クラウド内において、プログラム実行時，暗号化されたデータは暗号鍵によって元**のデータに復号**される。その結果、元の**平文(original data)**となり，**データ**がクラウド事業者に明らか（open)となってしまう.

・Problem 1:

Cloud service users always have possibility of risks, abuses and leakages of data by cloud providers.
e.g.) Leakage of customer information by Benesse Holdings in 2014

クラウド事業者によるデータの悪用や漏洩が生じる可能性がある。 e.g.) ベネッセの顧客情報漏洩事件

・Problem 2:

If the database and programs are deployed in the same cloud, the decryption key that is used when data are processed comes tobe known to the cloud provider.
Consequently, the provider can know the whole database.

復号時に使用した暗号鍵がクラウド事業者に知られる。この結果、処理の対象となる**データの全体**が明らかとなってしまう.

Related Researches

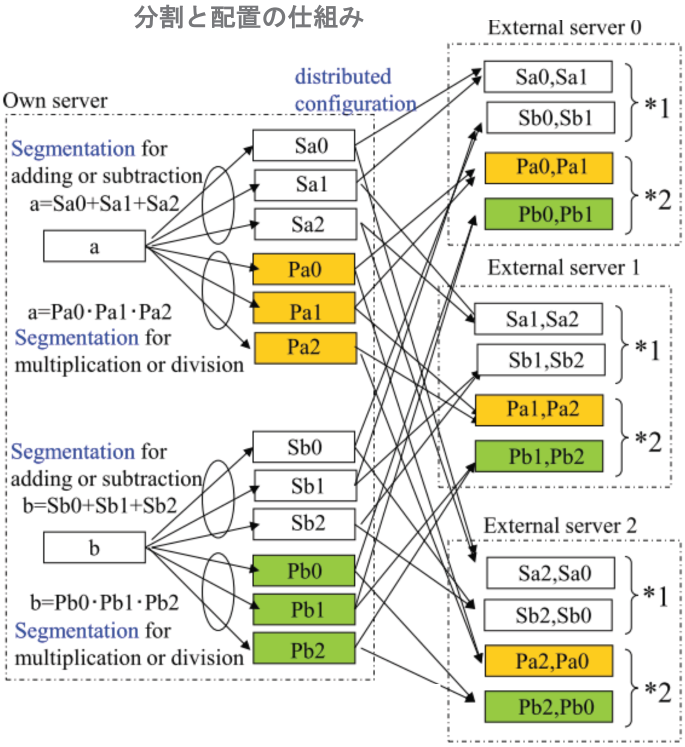
(1) Secret Sharing Methods
秘密分散法

(2) Secure Computation Methods
秘密計算法

(3) Our Research: Methodsof Dividing Data and Program
本研究：データ・プログラムの分割による方法

Proposed Method: Simple Secure Computation
簡易的秘密計算法（提案）

- Mechanism of Divisions and Deployment-
分割と配置の仕組み



*1 : Segmentation for adding or subtraction
*2 : Segmentation for multiplication or division

Addition (Existing method):
(Sa0+Sb0)+(Sa1+Sb1)+(Sa2+Sb2)
＝(Sa0+Sa1+Sa2)+(Sb0+Sb1+SPb2)
＝a+b

Subtraction (Existing method):
(Sa0-Sb0)+(Sa1-Sb1)+(Sa2-Sb2)
＝(Sa0+Sa1+Sa2)-(Sb0+Sb1+SPb2)
＝a-b

Multiplication (Proposal):
(Pa0･Pb0)･(Pa1･Pb1)･(Pa2･Pb2)
＝(Pa0･Pa1･Pa2)･(Pb0･Pb1･Pb2)
＝a･b

Division (Proposal):
(Pa0/Pb0)･(Pa1/Pb1)･(Pa2/Pb2)
＝(Pa0･Pa1･Pa2)/(Pb0･Pb1･Pb2)
＝a/b

Application Example of Proposed Method（k=2， n=2 ）−1

提案方式の応用例（k=2，n=2 ）−1

Take example of grade reports of students

学生の成績表を例にする

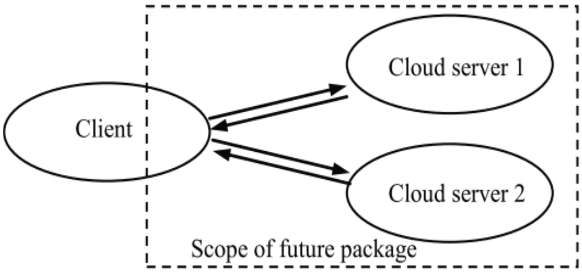


Fig. 3. System Configuration of an Application Example.

TABLE II Grade table of ten students

STUDENT	Score of subject 1	Score of subject 2
ID	x_1	x_2
1	50	80
2	40	58
3	65	62
4	73	92
5	81	50
6	45	82
7	57	51
8	78	70
9	92	68
10	62	78

Application Example of Proposed Method −3

Calculation Procedure （計算方法）：

(1) As statistical processing, it is necessary to calculate average value of each subject. Calculates total sum vertically.

(1) 統計処理として，各科目点数の平均値を求める．縦方向の総和を求めればよい．

Own Server： Requests to start the average value calculation program designating subject g（g=1ｏｒ2）, to cloud server 1 and cloud server 2.

自社サーバ：科目g（g=1ｏｒ2）を指定して，クラウド1，2に平均値計算プログラムを要請する．

Cloud1(k=1)： Calculates the total sum Xg1 vertically for designated subject g of all students' score xg1s. Then, it answers back to the client.

クラウド1(k=1)：指定科目gについて，xg1sの縦方向への総和を求め，結果を要求元（自社サーバ）に返答する．

Cloud2(k=2)： Calculates the total sum Xg2 vertically for designated subject g of all students' scores xg2s .

クラウド2(k=2)：指定科目gについて，xg2sの縦方向への総和を求め，結果を要求元（自社サーバ）に返答する．

Own Server： Add the calculation result from cloud server 1 and 2, and divide by the number of students I (in this example, I=10),then its answer E(xg) is the expected value or the averagevalue of subject g.

自社サーバ：クラウド1，クラウド2からの計算結果を加算し，学生数（この例では10）で除算して，平均値E(xg)とする．

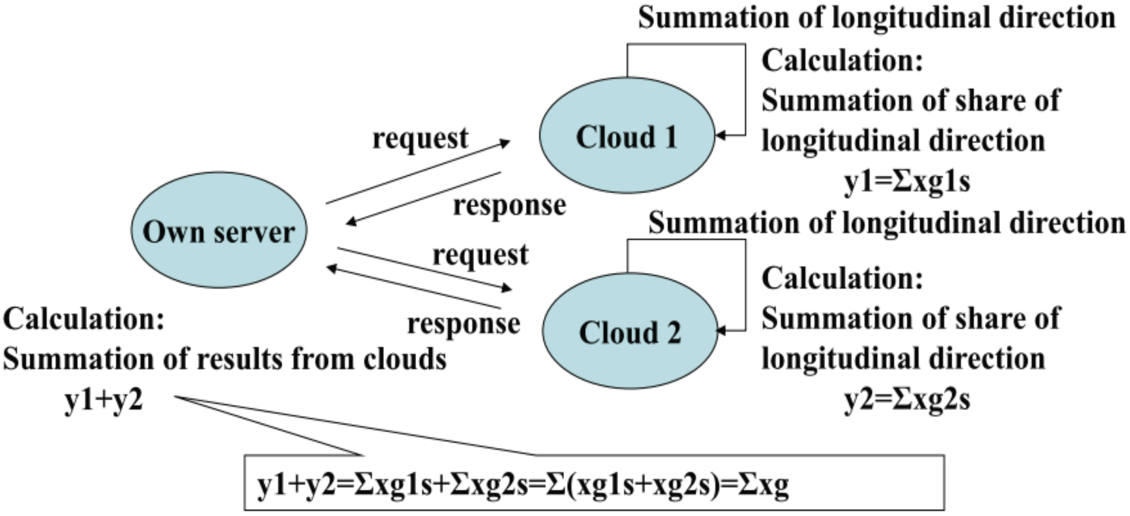
E is average value (mean value or expected value).

Application Example of Proposed Method −4

Calculation Procedure（計算方法）：

(1) As statistical processing, it is necessary to calculate average value of each subject. Calculates total sum vertically.

(1) 統計処理として，各科目点数の平均値を求める．縦方向の総和を求めればよい．

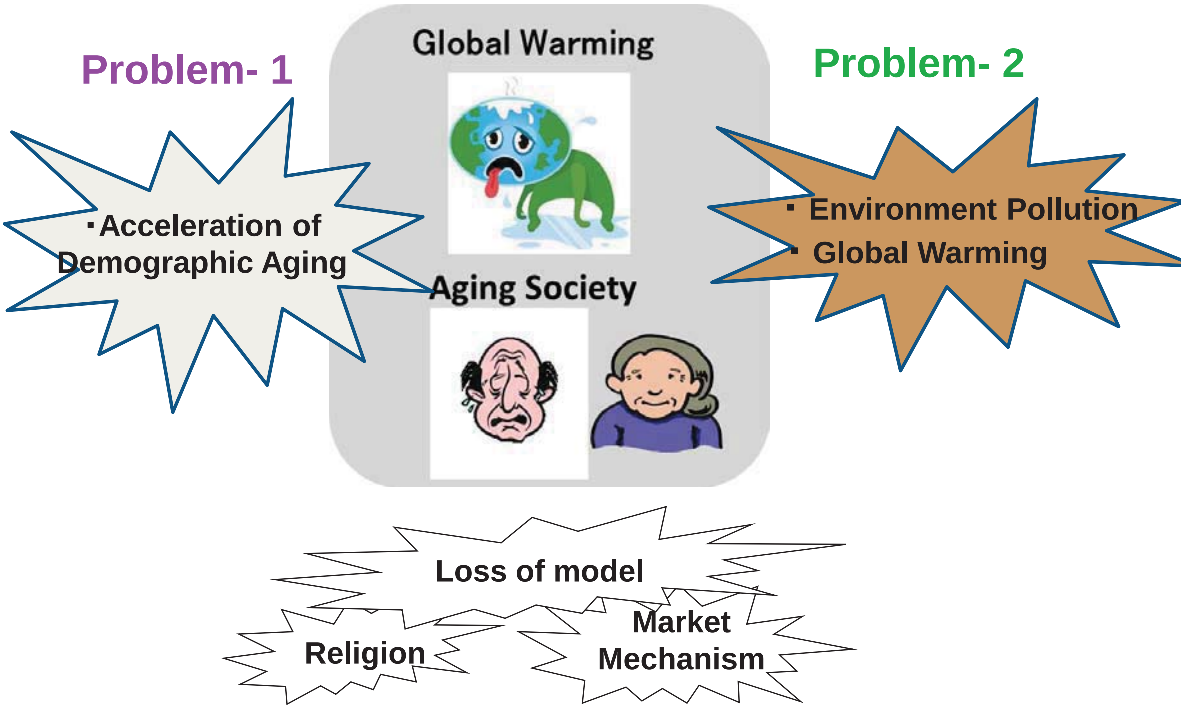


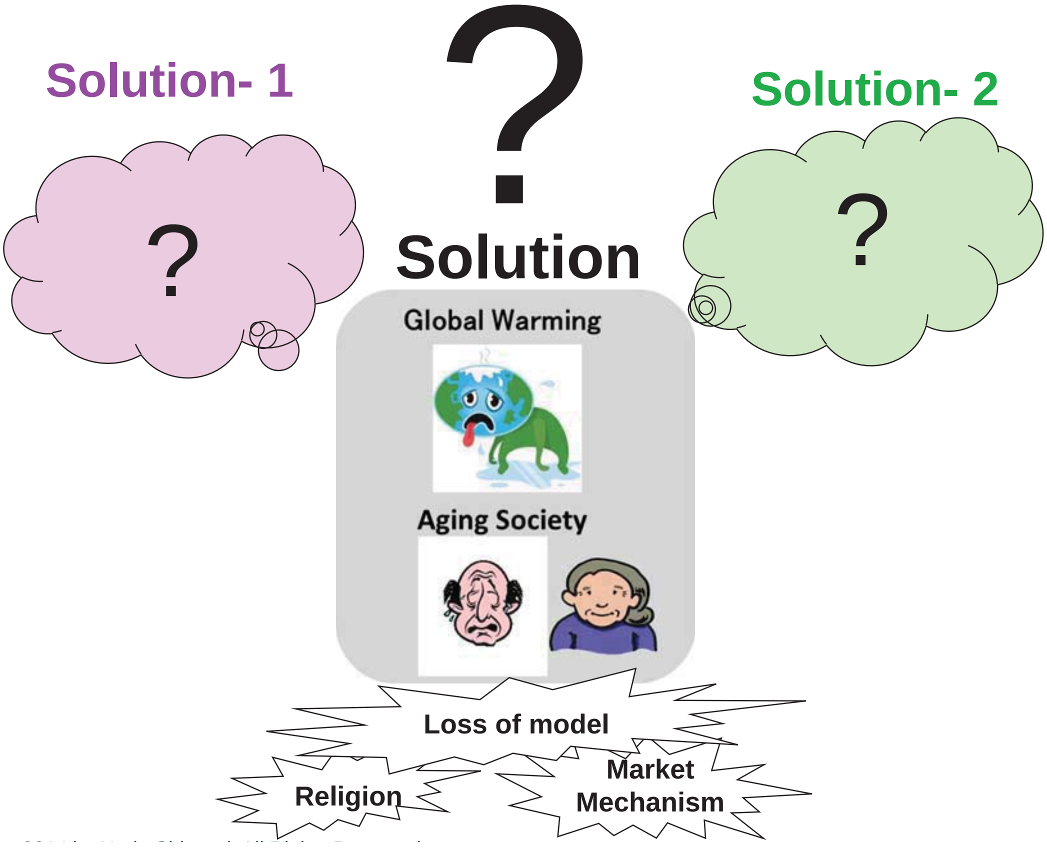
Summation of longitudinal direction
Calculation:
Summation of share of longitudinal direction
y1=Σxg1s
Cloud 1
request
response
Own server
request
response
Summation of longitudinal direction
Calculation:
Summation of share of longitudinal direction
y2=Σxg2s
Cloud 2
Calculation:
Summation of results from clouds
y1+y2
y1+y2=Σxg1s+Σxg2s=Σ(xg1s+xg2s)=Σxg

