International Journal of Informatics Society

06/15 Vol. 7 No.1 ISSN 1883-4566

Informatics Society

Aims and Scope

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its quality and value as a resource. Informatics also referred to as Information science, studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

Guest Editor's Message

Tomoo Inoue

Guest Editor of Nineteenth Issue of International Journal of Informatics Society

We are delighted to have the nineteenth issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Eighth International Workshop on Informatics (IWIN2014), which was held at Prague, Czech Republic, Sep. 10-12, 2014. The workshop was the eighth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 24 papers were presented in five technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of networking, business systems, education systems, design methodology, groupware and social systems.

Each paper submitted IWIN2014 was reviewed in terms of technical content, scientific rigor, novelty, originality and quality of presentation by at least two reviewers. Through those reviews 15 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. This volume includes three papers among the accepted papers, which have been improved through the workshop discussion and the reviewers' comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners over the world.

Tomoo Inoue is Professor of the Faculty of Library, Information and Media Science of University of Tsukuba. His research interests include HCI, CSCW, and Technology-enhanced learning. He received his Ph.D. in Engineering from Keio University in 1998. He is a recipient of awards including Best Paper Award, Activity Contribution Award and SIG Research Award from Information Processing Society of Japan (IPSJ). He has served a number of academic committees, including IPSJ Journal Group Editor-in-Chief, IPSJ Transactions on Digital Content Managing Editor-in-Chief, IEICE SIG Human Communication Science Steering Board, IPSJ SIG Groupware and Network Services Steering Board, ACM CSCW Papers Associate Chair, and IEEE TC CSCWD.

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 cannot 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 cannot 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. Furthermore, 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 efficient 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 a great deal of data (here in after, "batch update"), that is the batch processing [5]. It is used widely in various fields: the settlement of account in the accounting systems, a great deal of account transfer for entrust company in the banking systems, and so on. Here, since its target data are 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: 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 are entered from the ATMs (Automatic Teller Machines). We call this entry "online entry." As for the online entry, small amount of data are entered at each time, and they are reflected into the database immediately. And, their concurrent execution is controlled by the transaction processing [5], [13] 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 is maintained. Moreover, the concurrency control is performed by the DBMS using the lock method 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 method is used by the batch update for a great deals of data, then the conflicting online entries are made to wait for a long while. That is, though the batch update can be 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. It splits the batch update to the short time transactions, and executes them sequentially [5]. However, the atomicity and isolation of the ACID properties cannot be maintained as for the whole mini-batch. That is, since the updated data are committed one after another, the rollback of the whole target data cannot be executed even in the case of the failure. That is, the update must be cancelled 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 are not 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 cannot 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 [9]. However, we have not evaluated this method in the case of failure yet.

So, our goal in this paper is to evaluate the recovery function of the temporal update for the transaction failure assum-

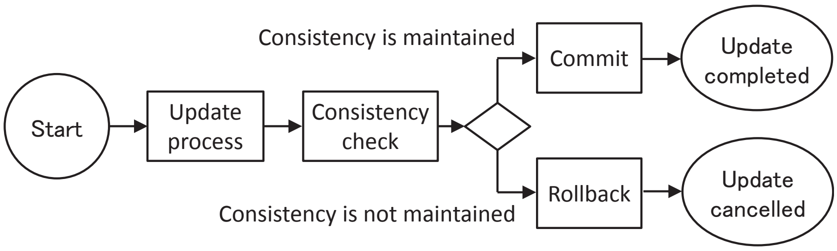


Figure 1: State transition of update transaction.

ing the actual business 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 entries in these system operations. Next, based on these requirements, we perform the experiment about this recovery method 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 a transaction, 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 Fig. 1. For example, in the banking systems, if the updated result of the account transfer becomes minus, it cannot be executed. So, the rollback of the transaction is executed, and the process is cancelled. 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 Fig. 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 is not maintained, the target table is recovered by the backup data. That is, the update is cancelled. 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.

Here, there are other recovery methods, for example, the differential backup function of the DBMS, compensating transaction. However, the recovery by the former cancels the update of the other transactions executed simultaneously, and the latter needs the extra program. So, the recovery of the batch update is generally executed by above-mentioned process.

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 completion; the state of the update cancelled 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. Then, the other transactions can access the table with the consistency. As for the isolation, the target table of the update must not 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 lock method for the target table or data during the batch update. Though the ACID properties of the batch update are able to be maintained by this method, there is a problem that the online entries to access the target data must wait 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 [13],[14]. 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 cannot be maintained 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], [10]. So, there is the same problem.

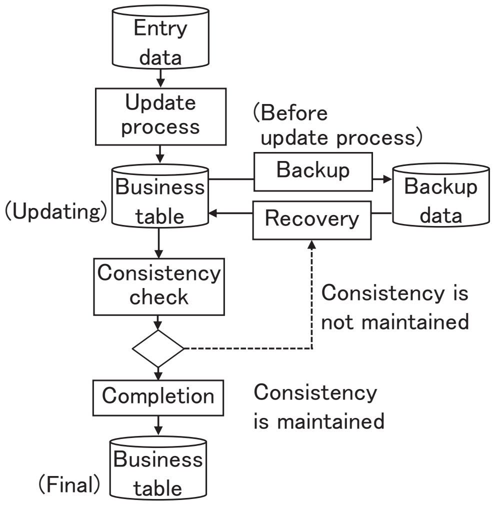


Figure 2: Data flow of batch update process.

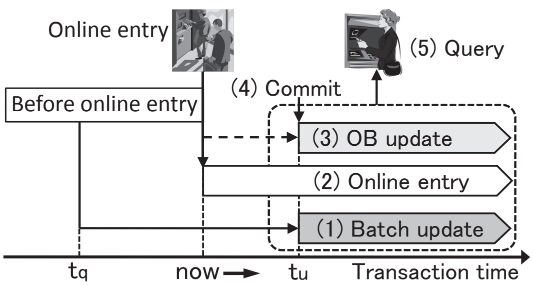


Figure 3: Outline of temporal update method.

2.2 Temporal Update Method

To solve this problem, we have proposed the temporal update method [9]. In this method, we use the extended transaction time. Here, the conventional transaction time expresses when a fact existed in the database [11], [4]. 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 has not been deleted yet, the instance of T_d is expressed by now, which is the current time to query the database [12]. And, we extend this time to the future.

In Fig. 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 does not 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. Thus, the online entry result is updated similarly to the batch update separately as shown by (3). We call this update "online batch update" (hereinafter, "OB update"). 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 exist after the commit, valid data have to be queried as shown by (5). The target data are selected as follows: firstly, the latest update data are selected for each primary key, and they can be identified by the above-mentioned time attribute T_a; secondly, if there are still plural data, the target data are 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.

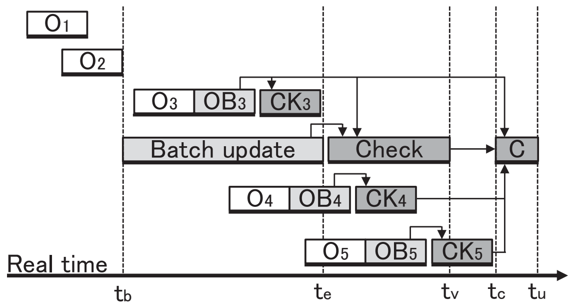


Figure 4: Execution timing of transactions.

We have shown the following in our previous studies of the temporal update method [6],[7]. Firstly, we can execute the batch update about the data related to each other concurrently with the online entry, though the both are executed separately by using the conventional methods. Secondly, we can execute the batch update more efficiently than the mini-batch, which is the conventional method to update the data unrelated each other in a lump, in both of the centralized and distributed database environment. Moreover, we have shown that we can apply this method even if the completion time cannot be predicted at the batch update starting [8]. On the other hand, it is necessary to consider the following to apply this method: the implementations of the OB update program, the transaction time and so on; in the case where the online entry frequency is high, its efficiency declines.

However, these previous studies assume only the case where the temporal update completed successfully. And, the following studies as for the transaction failure have not performed yet: 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 Fig. 1.

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 Fig. 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 due to the resource restrictions. On the other hand, at the batch update completion time t_e, from the viewpoint of the transaction processing of DBMS, the whole batch update and the completed OB update results must not 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 Fig. 2. And, if their consistencies are not 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 is not maintained, both of its rollback and the corresponding online entry O_i's rollback are done.

Thus, 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 Fig. 4. Here, the consistency has to be checked 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 must not 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 must not 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 is not executed if the latter is failed. Conversely, when a failure is detected in the OB update or batch update, it is necessary to execute their rollbacks in each case. 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 cancelled by the rollback.

For 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

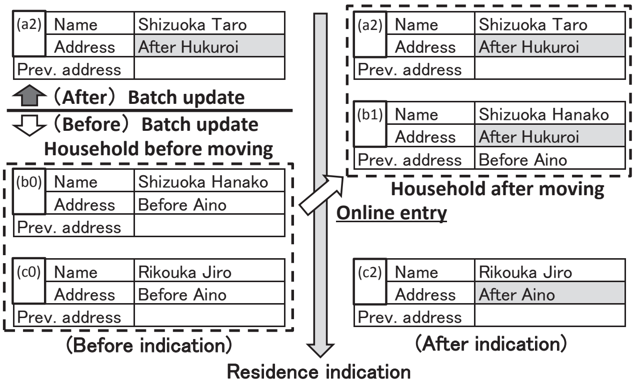
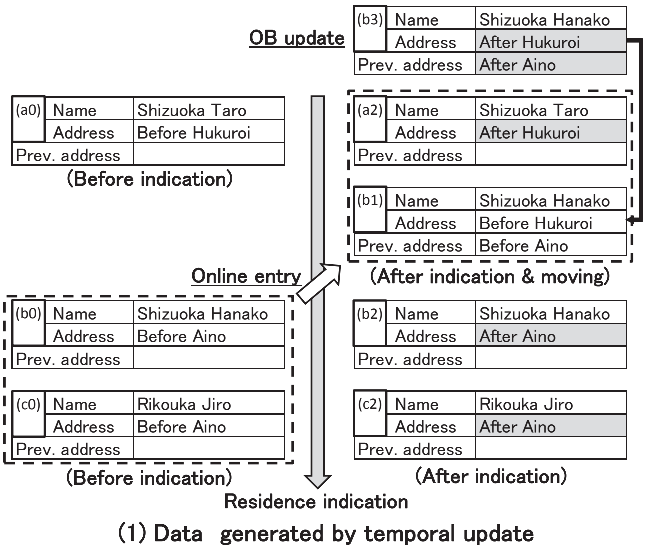


Figure 5: Update example of mini-batch.



Name	Normal end of temporal update	Failure of OB update / online entry	Failure of batch update
Shizuoka Taro	(a2)	(a2)	(a0)
Shizuoka Hanako	(b3)	(b2)	(b1)
Rikoouka Jiro	(c2)	(c2)	(c0)

(2) Update result of normal end and failure

Figure 6: Update example of temporal update.

must have the same current address, and a resident after moving has the previous address, which must be consistent with the current 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 cannot be executed by the mini-batch concurrently with the online entries. We show this example in Fig. 5. Here, "Address" shows the current address, and "Prev. Address" shows the previous address in it. In this case, household addresses are changed by the mini-batch update. Concurrently, by the online entry, a member of a household (b0) that has not 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 does not reflect the residence indication. However, the current address reflects it, though it is same timing with the previous address. Therefore, an anomaly between the current address and previous address occurs.

On the other hand, we show the case of executing the same processing by the temporal update method in (1) of Fig. 6. As for this method, its batch update results are not queried until the commit. So, as shown by (b1) in this figure, both of the current address and the previous address of the online entry result are based on the data before the residence indication. Similarly, the following data are generated: the batch update result (b2), which does not 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 as shown in (2) of Fig. 6. First, if the OB update fails, then the current address shown by (b1) remains. So, it causes anomaly among (b1) and (a2), that is, the residents in the same household have different current addresses. Therefore, both of the OB update and the online entry must be cancelled 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, its result is same as this. Third, if the batch update fails by some reason, the batch update and OB update are cancelled. They are (a2), (b2), (c2) and (b3). 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 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 Fig. 7, we show the program composition of the temporal update to realize the data manipulation shown in Fig. 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.

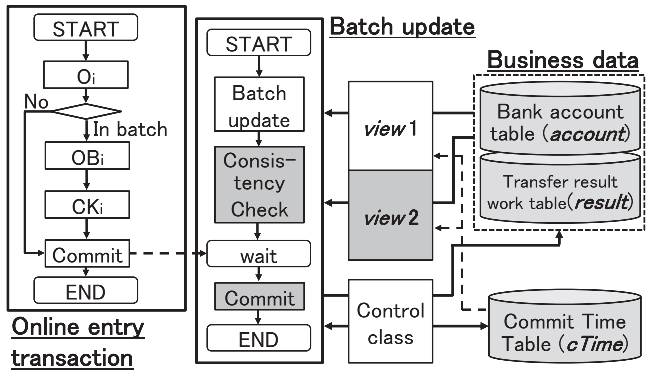


Figure 7: Program composition of temporal update.

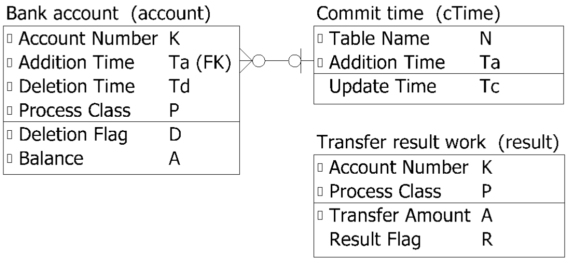
As for the table of database, we use the following three tables. Bank account table (hereinafter, "$account$") stores the deposit balance of the account, and Commit time table (hereinafter, "$cTime$") stores the data to manage the temporal update. In addition, Transfer result work table (hereinafter,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

Here, the data of $account$ are queried via following two views [3] as shown in Fig. 7, and $cTime$ is used for these views. First, "$view1$" is used by the online entry transactions, and the data of the batch update and OB update are not queried until the commit of the batch update. Of course, the transaction can query the OB update data made by itself. Second, "$view2$" 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 even before they are committed.

In Fig. 8, we show the table composition and its query result via views. Firstly, as for $account$, if we express its relation scheme by R_a, then its attributes are expressed by the following. 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.

$$R_a(K, T_a, T_d, P, D, A) \quad (1)$$

- 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 Fig. 8. It is for the case where the completion time of the batch update cannot 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 does not be queried. It is used by the OB update to hide



(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*1

Bank account (*account*) and Views

K	*Ta*	*Td*	*P*	*D*	*A*	*view*1	*view*2
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		

Commit Time (*cTime*)

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

(c) Query result of account via views

Figure 8: Data and views of temporal update.

the corresponding batch update result, which data was deleted by the online entry.

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

Next, as for $cTime$, if we express its relation scheme by R_c, then its attributes are expressed by the following.

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

- 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 views. 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 are queried as follows. We show its SQL expression in (b) of Fig. 8. First, as for the data $P = \mathrm{P_b}$ (batch update result) or $P = \mathrm{P_{ob}}$ (OB update result) in $account$, if $account.T_a = cTime.T_a$ then the value of $account.T_a$ is replaces by the one of corresponding $cTime.T_c$. Next, as for each K, the data having the largest P from among the data having the latest T_a are queried. Incidentally, "now" is the current time mentioned in Section 2.2.

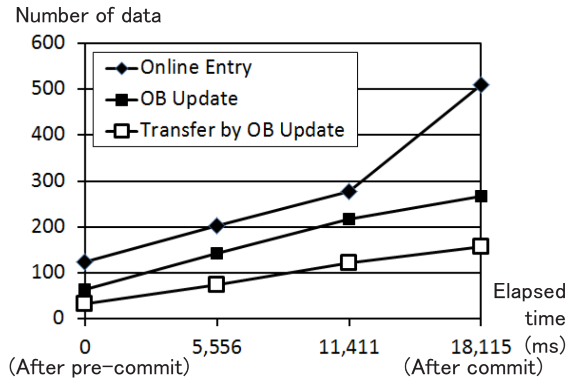


Figure 9: Transition of number of data via $view2$.

Similarly, we define $view2$. As for it, the data are 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 views in (c) of Fig. 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 is not 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 environment: CPU was Xeon CPU E5-1620 3.60 GHz with 8 GB memory, and its OS was Windows 7 (64 bit); DBMS was 5.6.17 version of MySQL; the transaction control was performed by its database engine InnoDB; the concurrent execution was implemented by Thread class of Java.