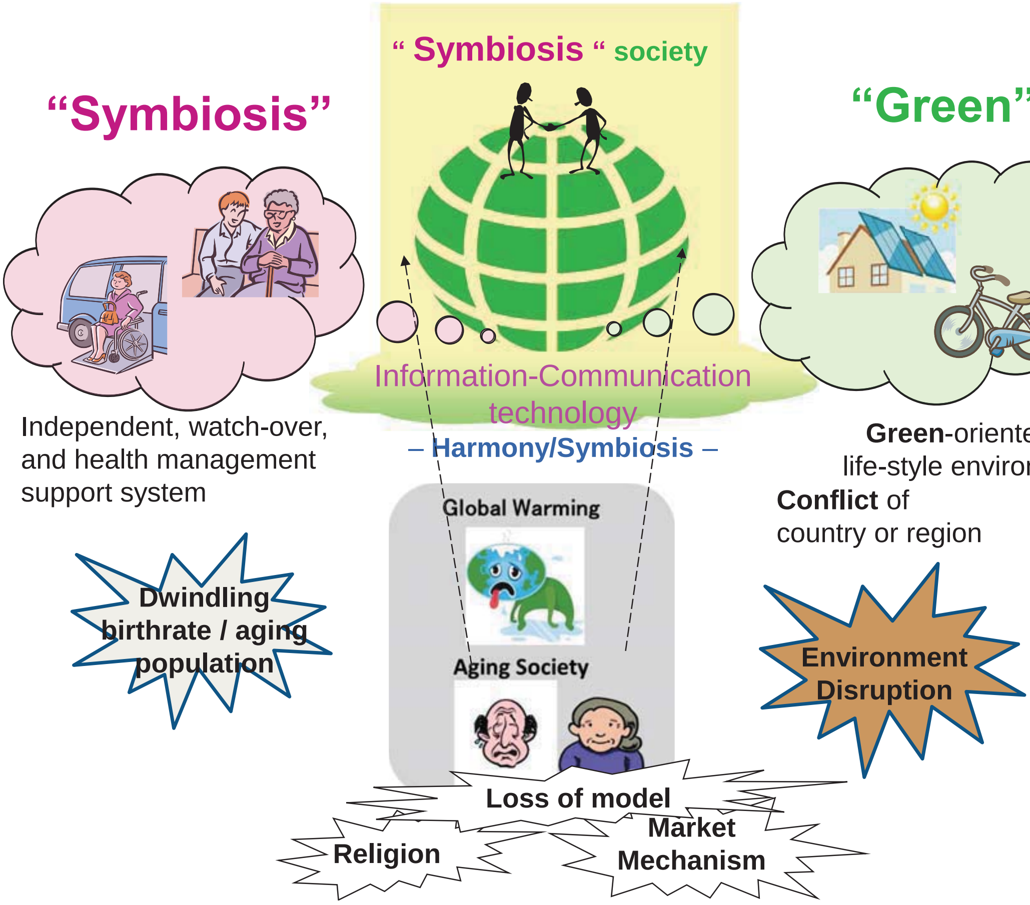
4. Conclusion

-Towards a Sustainable Information Society-

	Modern (Industrial Society)	**Solutions**: Post Modern (Symbiotic Society)
Evaluation Criterion	■ Rationality • Economy • Efficiency • Function	■ Rationality + α α : new criteria for relationship amon human, society and nature achieved by **I T** ・Our proposal : α = **Symbiosis**
View Point	Producer	User
Industry	large-scale production and consumption	Various, and small-scale production, recycle
Problems	･Environmental contamination ･Global Warming ･Natural Disaster ･Aging Society	・Symbiosis among human, society and nature achieved by **I T**
21st century	Post rationalism <20th century: Human conquers nature>	・Symbiotic Thought <Human assimilates into nature>


Problem- 1
･Acceleration of Demographic Aging
Global Warming
Aging Society
Problem- 2
･Environment Pollution
･Global Warming
Loss of model
Religion
Market Mechanism

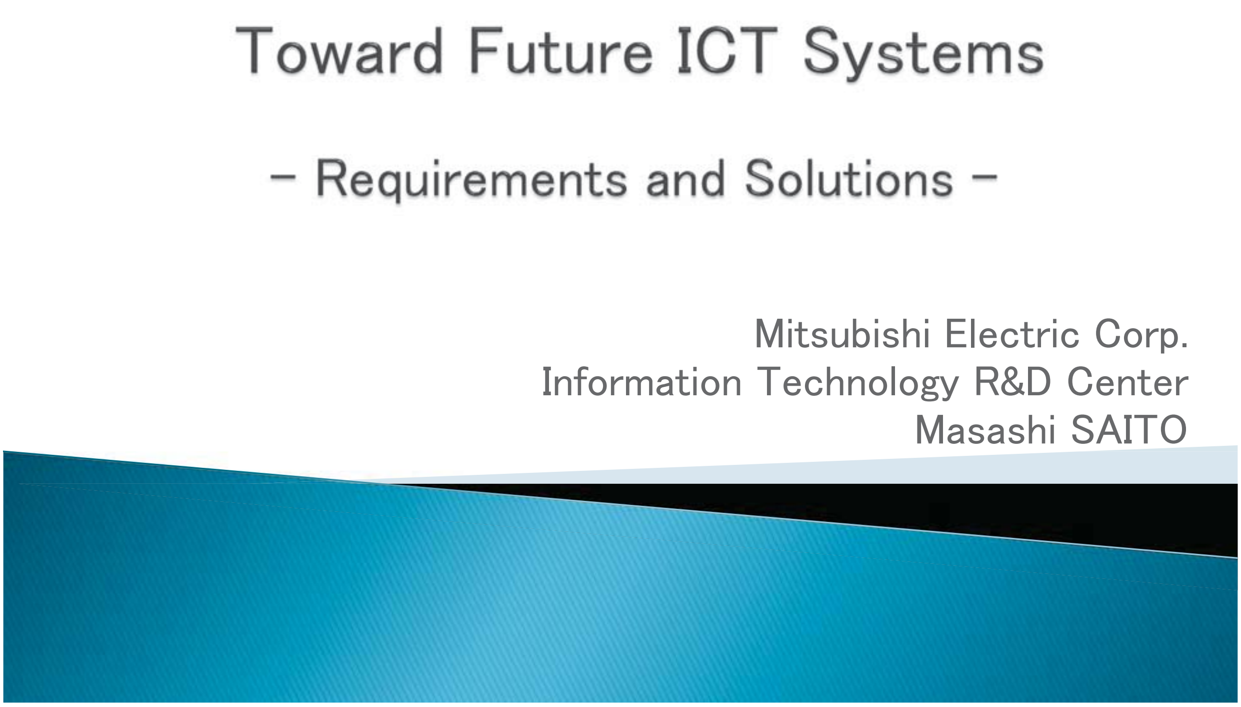

Solution- 1
?
?
Solution
Solution- 2
?
Global Warming
Aging Society
Loss of model
Religion
Market Mechanism


“ Symbiosis “ society
“Symbiosis”
“Green”
Information-Communication technology
– Harmony/Symbiosis –
Independent, watch-over, and health management support system
Green-oriented life-style environment
Conflict of country or region
Global Warming
Aging Society
Dwindling birthrate / aging population
Environment Disruption
Loss of model
Religion
Market Mechanism

Thank you for your attention !

Please visit my website:

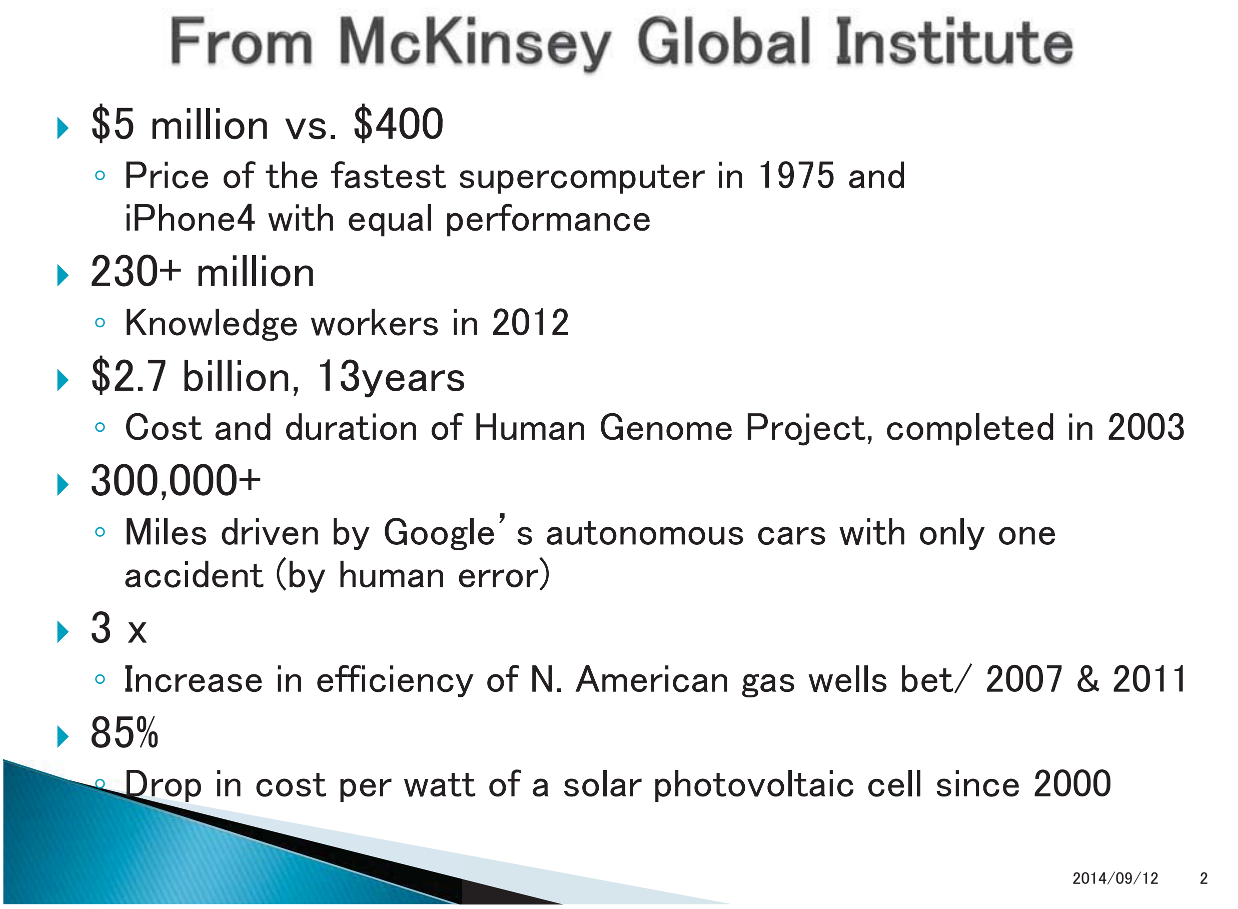
http://www.gits.waseda.ac.jp/research/faculty.php?id=3d4ffc&lang=en&ref=0