4.2.1 Consistency Check before Commit

To evaluate the requirement mentioned in Section 3, we performed experiments using the tables shown in Fig. 8. Prior to each experiment, we set 10 thousand data to the $account$, which bank balance was one million; set 9 thousand data of transfer amount A to $result$ to perform the bank transfer, and their result flag R (transfer result) was set to "null." Here, to examine the account transfer about both of the success and failure cases, we calculated the value of A by the following

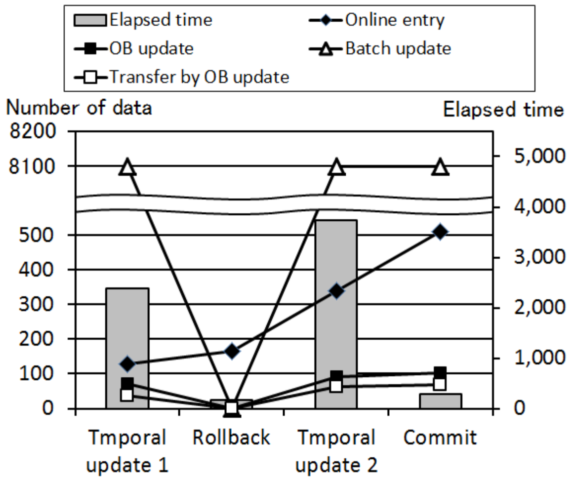


Figure 10: Transition of number of data about rollback.

function of K.

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

Concurrently, the online entries were executed from 5 terminals, by which 50 % of the account balance of each bank account was transferred to its corresponding bank account. Each was executed in the interval of about 0.3 sec. And, if the online entry conflicted with the batch update, then its OB update was 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 Fig. 9, we show the experimental results from the pre-commit until the commit completion in the temporal update. In this figure, "Online Entry" shows the change of the number of the data in $account$ updated by the online entries with the elapsed time; "Transfer by OB Update" shows the same number updated by the OB update. Next, "OB Update" shows the change of the number of the data in $result$ inserted by the OB update. Here, even though the bank transfer is failed due to the lack of the account balance, the result is inserted to $result$. But, $account$ 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 pre-commit of 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 was as follows. In addition, the above-mentioned online entries were executed over this experiment, and the data of target tables were queried after the each stage. Firstly, the temporal update and its pre-commit were executed; secondly, the rollback of the batch and OB update was executed; thirdly, the temporal update was rerun for the data of this time, and its pre-commit was executed; finally, the commit of the temporal update was executed.

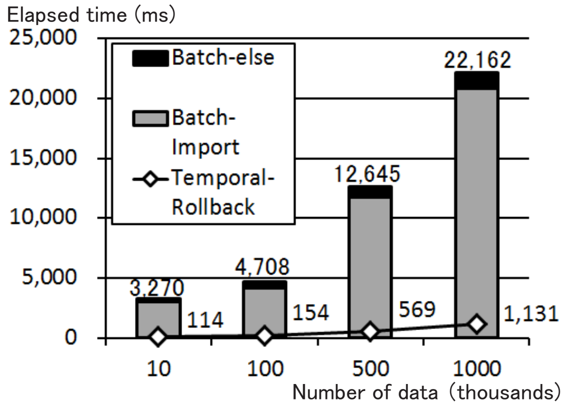


Figure 11: Comparison of elapsed time for recovery.

We show the experimental result in Fig. 10 similar to Fig. 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. Fig. 10 shows, in addition to the result shown in Fig. 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 cancelled at this stage, could be queried. On the other hand, as for the requirement about the affecting to the online entries, since this rollback did not 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 Fig. 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 Fig. 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 are cleared, and the target table is restored by the backup.

In Fig. 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 increase, 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 Fig. 7, both of the online entry and OB update could be composed as a single transaction. Second, as shown in Fig. 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 Fig. 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 cancelled by the rollback. Finally, as shown in Fig. 11, this rollback is very efficient comparing with the conventional recovery method: about 20 times in the experimental case. Moreover, as shown in Fig. 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 Fig. 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 Fig. 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 should be checked before their commit. And, as above-mentioned, 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 views 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, for example, unlike the mini-batch, it is not necessary to split the update into the short time transactions, or to implement the recovery methods such as the compensating transactions. So, we consider that the temporal update is useful in the case where the implementation like this is necessary in other methods.

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 is not a transaction as the whole processing. So, the latter cannot do its rollback and cannot maintain the isolation. And, the data have 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 Fig. 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 Fig. 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 cannot 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] H. Berenson et al., “A Critique of ANSI SQL Isolation Levels,” Proc. ACM SIGMOD 95, pp. 1-10 (1995).

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

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

[4] N. Edelweiss, P.N. Hübler, M.M. Moro, and G. Demartini, “A Temporal Database Management System Implemented on top of a Conventional Database,” Proc. XX International Conference of the Chilean Computer Science Society, pp. 58–67 (2000).

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

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

[7] T. Kudo, Y. Takeda, M. Ishino, K. Saotome, and N. Kataoka, “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).

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

[9] T. Kudo, Y. Takeda, M. Ishino, K. Saotome, and N. Kataoka, “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).

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

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

[12] B. Stantic, J. Thornton, and A. Sattar, “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).

[13] T. Wang, J. Vonk, B. Kratz, and P. Grefen, “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).

[14] D.S. Yadav, R. Agrawal, D.S. Chauhan, R.C. Saraswat, and A.K. Majumdar, “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).

(Received November 15, 2014)

Tsukasa Kudo received the M.E. from Hokkaido University in 1980 and the Dr.Eng. in industrial science and engineering from Shizuoka University, Japan in 2008. In 1980, he joined Mitsubishi Electric Corp. He was a researcher of parallel computer architecture, an engineer of application packaged software and business information systems. Since 2010, he is a professor of Shizuoka Institute of Science and Technology. Now, his research interests include database application and software engineering. He is a member of IEIEC, Information Processing Society of Japan and The Society of Project Management.

Masahiko Ishino received the master’s degree in science and technology from Keio University in 1979 and received the Ph.D. degree in industrial science and engineering from graduate school of Science and technology of Shizuoka University, Japan in 2007. In 1979, he joined Mitsubishi Electric Corp. From 2009 to 2014, he was a professor of Fukui University of Technology. Since 2014, he belong to Bunkyo University. Now, His research interests include Management Information Systems, Ubiquitous Systems, Application Systems of Data-mining, and Information Security Systems. He is a member of Information Processing Society of Japan, Japan Industrial Management Association and Japan Society for Management Information.

Kenji Saotome received the B.E. from Osaka University, Japan in 1979, and the Dr.Eng. in Information Engineering from Shizuoka University, Japan in 2008. From 1979 to 2007, he was with Mitsubishi Electric Corp., Japan. Since 2004, he has been a professor of Hosei business school of innovation management. His current research areas include LDAP directory applications and single sign-on system. He is a member of the Information Processing Society of Japan.

Nobuhiro Kataoka received the master’s degree in electronics from Osaka University, Japan in 1968 and the Ph.D. in information science from Tohoku University, Japan in 2000. From 1968 to 2000, he was with Mitsubishi Electric Corp. From 2000 to 2008, he was a professor of Tokai University in Japan. He is currently the president of Interprise Laboratory. His research interests include business model and modeling of information systems. He is a fellow of IEIEC.

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 technologies 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 show the concrete example of activity that successfully utilizes the geographical feature of Matsue City to teach children the concept of computing bottlenecks. In this activity we regard the geographical feature of Matsue as one of its regional materials.[1]

Keywords: Computer Science Unplugged, Teaching Methods, Regional Materials, Bottlenecks

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.

When a CS Unplugged activity is practiced by the pupils in an elementary school or a junior high school, most of those pupils could be local children in general. In this sense, utilizing regional or local materials as the content of the activity is considered to be effective, because the children must feel familiar with it.

In this context we have developed the new activity in which children can learn the phenomenon of bottlenecks [5]. As far as we have investigated, there is no activity that deals with bottlenecks. This activity utilizes the geographical feature of Matsue City in Japan, which we regard as one of its regional materials. It is a two-player board game and through tackling the 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 activity and also show the result of its practices by the children in a junior high school and a college of technology, both in Matsue City.

The rest of this paper is organized as follows. Section 2 describes the outline of CS unplugged. From Sections 3 through 5, the details of the newly proposed activity are shown. Section 6 presents the feedback from children. Section 7 deals with one of the future works with regard to utilizing regional materials in CS Unplugged. 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 Table 1. 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 activity. The CS Unplugged activities are characterized by features such as the followings [6].

[1]The work reported in the paper was supported by JSPS KAKENHI Grant Number 25350275.

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

- 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 TARGET CONCEPT OF NEW ACTIVITY

The concept addressed in the new activity is computing bottlenecks [7]. It is believed to be a new activity, as a preliminary survey of the current CS Unplugged activities did not show the same concept.

3.1 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.

Figure 1: Ohashi-gawa River vicinity in Matsue City

3.2 Geographical Bottleneck in Matsue City

Matsue City, where one of the authors' 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 Fig. 1. 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.

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.

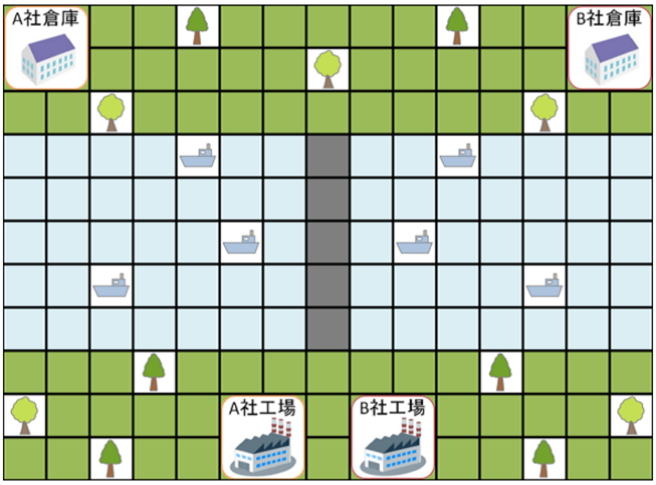
4 DETAILS OF THE ACTIVITY

4.1 Overview

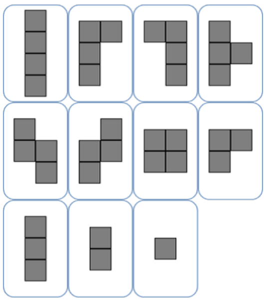
In this activity, two students compete against each other in a board game to see which one's pieces can reach the goal more quickly. The learning materials are the board and the pieces. In the game, they experience bottlenecks and devise strategies to relieve them.

4.2 Learning Materials and Rules

Figure 2 shows the following learning materials used in this activity: the board, eight pieces, and eleven bridge-building cards.



(a) Board



(c) Cards

(b) Pieces

Figure 2: Learning materials

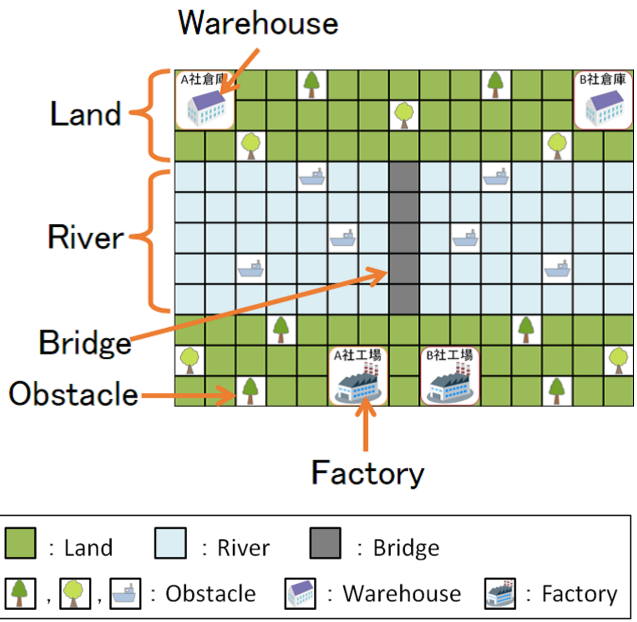


Figure 3: Board components

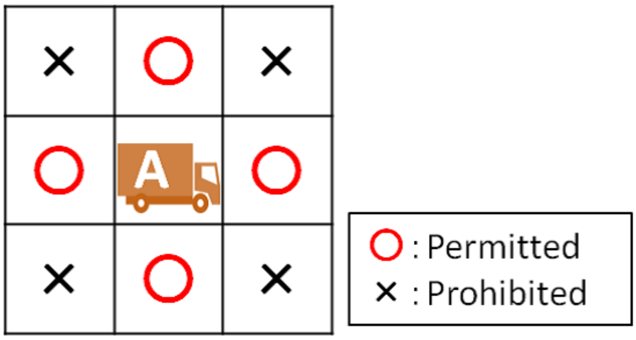


Figure 4: Permissible moves

4.2.1 Board

The board represents a city divided into north and south sides separated by a river, much like Matsue City. As shown in Fig. 3, 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.

4.2.2 Pieces

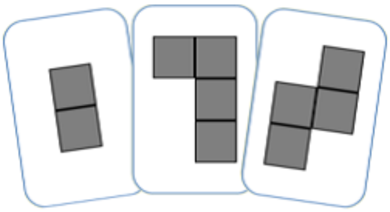
Each student has four pieces, which represent trucks that transport production materials and products between the warehouse and the factory of his or her company. The students take turns, with each student moving all four of his or her pieces per turn throughout the game. Each piece can be moved just 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 Fig. 4, 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.

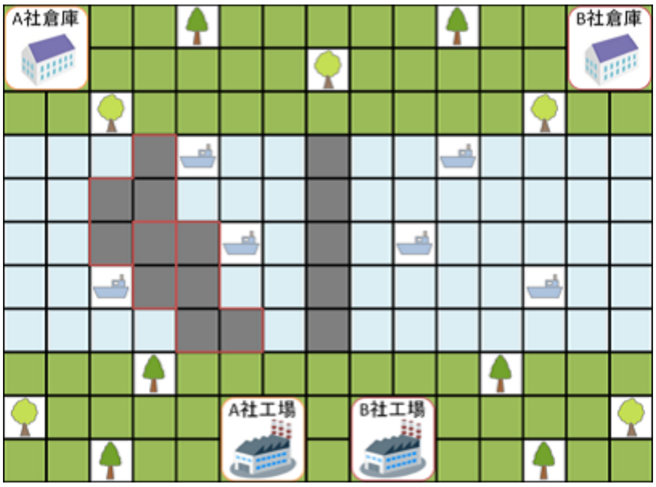
During one turn, the student must move each of the four pieces one space unless if all adjacent squares are already occupied or represent the river, obstacles, or the other company's facilities.

4.2.3 Bridge-building Cards

Each bridge-building card shows at most four squares arranged in a different block configuration, such as those shown in Fig. 2(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



(a) Chosen cards



(b) A possible bridge configuration

Figure 5: Bridge-building cards and their usage

shown on the card. The pieces of each student can be moved to any bridge addition space marked by either student.

Assuming that a student chose three cards as shown in Fig. 5(a), one of the possible bridge configurations during the game can be shown in Fig. 5(b). In this case, the student has used a card three times.

5 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.

5.1 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 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.

5.2 Practice

After hearing the story, the students first perform a practice game without the bridge-building cards. The board therefore

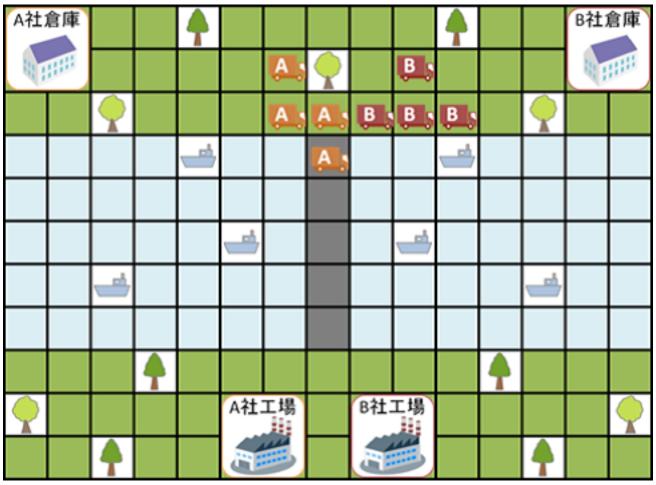


Figure 6: A traffic jam in the practice game

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 Fig. 6, 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.

5.3 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 7 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 Fig. 7, 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.

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.

6 EXPERIMENTS

We conducted three experiments for the activity mentioned above. The specifications of these experiments are shown in Table 2. A snapshot taken in the third experiment is shown in Fig. 8.

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

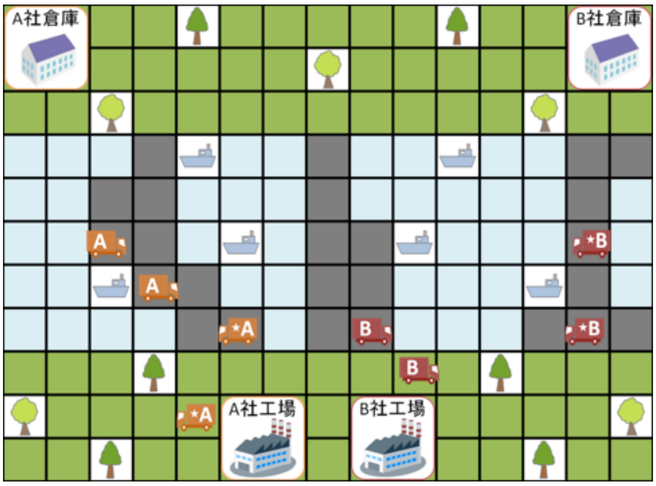


Figure 7: A game in progress with alleviated traffic congestion

Figure 8: Snapshot of the third experiment

experiments.