Toward Future ICT Systems

– Requirements and Solutions –

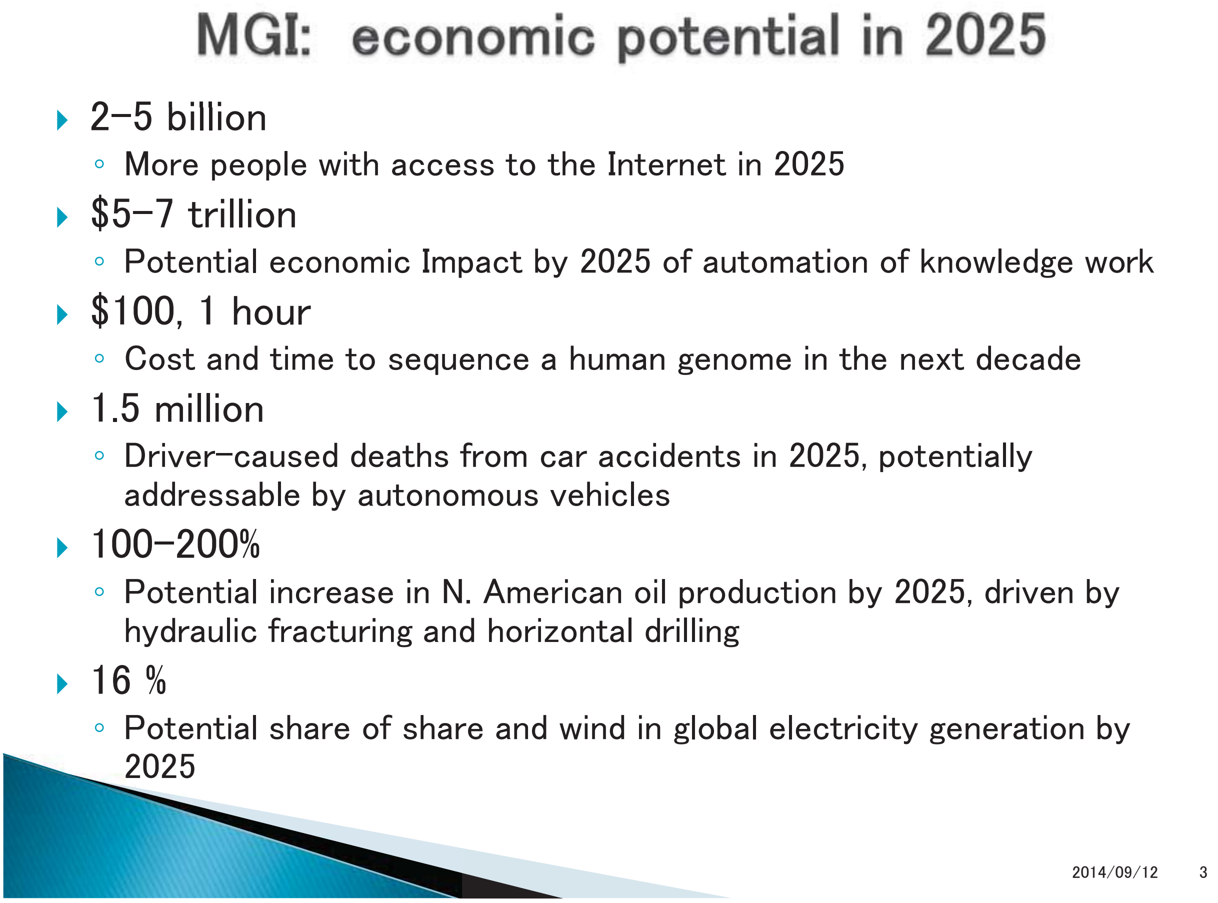
Mitsubishi Electric Corp.
Information Technology R&D Center
Masashi SAITO



From McKinsey Global Institute

- $5 million vs. $400
  - Price of the fastest supercomputer in 1975 and iPhone4 with equal performance
- 230+ million
  - Knowledge workers in 2012
- $2.7 billion, 13years
  - Cost and duration of Human Genome Project, completed in 2003
- 300,000+
  - Miles driven by Google's autonomous cars with only one accident (by human error)
- 3 x
  - Increase in efficiency of N. American gas wells bet/ 2007 & 2011
- 85%
  - Drop in cost per watt of a solar photovoltaic cell since 2000

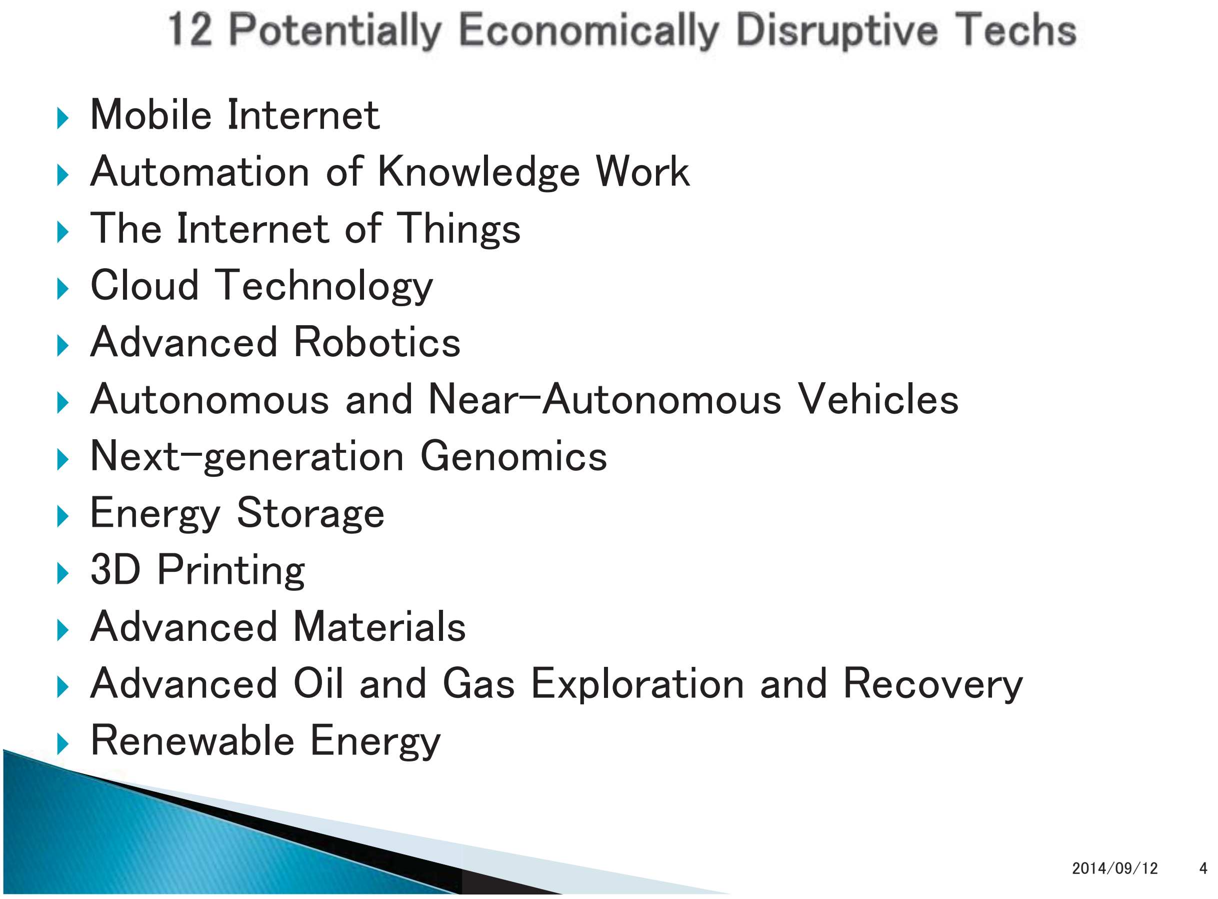
2014/09/12 2



MGI: economic potential in 2025

- 2–5 billion
  - More people with access to the Internet in 2025
- $5–7 trillion
  - Potential economic Impact by 2025 of automation of knowledge work
- $100, 1 hour
  - Cost and time to sequence a human genome in the next decade
- 1.5 million
  - Driver-caused deaths from car accidents in 2025, potentially addressable by autonomous vehicles
- 100–200%
  - Potential increase in N. American oil production by 2025, driven by hydraulic fracturing and horizontal drilling
- 16 %
  - Potential share of share and wind in global electricity generation by 2025

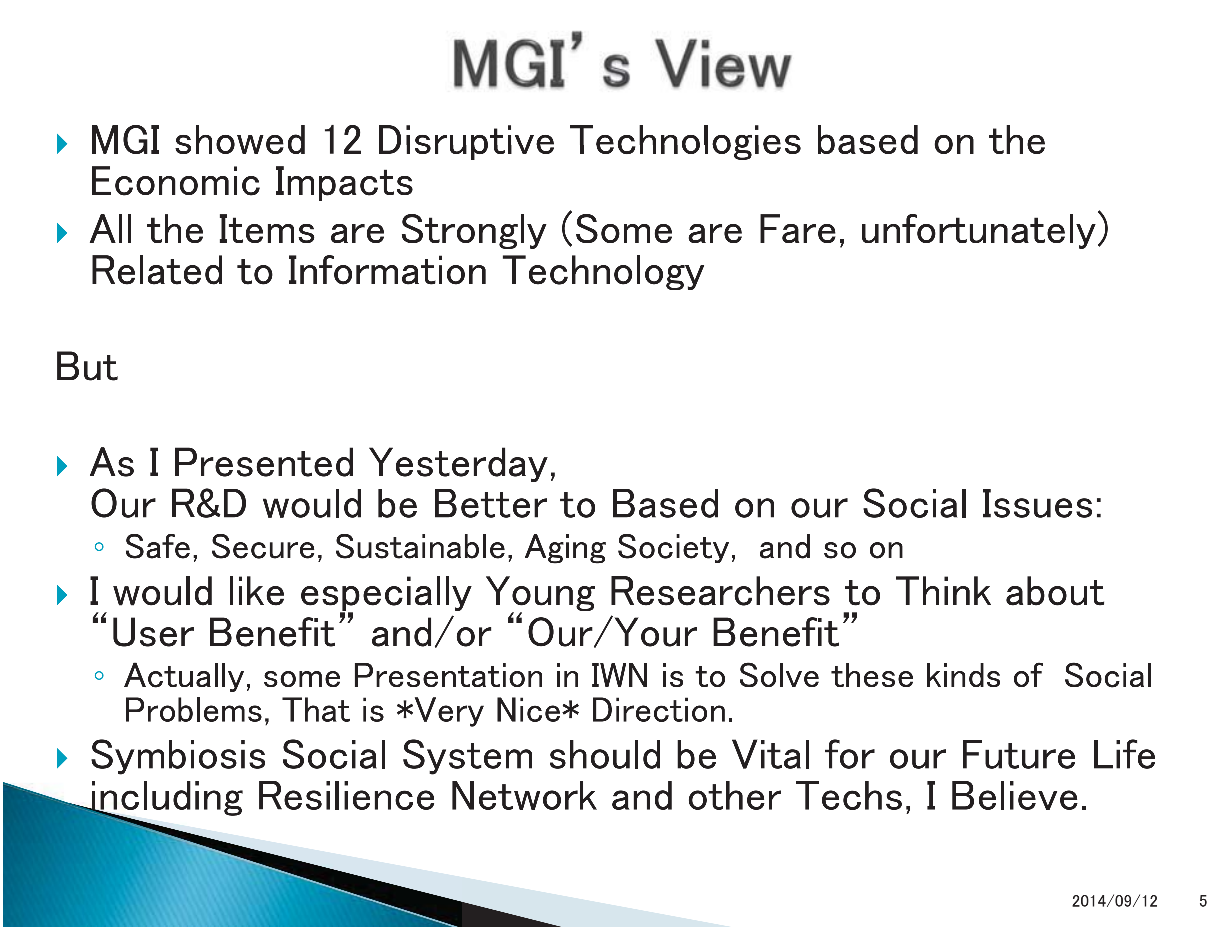
2014/09/12 3



12 Potentially Economically Disruptive Techs

- Mobile Internet
- Automation of Knowledge Work
- The Internet of Things
- Cloud Technology
- Advanced Robotics
- Autonomous and Near-Autonomous Vehicles
- Next-generation Genomics
- Energy Storage
- 3D Printing
- Advanced Materials
- Advanced Oil and Gas Exploration and Recovery
- Renewable Energy

2014/09/12 4



MGI's View

- MGI showed 12 Disruptive Technologies based on the Economic Impacts
- All the Items are Strongly (Some are Fare, unfortunately) Related to Information Technology

But

- As I Presented Yesterday,
  Our R&D would be Better to Based on our Social Issues:
  - Safe, Secure, Sustainable, Aging Society, and so on
- I would like especially Young Researchers to Think about "User Benefit" and/or "Our/Your Benefit"
  - Actually, some Presentation in IWN is to Solve these kinds of Social Problems, That is *Very Nice* Direction.
- Symbiosis Social System should be Vital for our Future Life including Resilience Network and other Techs, I Believe.