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 the concept of bottlenecks. The most important question was "Did you understand what bottleneck is?" The children were given four choices for this question: 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.

7 FUTURE WORK

As for utilizing regional materials in CS Unplugged, we could take a different approach from the one described above. This approach replaces the content of the existing activities with some regional materials [8].

One example 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 9 shows some examples. These images represent the materials local to Shimane Prefecture.

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

Table 3: Understanding level of participants

Did you understand what bottleneck is?		
Answer	Number	Percentage
Yes	24	71
Rather yes	8	24
Rather no	0	0
No	0	0
No answer	2	6

Another example 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 in some specific region, such as lullabies and folk songs.

We would like to take this approach as one of the future works for the study of utilizing regional materials in CS Unplugged, because its effectiveness should be confirmed through several experiments.

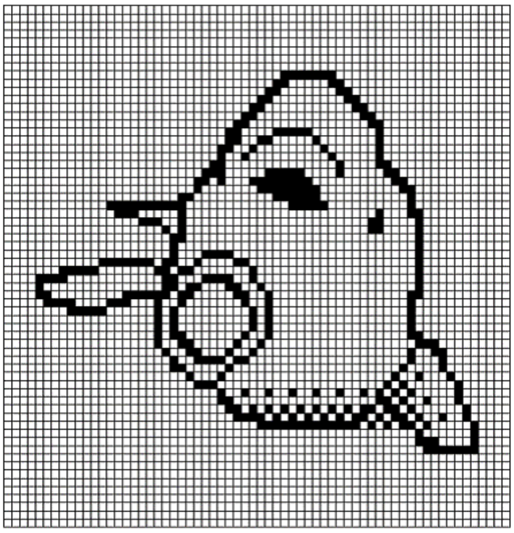
8 CONCLUSION

In this paper, we have proposed a new CS Unplugged activity utilizing regional materials. It has been 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.

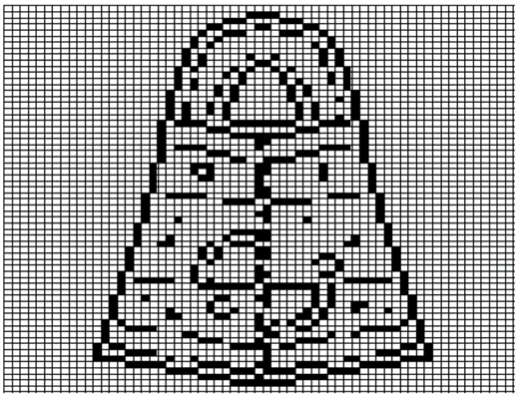
The target concept of the proposed activity is a computing bottleneck. For the learning experience, we designed a board game and materials in which the board is modeled after the geographical features of Matsue City, where one of the authors' former campus is located, and, in particular, on its bottleneck-inducing 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



(a) White dolphine and bubble ring



(b) Dotaku(Bell-shaped vessel)

Figure 9: Images local to Shimane

more easily and enjoyably.

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)," eText 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] 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).

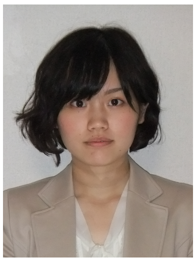
[6] T. Nishida, and S. Kanemune, "Learning Computer Science with Fun," IPSJ Magazine, Vol.50, No.10, pp. 980-985 (2009) (in Japanese).

[7] http://en.wikipedia.org/wiki/Bottleneck (2014.6.10).

[8] 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).

(Received November 20, 2014)
(Revised April 9, 2015)

Hisao Fukuoka received the B.E. and the M.E. in electrical engineering from Kobe University in 1976 and 1978, respectively. He joined Mitsubishi Electric Corporation in 1978. He received the Ph.D. degree in computer science from Shizuoka University in 1999. From 2003 to 2013, he had been a Professor of National Institute of Technology, Matsue College. He is currently a Specially Appointed Professor in the Department of Information Environment, Tokyo Denki University. His research interests include educational technology, embedded systems, and information theory. He is a member of IPSJ, IEICE, and IEEE.



Akane Kawakami graduated from National Institute of Technology, Matsue College in 2011. She received the B.E. in information engineering from National Institute of Technology, Advanced Course of Matsue College in 2013. She joined WACOM-IT Co.Ltd. in Shimane Prefecture, Japan in 2013. She is currently engaged in the development of Web application systems. Her research interests include educational technology, programming languages, and Web application systems.

Yuji Wada received the B.E. and the M.E. in electrical engineering from Waseda University in 1974 and 1976, respectively. He joined Mitsubishi Electric Corporation in 1976. He received the Ph.D. degree in computer science from Shizuoka University in 1997. He is currently a Professor in the Department of Information Environment, Tokyo Denki University. His research interests include database systems, data mining, and recommendation. He is a member of IPSJ, IEICE, JSAI, JSSST, and DBSJ.

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

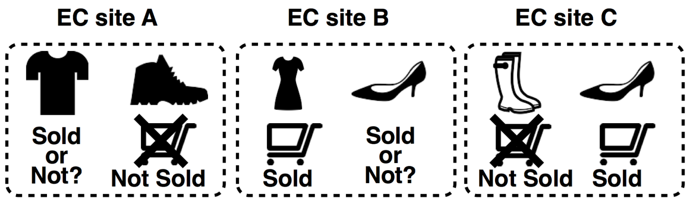


Figure 1: An illustration of our research settings. Traditional item purchase behavior research focuses on single EC site's information (in the above figure, EC site A). Our proposal collects purchase of multiple EC sites (in the above figure, EC sites A, B, and C) and constructs a model to predict whether the item would sell or not.

transfer. The algorithm is based primarily on *TrBagg*, which is an extension of bagging method, and efficiently drops inadequate base classifiers in aggregation phase.

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/

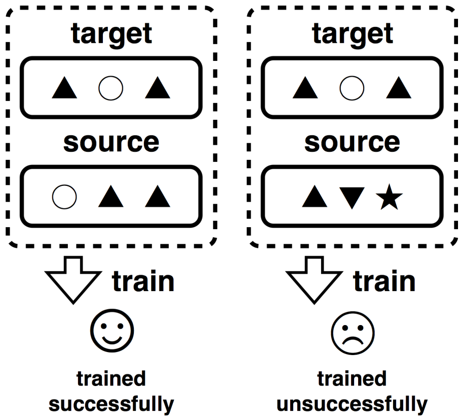


Figure 2: An example of negative transfer. In the left image, the target and source data have similar distribution and model training is successful. In contrast, the right image shows different distributions triggering negative transfer and the failure of model training.

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 $\mathcal{F}^* : \{\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$ $(n \leq N)$

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

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** $\mathcal{F}^* : \{\hat{f}_0, \hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n\}$ $(n \leq N)$

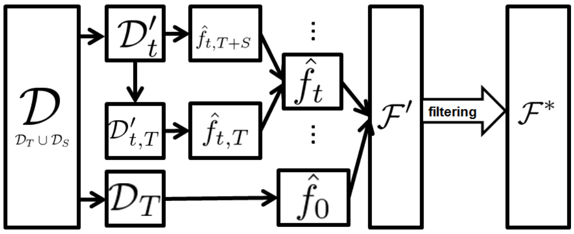


Figure 3: An overview of OptTrBagg. $\mathcal{D}'_t$ is generated by bootstrap sampling from both target and source data $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_S$. $\mathcal{D}'_{t,T}$ is extracted from $\mathcal{D}'_t$ by intersection of target data $\mathcal{D}_T$. Classifier $\hat{f}_t$ is 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.

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 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.

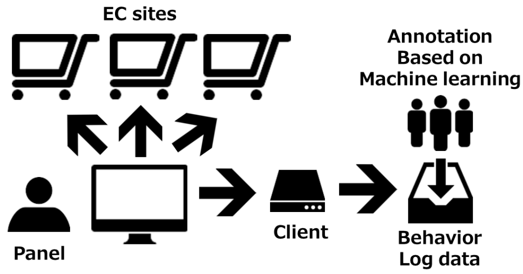


Figure 4: Data crawling scheme used by the project. Client software installed on Panel member's personal computers captures information about purchase behavior such as items that the Panel bought or queries they entered in Google, and sends them to our server. These data are converted into format suitable 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.

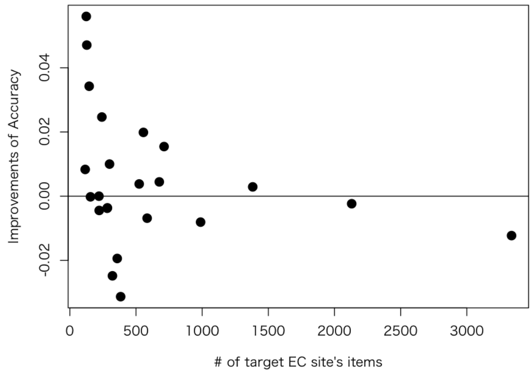


Figure 5: A scatter plot of the item size of target EC sites (X axis) and the accuracy improvements offered by OptTrBagg (Y axis).

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

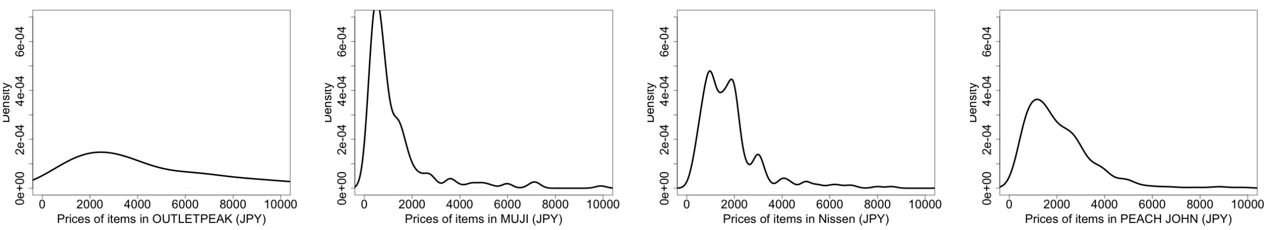
For Experiment 2, we prepared to set 3 parameters; the number of initial base classifier N, the top K size using name based features, and the data size of bootstrap $|\mathcal{D}'_t|$. At first, the number of initial base classifiers N was fixed to 100. In name based features, the top $K = 10$ items 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 supe-

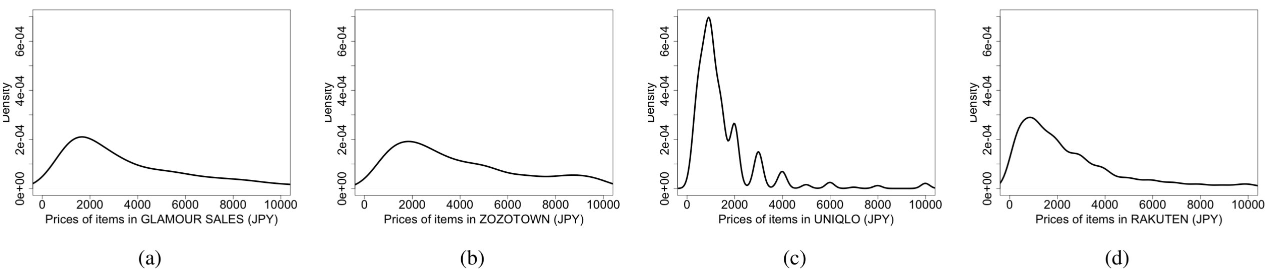


Figure 7: Price distribution of items in each source EC site used in Experiment 2, GRAMOUR SALES (Fig. 7(a)), ZOZOTOWN (Fig. 7(b)), UNIQLO (Fig. 7(c)), and RAKUTEN (Fig. 7(d)).

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

riority 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 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, pp. 1377–1386 (2009).

[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, pp. 1099–1100 (2009).

[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 (2003).

[5] W. W. Moe and P. S. Fader, "Dynamic conversion behavior at ecommerce sites," Manage. Sci., vol. 50, no. 3, pp. 326–335 (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, pp. 130–137 (2010).

[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, pp. 1805–1814 (2011).

[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, pp. 1029–1036 (2010).

[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, pp. 409–418 (2012).

[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, (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, pp. 148–156 (1996).

[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, pp. 159–166 (2010).
[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, pp. 1077–1078 (2010).
[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, (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, pp. 863–870 (2010).
[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, pp. 287–296 (2011).
[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, pp. 627–634 (2008).
[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, pp. 219–228 (2009).
[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, pp. 193–200 (2007).
[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, pp. 464–469 (2006).
[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, pp. 256–263 (2007).
[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 (1998).
[26] J. R. Quinlan, "C5.0," http://mloss.org/software/view/329/ (2011).
[27] R. Quinlan, "C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)," 1st ed. Morgan Kaufmann, (1992).

(Received November 20, 2014)
(Revised March 10, 2015)

Noriko Takaya Senior Research Engineer, Supervisor, Co-innovation Promotion Project, NTT Service Evolution Laboratories. She received the B.A. degree in in English and American Literature from the Gakushuin University in 1987. In 1995, she joined NTT Advertising, Inc. Since 2010, she has been in NTT Service Evolution Laboratories. She is a member of Information Processing Society of Japan, Japan Marketing Academy and Japan Institute of Marketing Science.

Yusuke Kumagae Yusuke Kumagae received his B.E. and M.E. degrees from University of Tsukuba in 2009 and 2011, respectively. He now works at NTT Service Evolution Laboratories. His research interests include machine learning and consumer behavior.

Yusuke Ichikawa Senior Research Engineer, 2020 Epoch–making Project, NTT Service Evolution Laboratories. He received the B.E. and M.E. degrees in measurement engineering from Keio University, Kanagawa, in 1995 and 1997, respectively. He joined NTT Multimedia Network Laboratories in 1997. Since 1998, he has been engaged in R&D of recommendation systems. He is a senior member of the Information Processing Society of Japan (IPSJ) and received the 2005 IPSJ Yamashita SIG Research Award.

Hiroshi Sawada Hiroshi Sawada received the B.E., M.E. and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993 and 2001, respectively. He joined NTT Corporation in 1993. From 2009 to 2013, he was the group leader of Learning and Intelligent Systems Research Group at the NTT Communication Science Laboratories, Kyoto, Japan. He is now a senior research engineer, supervisor at the NTT Service Evolution Laboratories, Yokosuka, Japan. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable model, graph-based data structure, and computer architecture.

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 this information through a network is called a "probe information system". Recently, such probe information systems have attracted attention and been used to share traffic information. In this study, we focus on road surface conditions concerning driving comfort and ride quality. We need to share data, 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 changes in road conditions changed in winter, even within the same segment.

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

1 INTRODUCTION

Recently small, high-performance sensors have become widespread due to the development of MEMS (Micro Electric Mechanical Systems), and are embedded in various kinds of object in our living environment such as personal computers, beacons, cars and smartphones. Many researchers tackle advanced studies in the field of mobile sensing. 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 at 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 this 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 driving comfort and ride quality. In winter, uneven surfaces 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.

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.

2.1 An Approach Using Fixed Cameras

There is a study that uses fixed cameras on poles on a road to estimate 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.

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 application 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 a 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

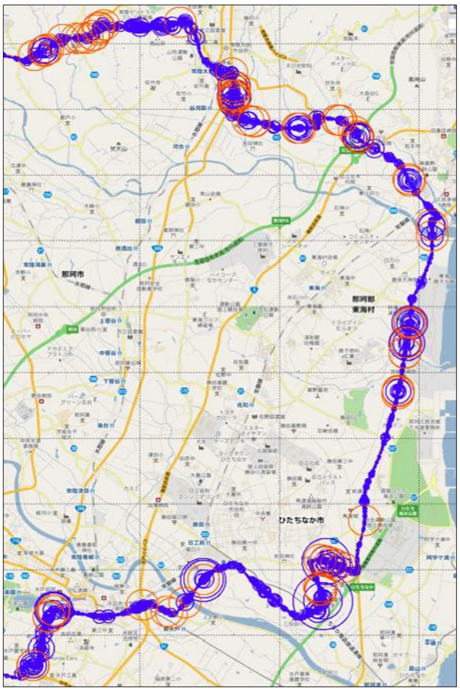


Figure 1: The relationship between estimation result and visual check (Quoted from [16]).

granularity. However, the method cannot estimate 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]. Method [11] is based on iterating multibody analysis and uses a multibody vehicle model. Methods [12]-[14] use three-axis acceleration sensors and a GPS sensor embedded in a vehicle. Methods [15][16] involve 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 a GPS sensor, and defines it as the length of a bump. In Fig. 1, results of estimation of road bumps and actual visual confirmation of road bumps are shown. The results of estimation are shown as blue circles, and visual confirmation is shown as orange circles.

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

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

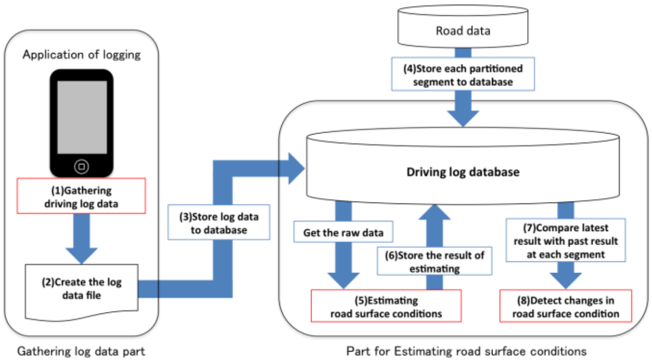


Figure 2: An overview of the proposed system.