2014/09/12 5

Thank you very much.

2014/09/12 6

Towards Future ICT
- Requirements and Solutions -

IWIN 2014
Prague, Czech Republic, Sep. 12, 2014

Yoh Shiraishi

School of Systems Information Science

Future University Hakodate, Japan

Topics

- My brief introduction
  - Motivation
- Human based computing
  - Participatory sensing and crowd sourcing
- Awareness?

My introduction

- Future University Hakodate, Japan (from 2009)
  - Associate Professor
  - The Center for Spatial Information Science, Univ. Tokyo (-2009)
  - Ph.D (received from KEIO University, Prof. Yuichiro Anzai)
- Research backgrounds:
  - Database, sensor network (fixed sensors), geographic information system (GIS), ubiquitous computing
- Current research interests:
  - Participatory sensing (mobile sensing)
  - ITS, probe information system
  - Smart transportation system
  - (artificial intelligence, spatial cognition)

Motivation: ITS

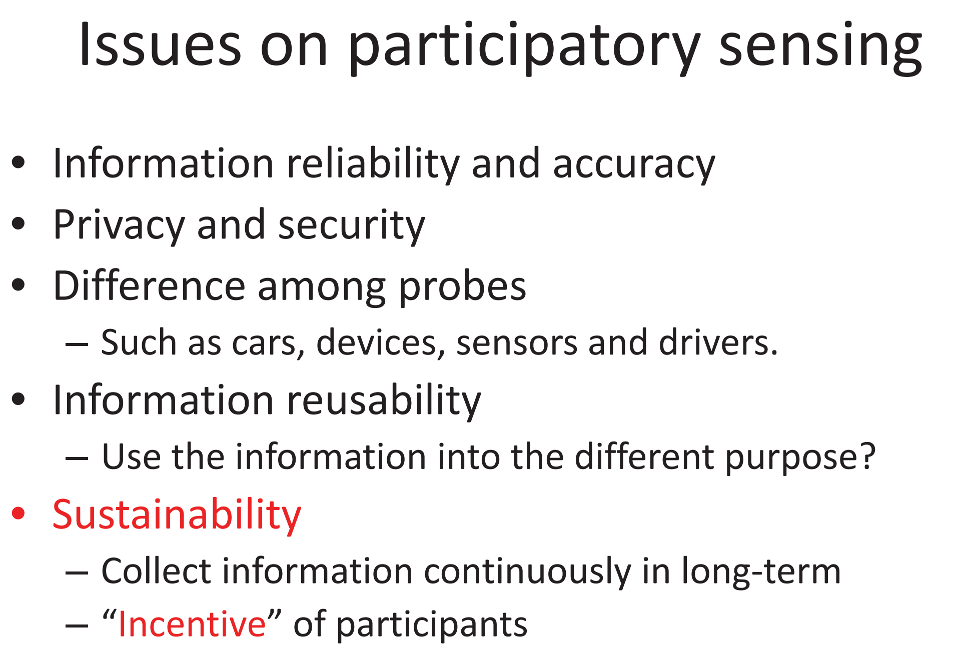
- Difficult to drive for beginners on roads
  - Irregular road infrastructure
  - In winter season
  - In spring season
    - Road cracks, bumpy roads

One solution for the problem

- “Implicit knowledge”
  - Students know the bottleneck points as the entrance of shopping centers and schools, intersections with no lane for right-turn
  - Human as a sensor
- Can we use our driving records?
  - As collective intelligence by collecting and sharing
- One solution:
  - Participatory sensing (probe information system, “human probe”)

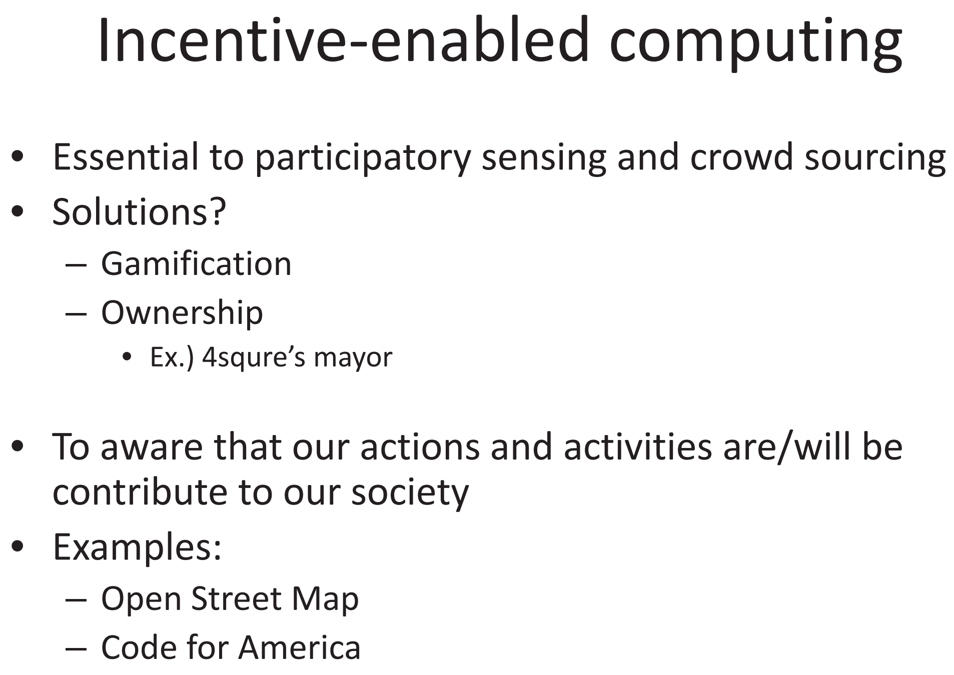
Participatory sensing

- Studies on ITS by our laboratory
  - To grasp the congestion situation by in-vehicle camera
  - To estimate the road conditions by a smartphone
- Collect various kinds of information with sensors and share these information.
- Such information are useful for the driver and other users as collective knowledge / additional values.
- Can solve many problems such as global warming, aging society and natural disaster.



Issues on participatory sensing

- Information reliability and accuracy
- Privacy and security
- Difference among probes
  - Such as cars, devices, sensors and drivers.
- Information reusability
  - Use the information into the different purpose?
- Sustainability
  - Collect information continuously in long-term
  - “Incentive” of participants



Incentive-enabled computing

- Essential to participatory sensing and crowd sourcing
- Solutions?
  - Gamification
  - Ownership
    - Ex.) 4squre’s mayor

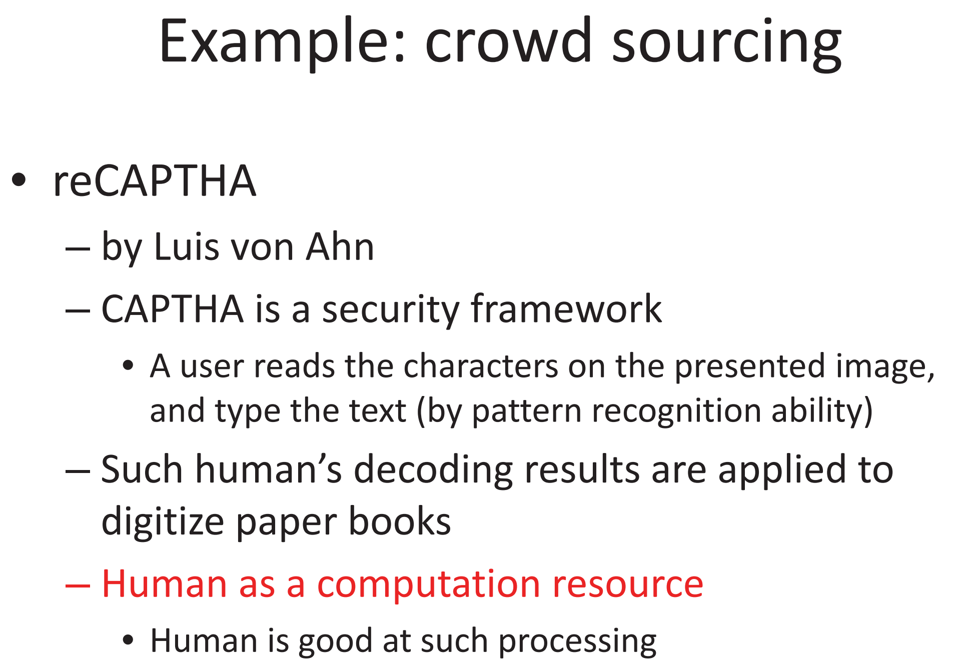
- To aware that our actions and activities are/will be contribute to our society
- Examples:
  - Open Street Map
  - Code for America

Example: crowd sourcing

- Open Street Map
  - In order to make and edit a map, users record and share their GPS trajectories
  - Potential to add useful information on each road
    - (such as traffic condition, road condition)
  - Capability to represent city dynamics by the mapping process

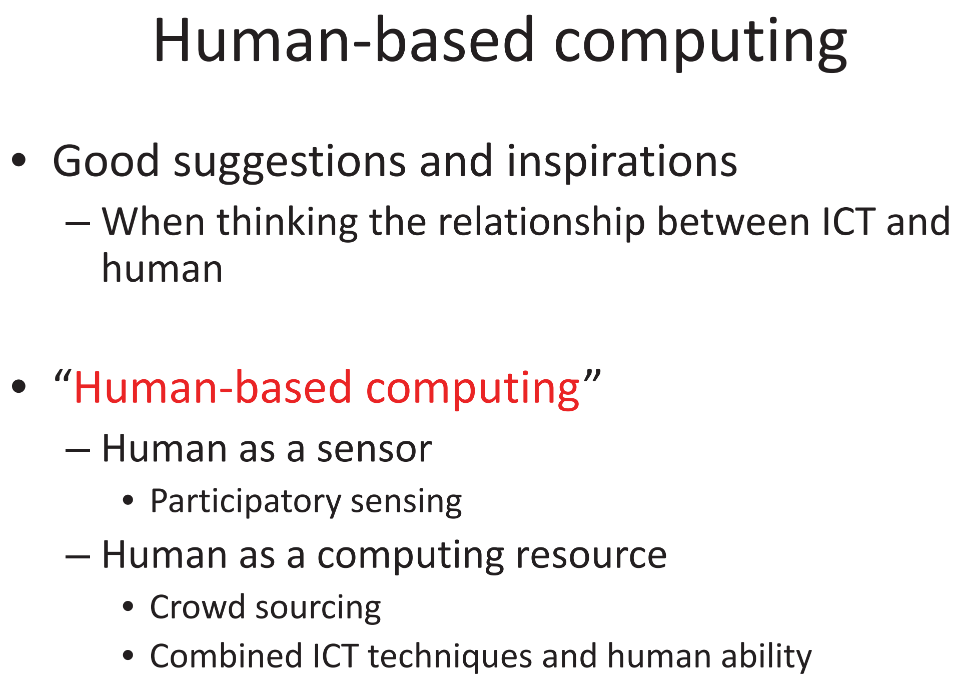
Example: crowd sourcing

- Code for America
  - All over the world
  - “Coding a better government”
    - By Jennifer Pahlka (TED conference)
  - Citizens discover the problems in the living environment (city, town)
  - Citizen hackers can solve these problems by programming
  - Open resources and programming by citizens



Example: crowd sourcing

- reCAPTHA
  - by Luis von Ahn
  - CAPTHA is a security framework
    - A user reads the characters on the presented image, and type the text (by pattern recognition ability)
  - Such human’s decoding results are applied to digitize paper books
  - Human as a computation resource
    - Human is good at such processing



Human-based computing

- Good suggestions and inspirations
  - When thinking the relationship between ICT and human

- “Human-based computing”
  - Human as a sensor
    - Participatory sensing
  - Human as a computing resource
    - Crowd sourcing
    - Combined ICT techniques and human ability

Awareness?

- “Context awareness”
  - The system is aware of user’s context and provide the service.
  - “Ubiquitous computing”
    - Calm computing, invisible computing (by Mark Weiser)
    - The existence of the devices is desirable to be hidden
  - May be an ideal case
  - But, the user may not perceive his/her context.
    - Uses the service with unconsciously and transparently

An example

- In the morning, the day before yesterday
  - The blackout happens in this hotel!
- Where were you?
  - At breakfast? Room? Elevator?
- Observations:
  - User A: put the button “open”
    - Cannot image our context
  - User B: put the button “0 floor”
    - Can image our context

How to improve the situation?

- Make a user to perceive his/her context
  - May be difficult to do such situation (disaster) due to the system down
- Make a user to image his/her context
  - Sometimes his/her context information may be provided by a system
  - Rough understanding of the system we use
    - Mechanism and behaviors of the system

User's ability to perceive and image the context

- What is "awareness"?
  - the ability to notice something using your senses
    - from Longman dictionary
- Some situations,
  - Human cannot perceive (or image) their context and situation, and cannot make decisions based on the context and situation
- A little sense to danger /crisis
  - Less opportunities to think
  - Too be convenient may degrade some human's abilities
    - Dare to be in-convenient?
- Benefits with inconvenience may be fine (sometimes)

Thank you very much for your attention

siraisi@fun.ac.jp

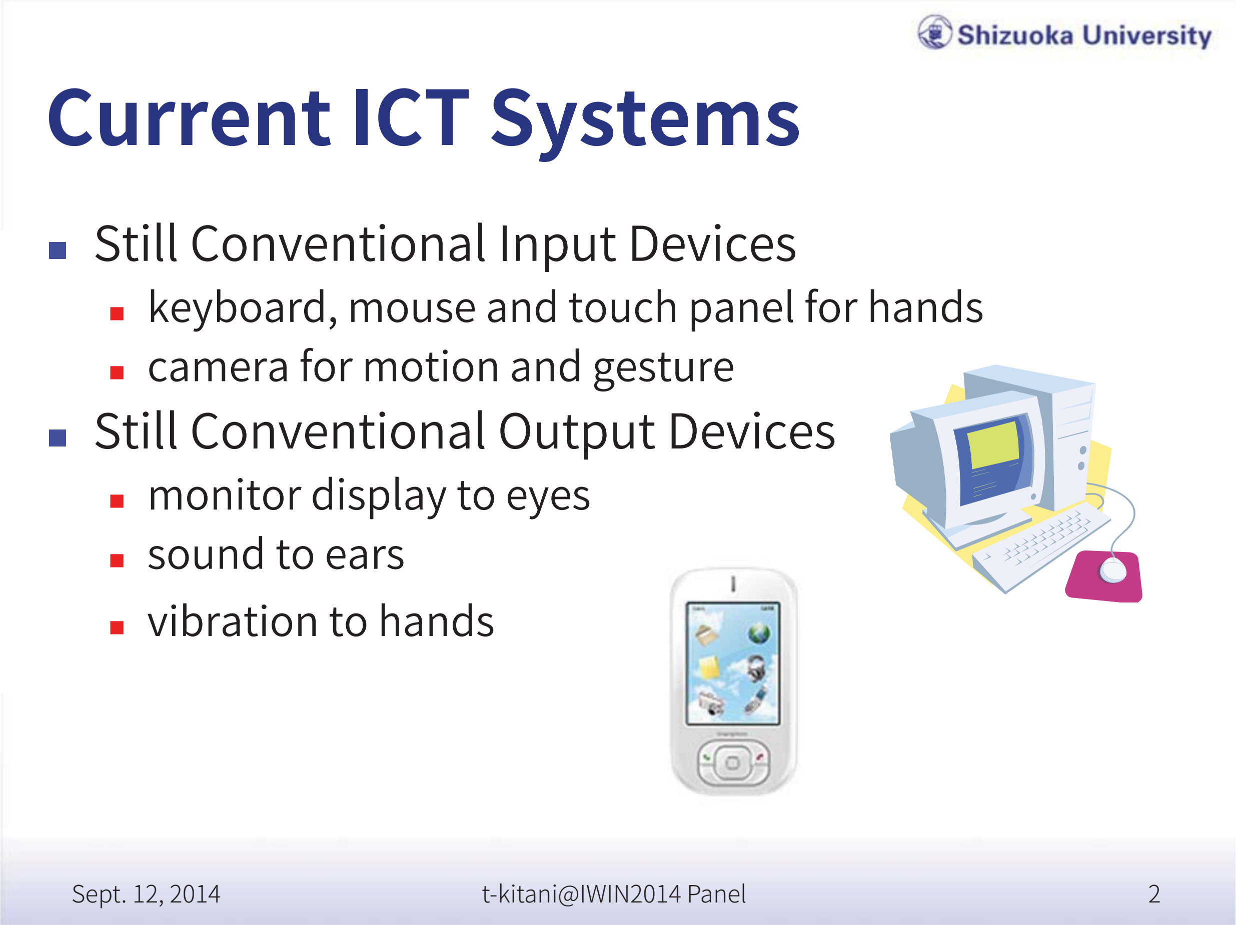
Shizuoka_univ. Kitani_lab.
BikeInformatics

Toward Future ICT Systems - Requirements and Solutions – Human-centric unconscious computing & Cyber (human)-Physical Systems

International Workshop on Informatics,
at Hotel Century Old Town Prague, Czech, on Sept. 12, 2014

Tomoya Kitani
Graduate School of Informatics, Shizuoka University
t-kitani@inf.shizuoka.ac.jp / t-kitani@kitanilab.org

Mikaeridai viewing platform, Rishiri island, Hokkaido, Aug. 21, 2014

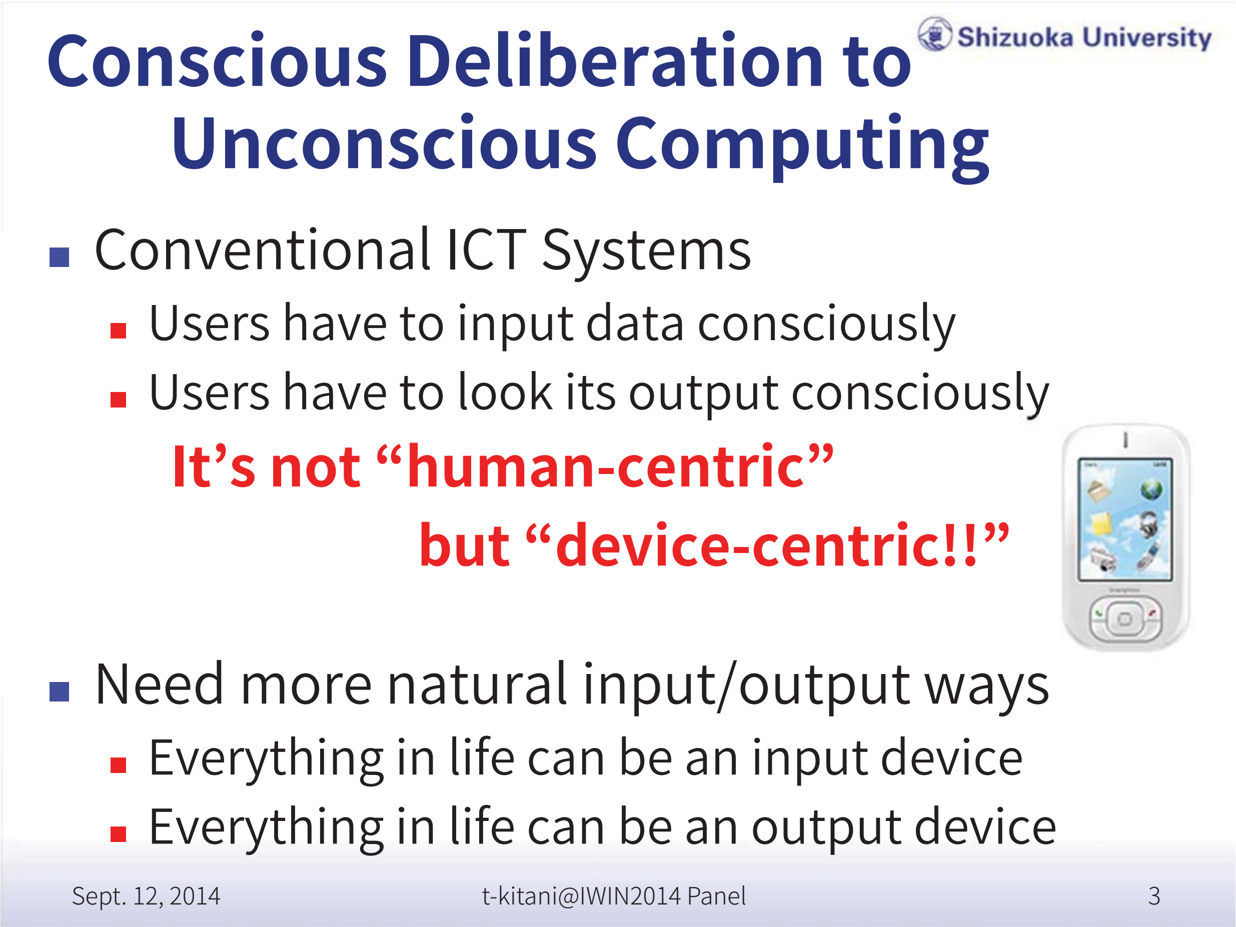


Shizuoka University

Current ICT Systems

- Still Conventional Input Devices
  - keyboard, mouse and touch panel for hands
  - camera for motion and gesture
- Still Conventional Output Devices
  - monitor display to eyes
  - sound to ears
  - vibration to hands

Sept. 12, 2014 t-kitani@IWIN2014 Panel 2



Shizuoka University

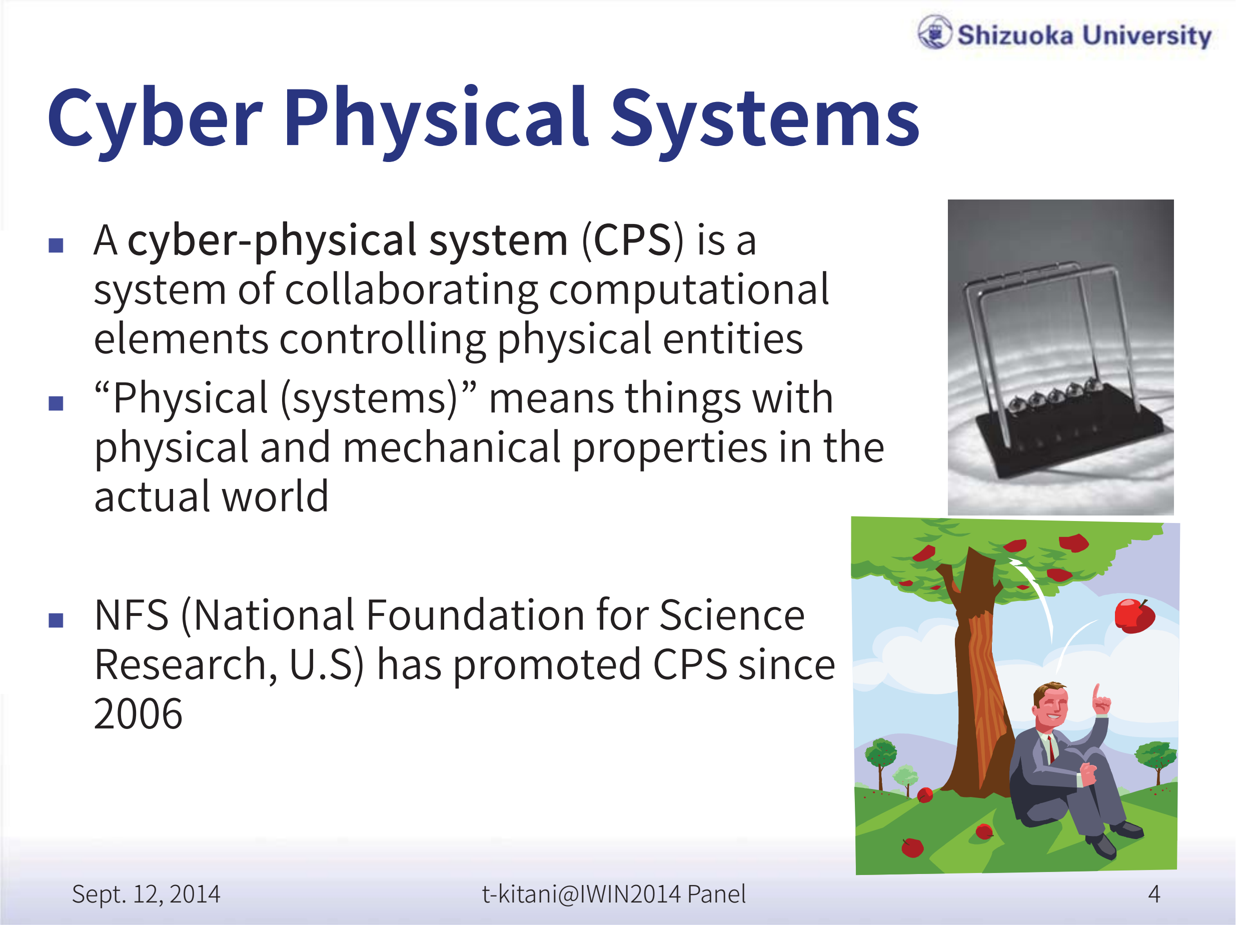
Conscious Deliberation to Unconscious Computing

- Conventional ICT Systems
  - Users have to input data consciously
  - Users have to look its output consciously

It's not "human-centric" but "device-centric!!"