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 with 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.

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 and give a detailed overview of the proposed method.

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 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. Smartphones are often used on a car dashboard 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 driving log data because they can be used in all weathers and at all times. 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 and longitude, and time stamps during driving. Gathered driving log data are managed in a database, and we estimate road surface conditions and detect changes in them by using the driving log data. An overview of the proposed system is shown in Fig. 2.

The proposed system consists of two stages: gathering log data and estimating road surface conditions. First, the system 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 system generates a log data file, and stores log data in the database (2), (3). Our method partitions a road from intersection to intersection into multiple segments, and creates a road segment table that is then stored in the database as a preliminary preparation (4). In the second stage, the system 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 system 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.

Table 3: Structure of log data table.

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

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

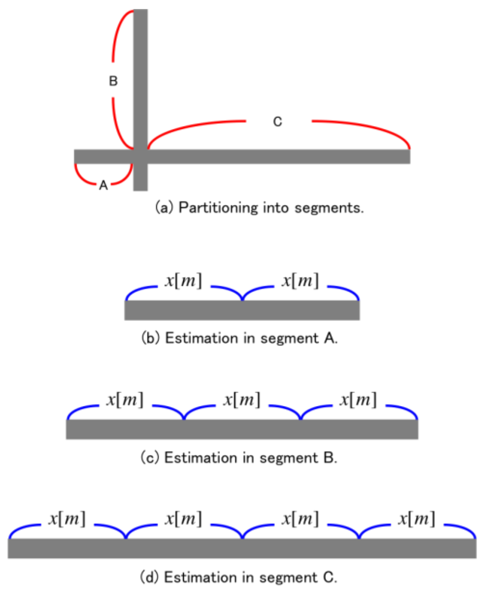


Figure 3: An illustration of partitioning into segments and estimating in the partitioned segment.

Table 5: 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
-	…	…	…	…	…

3.3.1. Gathering Driving Log Data

Our method gathers driving log data by using smartphones on car dashboards. 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 in a raw data table in the database, through the mail function of smartphones. The structure of the log data table is shown in Table 3.

This study focuses on the vertical component of acceleration values because it is the piece of log data most affected by the impact of road bumps. Accordingly, we estimate road surface conditions using changes in the vertical component of acceleration values.

3.3.2. Estimating Road Surface Conditions

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

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 the vertical component of acceleration values in a constant segment and setting thresholds. The thresholds are explained in Section 4.2. We need to estimate road surface conditions at every fixed interval when we use thresholds. However, lengths of actual roads and driving routes vary. In addition, we need to compare estimation results for each individual segment to detect changes in road surface conditions. For these reasons, the proposed method partitions a road from intersection to intersection into multiple segments and further partitions those segments into sub-segments of constant length. It then estimates road surface conditions of each partitioned segment, in order to solve the problem of varying length of segments. Figure 3 shows an illustration of partitioning into segments and estimating in each partitioned segment, and Table 5 shows an example of segment table data.

Actual lengths from intersection to intersection are different, as shown in Fig. 3 (a). Therefore, the proposed method continues to calculate every x $[m]$ until it reaches the end of a segment, as shown in Fig. 3(b), (c) and (d). 'x $[m]$'. The variable used for calculating the interval of variance is explained in Section 4.2. The estimated results at

Table 6: Structure of 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
-	…	…	…	…	…

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):

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

This $Est(seg\_id)$ is stored in the database table of estimation results with an estimation result ID, date and time stamp, as shown in Table.

In this way, we classify roads of various lengths into three levels. Moreover, recording estimation results in units of segments enables comparison of the results of each individual segment.

3.3.3. Detecting Changes in Road Surface Conditions

We detect changes in road surface conditions by using a table to compare estimation results 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 estimation results using the following process. We define the number of estimation results with 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 value of past results $Past\_Est(seg\_id)$ is calculated as follows (after the decimal point is rounded):

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

The reliability of estimation results is high when there is a wide range of past data to refer to. However, this changes depending on 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

Figure 4: Image of setting a smartphone.

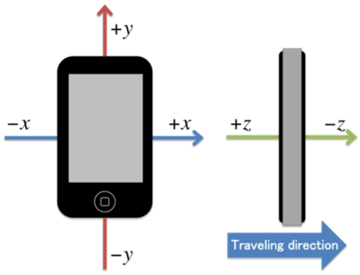


Figure 5: Acceleration axis.

varies according to season and purpose. Therefore, we define that range is set by the user, based on these variables.

Next, we compare latest estimation $Latest\_Est(seg\_id)$ with past estimation $Past\_Est(seg\_id)$. We can assume that road conditions remain constant if $Past\_Est(seg\_id)$ is equal to $Latest\_Est(seg\_id)$. However, if $Past\_Est(seg\_id)$ is not equal to $Latest\_Est(seg\_id)$ we can assume that road surface conditions have changed within the past few hours.

4 EXPERIMENTS AND DISCUSSIONS

In this section, we explain the experiments we conducted to confirm 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 using this data, and visualizes the results. We used Java and JDBC (Java Database Connection) to implement the processing of estimation of road surface conditions and detection of changes in these conditions, and we also used JavaScript to implement an application for visualizing estimation results.

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 Fig. 4. In this case, the y-axis of acceleration becomes the vertical component of acceleration values, as shown in Fig. 5.

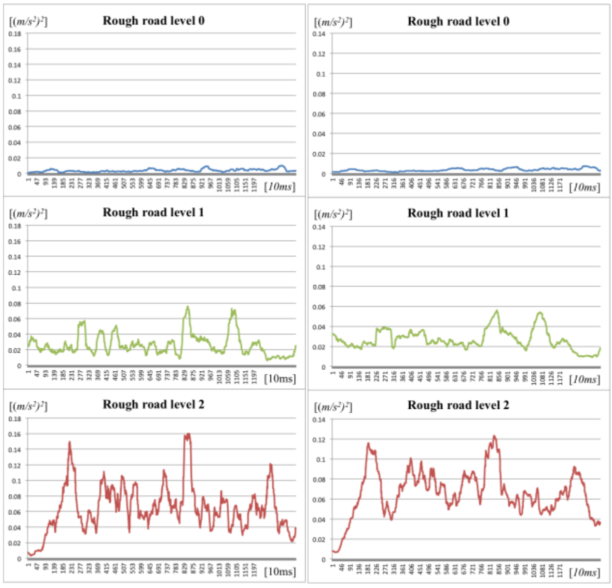


Figure 6: Variance in every 5 and 10 meters.

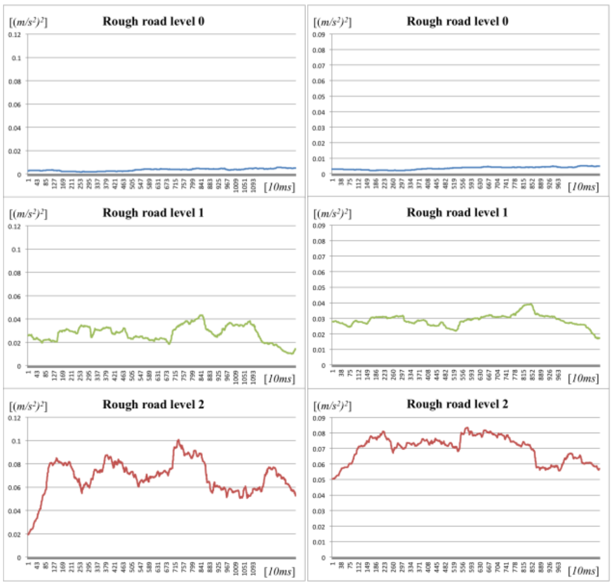


Figure 7: Variance in every 20 and 40 meters.

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

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 thresholds that classify the rough road level distinctively. On the other hand, 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

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

(a) Segment (b) Every 40 meters

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

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 to discriminate between rough road level 0 and 1 at $0.0190[(m/s^2)^2]$, and 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 using the calculation interval and thresholds established in Section 4.2. We analyzed the estimation result of each segment 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.

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 level 2 were successfully estimated but there were false estimations when a rough road level 1 was estimated as level 0. We analyzed the log data of false estimations to find the cause. The estimation result of a segment in which false estimation occurred is shown in Fig. 8(a). The estimation result of every 40 meter interval in the segment is shown in Fig. 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.