- Need more natural input/output ways
  - Everything in life can be an input device
  - Everything in life can be an output device

Sept. 12, 2014 t-kitani@IWIN2014 Panel 3

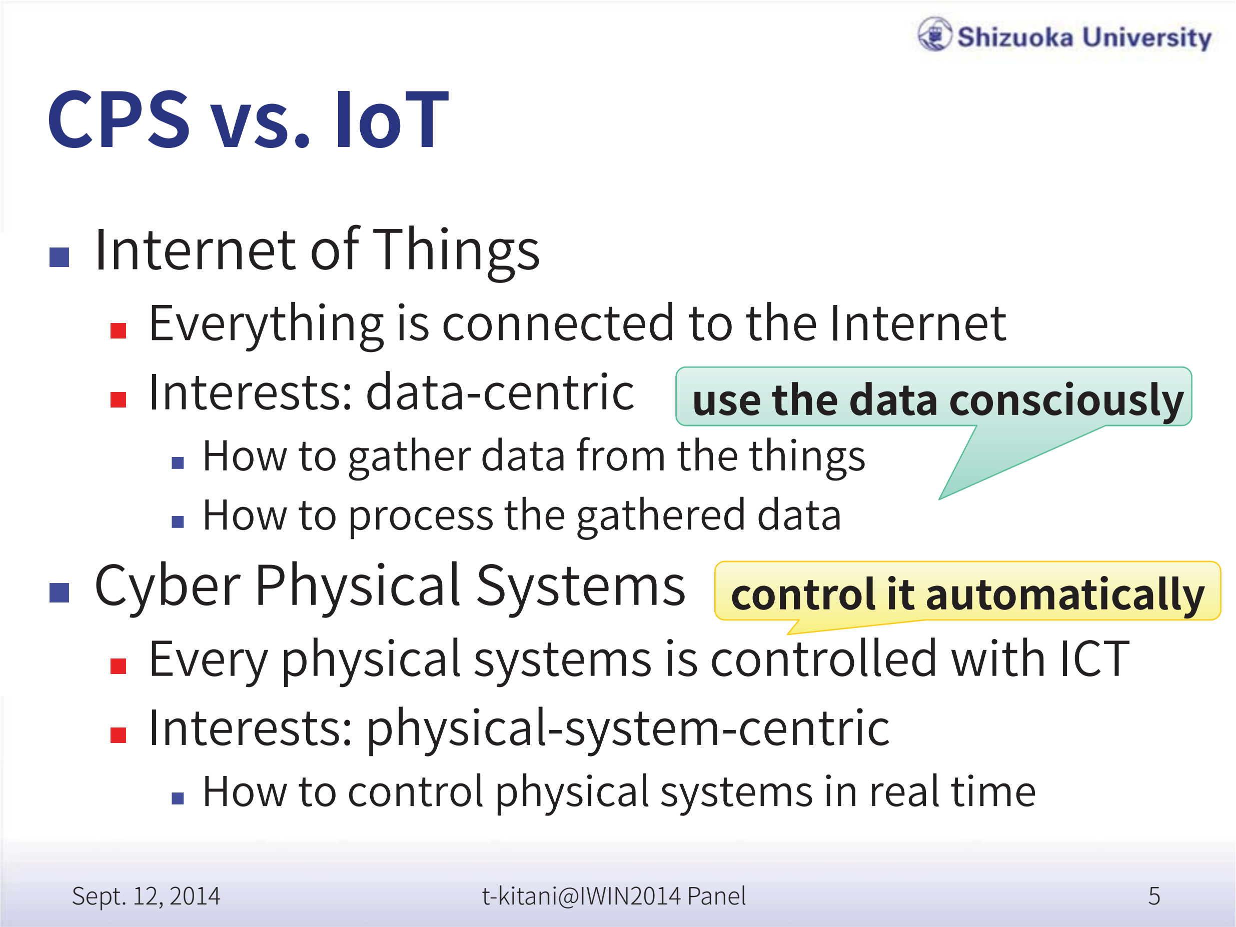


Shizuoka University

Cyber Physical Systems

- A **cyber-physical system** (CPS) is a system of collaborating computational elements controlling physical entities
- "Physical (systems)" means things with physical and mechanical properties in the actual world
- NFS (National Foundation for Science Research, U.S) has promoted CPS since 2006

Sept. 12, 2014 t-kitani@IWIN2014 Panel 4



Shizuoka University

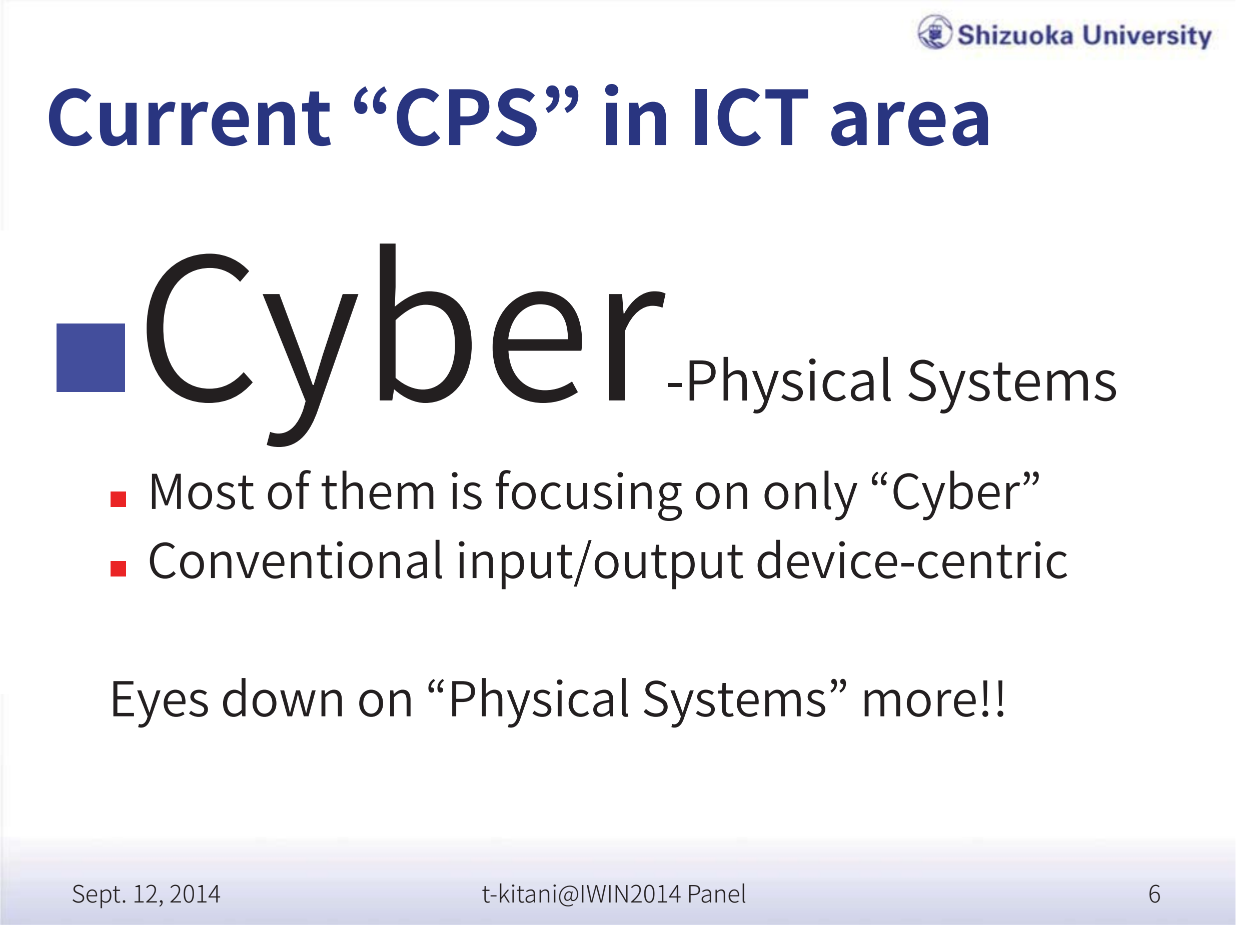
CPS vs. IoT

- Internet of Things
  - Everything is connected to the Internet
  - Interests: data-centric
    - How to gather data from the things
    - How to process the gathered data
- Cyber Physical Systems
  - Every physical systems is controlled with ICT
  - Interests: physical-system-centric
    - How to control physical systems in real time

use the data consciously

control it automatically

Sept. 12, 2014 t-kitani@IWIN2014 Panel 5



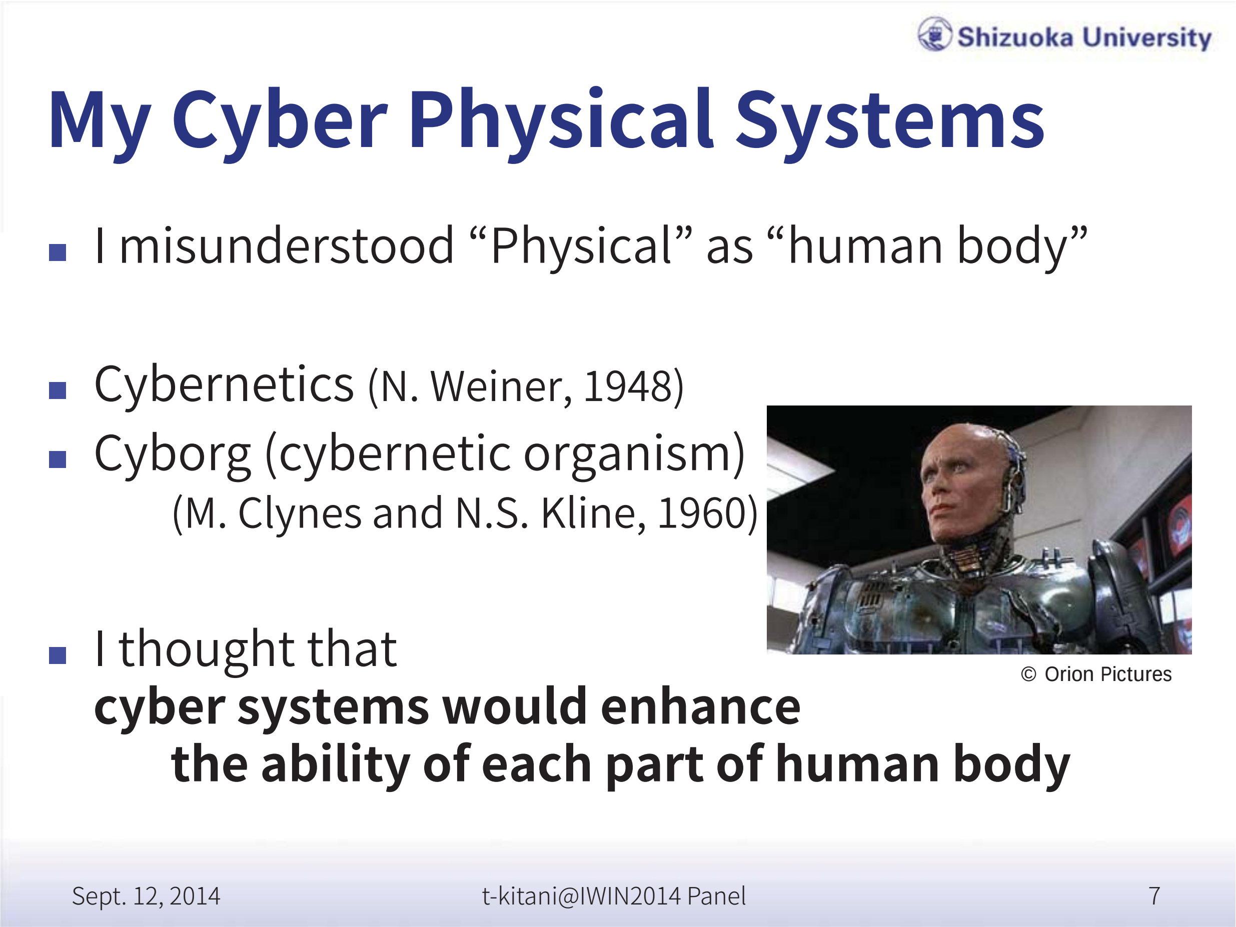
Shizuoka University

Current “CPS” in ICT area

- Cyber-Physical Systems
  - Most of them is focusing on only “Cyber”
  - Conventional input/output device-centric

Eyes down on “Physical Systems” more!!

Sept. 12, 2014 t-kitani@IWIN2014 Panel 6



Shizuoka University

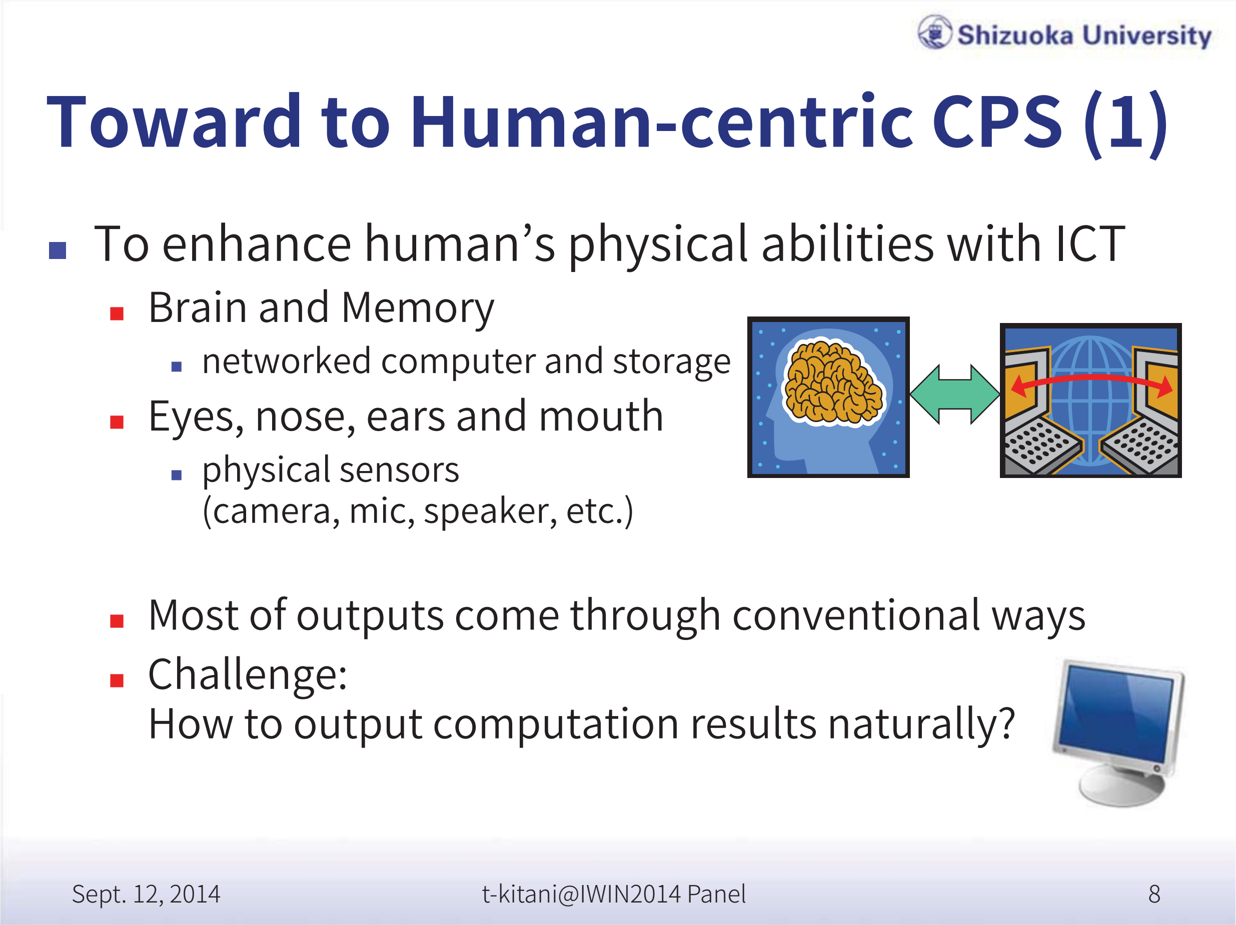
My Cyber Physical Systems

- I misunderstood “Physical” as “human body”

- Cybernetics (N. Weiner, 1948)
- Cyborg (cybernetic organism)
  (M. Clynes and N.S. Kline, 1960)

© Orion Pictures

- I thought that
  cyber systems would enhance
  the ability of each part of human body

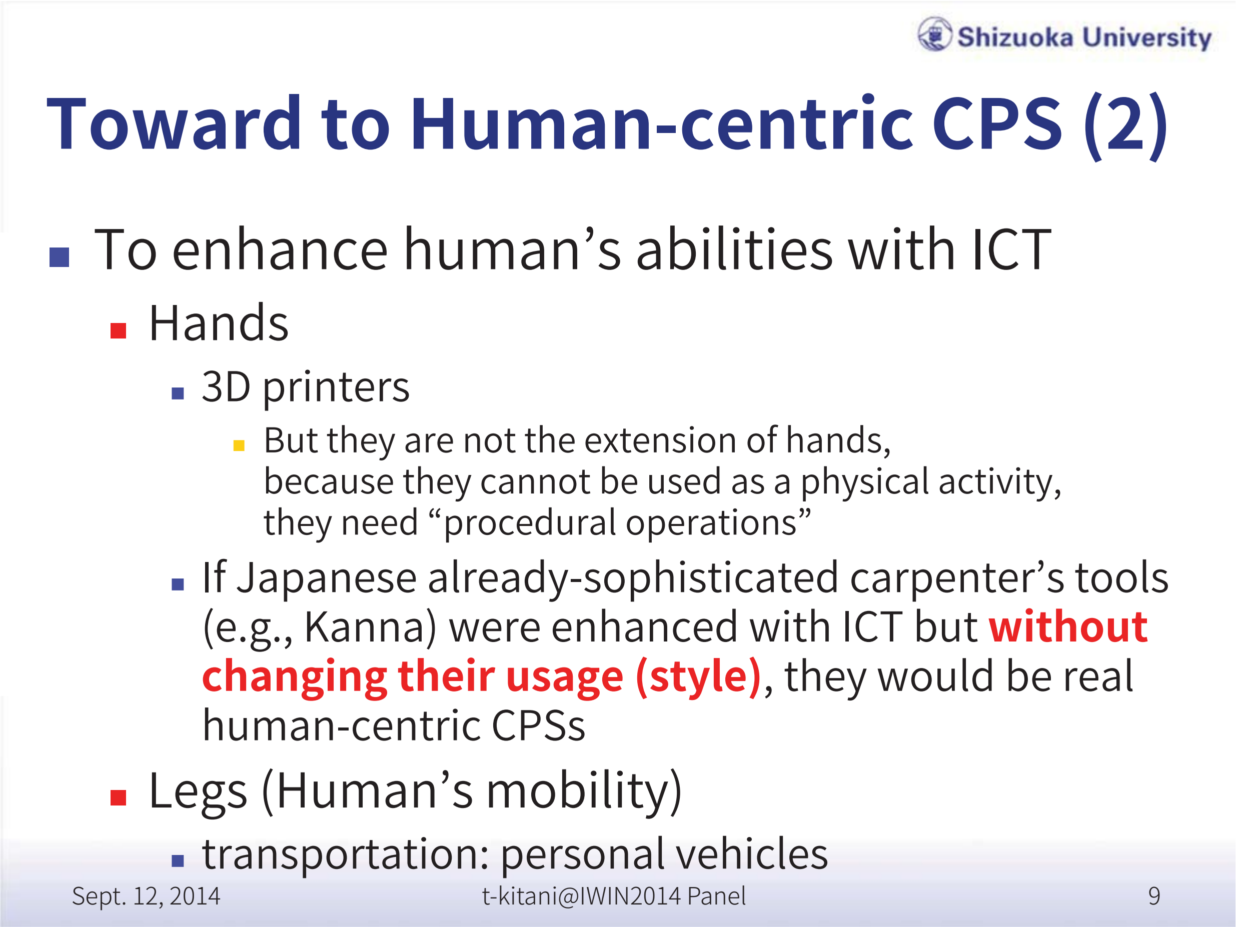


Shizuoka University

Toward to Human-centric CPS (1)

- To enhance human’s physical abilities with ICT
  - Brain and Memory
    - networked computer and storage
  - Eyes, nose, ears and mouth
    - physical sensors
      (camera, mic, speaker, etc.)

  - Most of outputs come through conventional ways
  - Challenge:
    How to output computation results naturally?

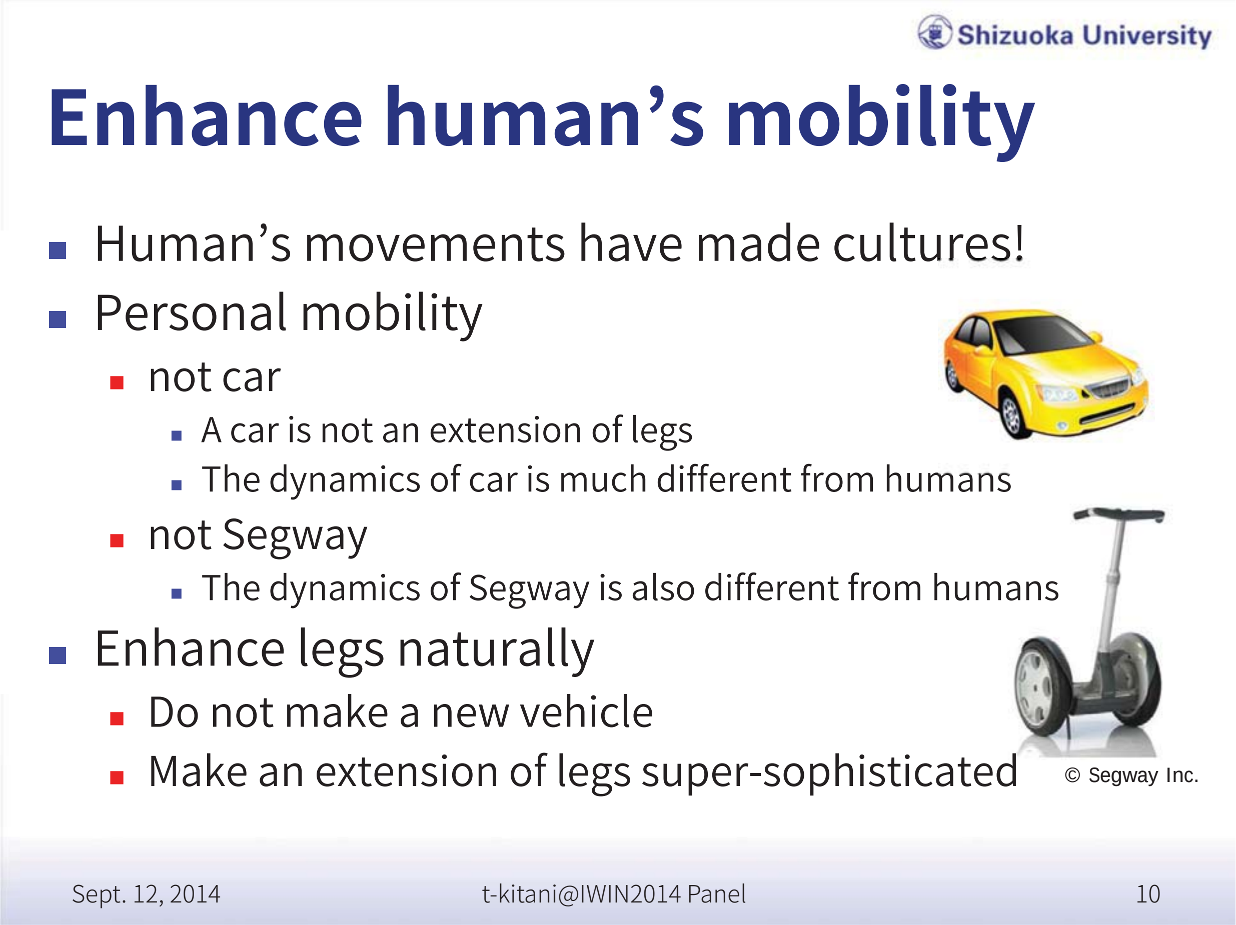


Shizuoka University

Toward to Human-centric CPS (2)

- To enhance human's abilities with ICT
  - Hands
    - 3D printers
      - But they are not the extension of hands,
        because they cannot be used as a physical activity,
        they need "procedural operations"
    - If Japanese already-sophisticated carpenter's tools (e.g., Kanna) were enhanced with ICT but **without changing their usage (style)**, they would be real human-centric CPSs
  - Legs (Human's mobility)
    - transportation: personal vehicles

Sept. 12, 2014 t-kitani@IWIN2014 Panel 9



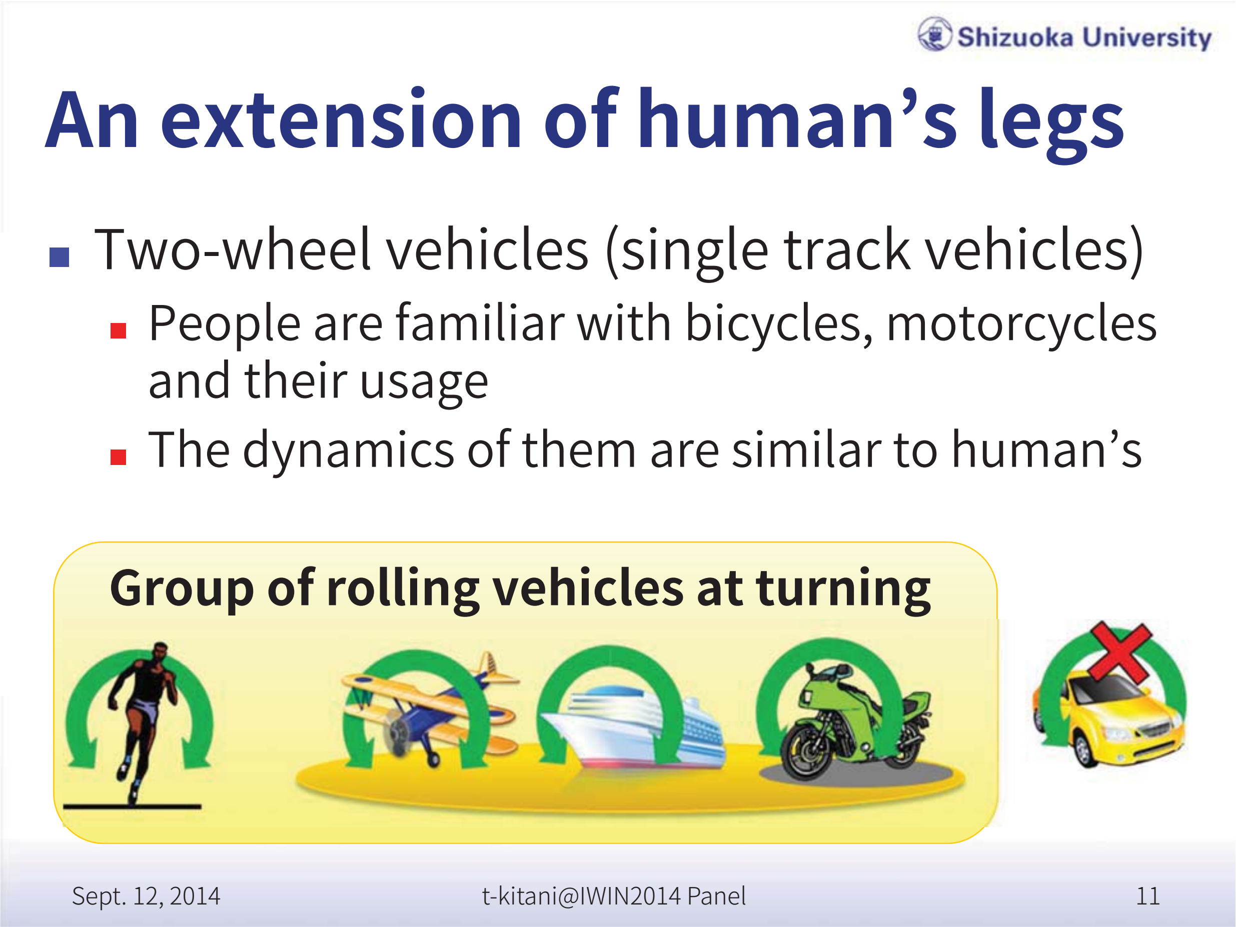
Shizuoka University

Enhance human's mobility

- Human's movements have made cultures!
- Personal mobility
  - not car
    - A car is not an extension of legs
    - The dynamics of car is much different from humans
  - not Segway
    - The dynamics of Segway is also different from humans
- Enhance legs naturally
  - Do not make a new vehicle
  - Make an extension of legs super-sophisticated

© Segway Inc.

Sept. 12, 2014 t-kitani@IWIN2014 Panel 10



Shizuoka University

An extension of human’s legs

- Two-wheel vehicles (single track vehicles)
  - People are familiar with bicycles, motorcycles and their usage
  - The dynamics of them are similar to human’s

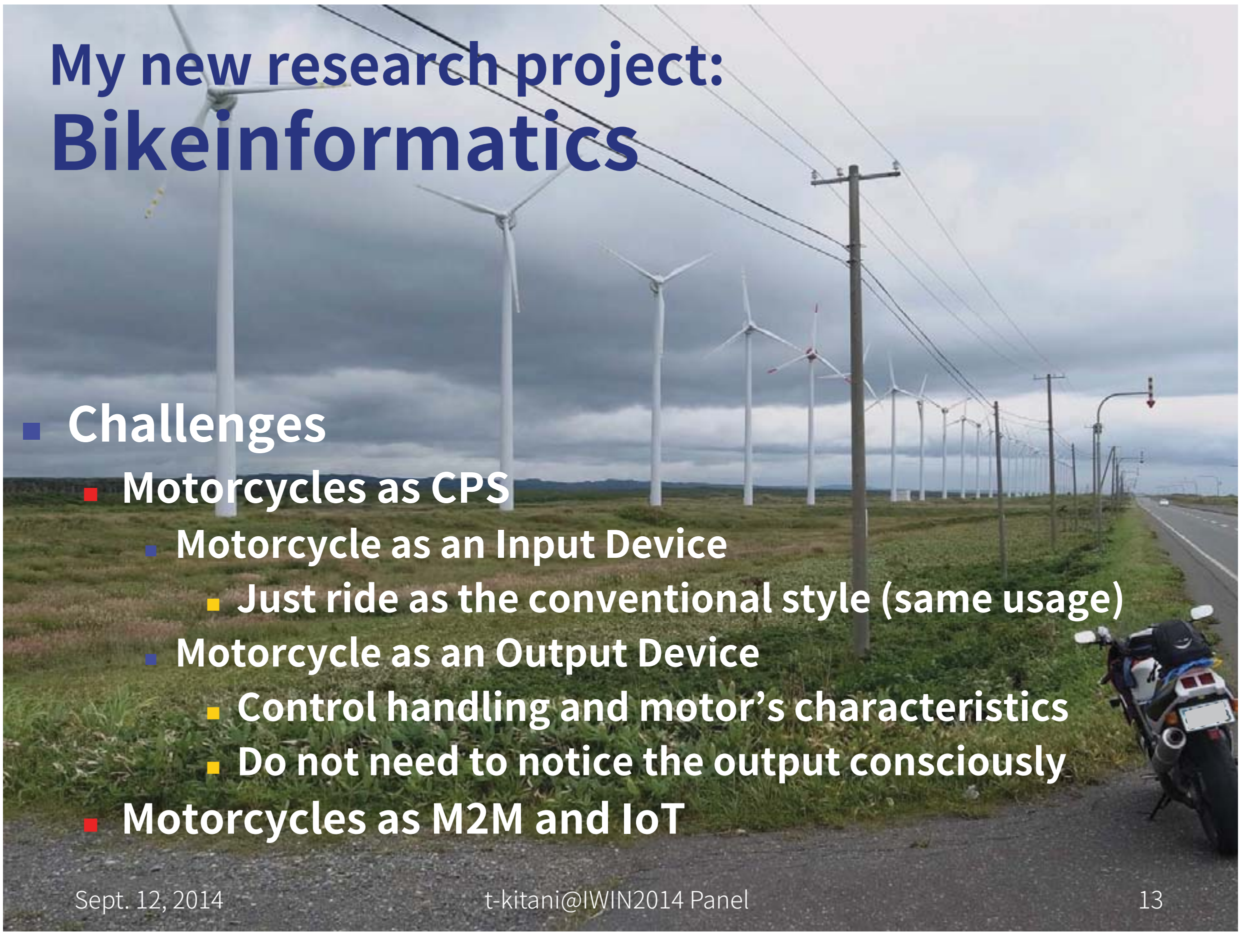
Group of rolling vehicles at turning

Sept. 12, 2014 t-kitani@IWIN2014 Panel 11

My new research project: Bikeinformatics

- To make motorcycles super-sophisticated
  - Current hardware are already sophisticated
  - Need good software to control motorcycles well

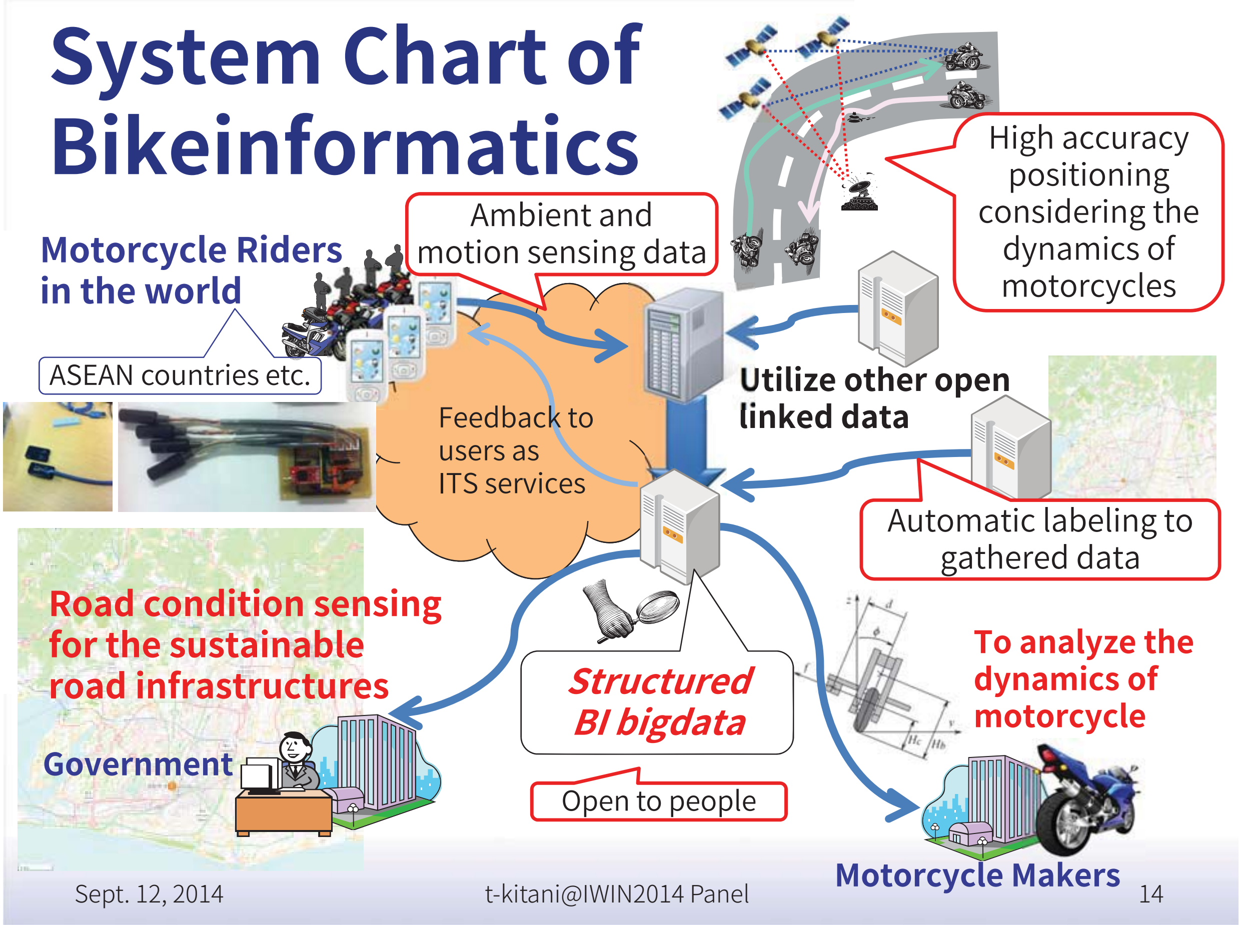
Sept. 12, 2014 t-kitani@IWIN2014 Panel 12



My new research project: Bikeinformatics

- Challenges
  - Motorcycles as CPS
    - Motorcycle as an Input Device
      - Just ride as the conventional style (same usage)
    - Motorcycle as an Output Device
      - Control handling and motor's characteristics
      - Do not need to notice the output consciously
  - Motorcycles as M2M and IoT

Sept. 12, 2014 t-kitani@IWIN2014 Panel 13



System Chart of Bikeinformatics

Motorcycle Riders in the world

ASEAN countries etc.

Ambient and motion sensing data

High accuracy positioning considering the dynamics of motorcycles

Utilize other open linked data

Feedback to users as ITS services

Automatic labeling to gathered data

Road condition sensing for the sustainable road infrastructures

Government

Structured BI bigdata

Open to people

To analyze the dynamics of motorcycle

Motorcycle Makers

Sept. 12, 2014 t-kitani@IWIN2014 Panel 14