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

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

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

that rough road level 1 was estimated at multiple sites, as shown in Fig. 8(b). However, the estimation result produced using the proposed method shows the segment 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 change depending on date and time. We used a visualization application to analyze driving log data in winter when road surface conditions often change. Estimation results of driving log data collected when we drove on a snowy segment at night are shown in Fig. 9(c), and estimation results of driving log data collected when we drove in the same segment in good weather on the following day are shown in Fig. 9(d). 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.

On the day in which the data used in Fig. 9(c) was collected, some roads were covered in snow due to snowfall on the previous day. The red line in Fig. 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 Fig. 9(d), because the temperature increased and the snow melted in the daytime.

4.5 Discussion

In this section, we discuss the experiment results and other factors affecting vertical acceleration values.

4.5.1. Discussion of 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 the vertical component of acceleration values. Moreover, we were also able to confirm that rough road level 0 and 2 can be estimated accurately. So we think the proposed method can indicate road surface conditions to a user more clearly than the method used in study [16]. 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 level 1. Therefore, the proposed method confuses level 1 with level 0 when it estimates the road as segments. Accordingly, we need to consider ways to improve our method, such as adding new parameters and further partitioning longer segments.

From the results of the elementary experiment for detecting changes, we think the proposed method is able to detect changes in rough road levels in winter, even within the same segment.

4.5.2. Discussion of Factors Affecting Acceleration Values

There are many factors that have an influence on vertical acceleration values. One of the factors is type of vehicle. Length of wheelbase and strength of suspension differ depending on the type of vehicle, and we think these aspects have an influence on vertical acceleration values. The other factor is a individual driving style. We think acceleration values change in a different way depending on driving technique and experience. We need to check the relationship between these factors and vertical acceleration values in the future.

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 its effectiveness. 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. Additionally, we have to research the relationship between IRI and rough road levels estimated by the proposed method.

REFERENCES

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy and M. B. Srivastava, "Participatory sensing," World Sensor Web Workshop, ACM Conference on Embedded Networked Sensor Systems (2006).

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

[3] H. Yamauchi, A. Tomono and A. 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] H. Yamauchi, F. Oka and H. 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] A. Kubota, N. Kitajima, Y. Kobayashi and S. 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] Y. Miyazaki, H. Okabe, K. Tanji, M. Otsuki and S. 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] T. Kuno, H. Sugiura and J. 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] M. Tanaka, T. Morie, S. Matsuoka, K. Iwase and Y. Yamamoto, "Dry/Wet judgment of road surface using Gabor filtering of vehicle camera images," ITE Technical Report, Vol. 35, No. 9, pp. 263-267 (2011) (*in Japanese*).
[9] T. Furukane, K. Shibata and Y. 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] Y. Hinagata, M. Bando and Y. Kawamata, "In-vehicle stereo camera system for off-road environments," Multimedia, Distributed, Cooperative, and Mobile (DICOMO2012) Symposium, vol. 2012, pp. 1375-1381 (2012) (*in Japanese*).
[11] N. Takahashi, T. Shiba, K. Morita and Y. Endo, "Estimation of road profile with multibody vehicle model," Proceedings of the 19th JSME Transportation and Logistics, pp. 105-108 (2010) (*in Japanese*).
[12] K. Chen, M. Lu, X. Fan, M. Wei and J. 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] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden and H. 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] K. D. Zoysa, C. Keppitiyagama, G. 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] K. Yagi, "Extensional smartphone probe for road bump detection," 17th ITS World Congress, pp. 1-10 (2010).
[16] K. 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*).

(Received November 20, 2014)
(Revised February 15, 2015)

Tomohiro Nomura received B.E degrees from Future University Hakodate in 2013. His research interests include mobile sensing and ITS. He currently works at his master's degree in Graduate School of Systems Information Science, Future University Hakodate.

Yoh Shiraishi received doctor's degree from Keio University in 2004. He is currently an associate professor at the Department of Media Architecture, School of Systems Information Science, Future University Hakodate, Japan. His research interests include database, mobile sensing and ubiquitous computing. He is a member of IPSJ, IEICE, GISA and ACM.

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. Here, some people would be bored to read all review comments and quit it in the halfway, but note that we actually need to read considerable amount of review comments if we wish to grasp the real evaluation from text-based evaluation 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]–[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

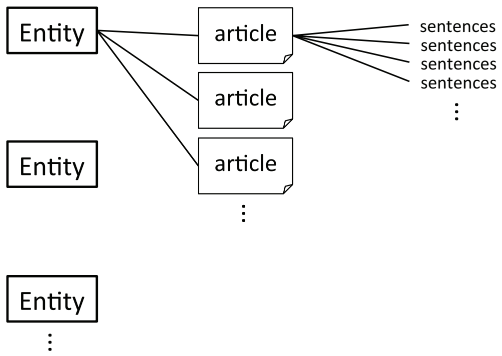


Figure 1: The Structure of Review-site Data

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 of restaurants, "Good restaurant for family with reasonable cost" can be a typical practical aspect description that is useful for many people. 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).

Figure 2 illustrates the overview of the proposed method. First, (a) we learn a dictionary of evaluation words from a

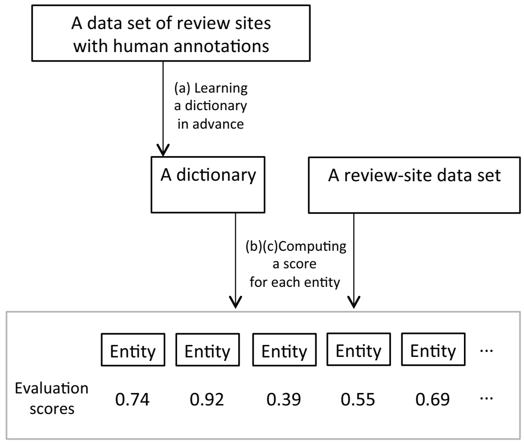


Figure 2: The Process of The Proposed Method

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 data-set 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.

(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.

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

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 $P_{w,a}$ 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.

Figure 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:

$$P_s = \begin{cases} \textit{positive}, & (\text{if } T_p < \frac{1}{|W_p^s|} \Sigma_{w \in W_p^s} P_{w,a}), \\ \textit{neutral}, & (\text{if } -T_p \leq \frac{1}{|W_p^s|} \Sigma_{w \in W_p^s} P_{w,a} \leq T_p), \\ \textit{negative}, & (\text{if } \frac{1}{|W_p^s|} \Sigma_{w \in W_p^s} P_{w,a} < T_p). \end{cases}$$

Figure 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*.

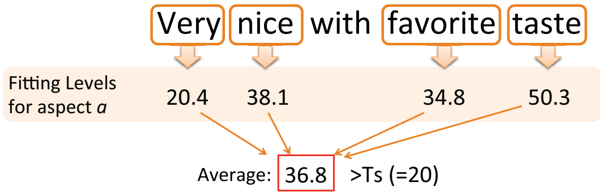


Figure 4: Judging Aspect for An Evaluation Sentence

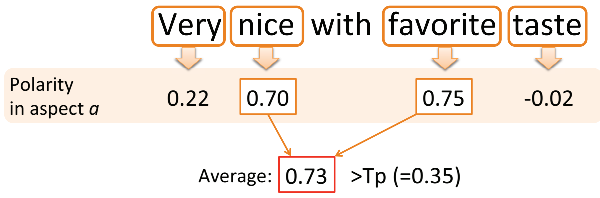


Figure 5: Judging Polarity for An Evaluation Sentence

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

$$Score(i) = \frac{|\{s | P_s = \textit{positive} \text{ and } s \in E_i\}|}{|\{s | P_s = \{\textit{positive} \text{ or } \textit{negative}\} \text{ and } s \in E_i\}|}$$

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

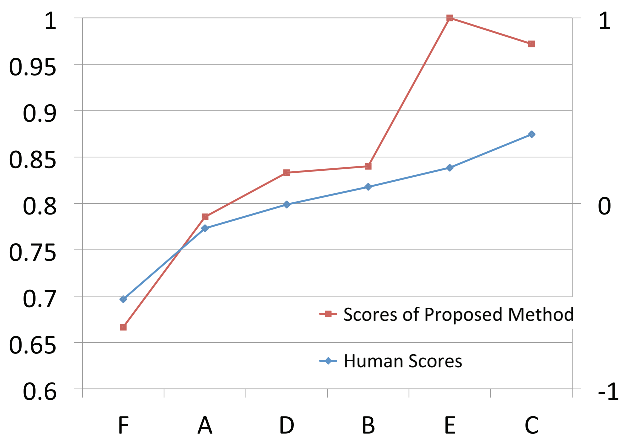


Figure 6: Human and Computed Scores for Restaurants under Aspect “Taste” (Experiment 1)

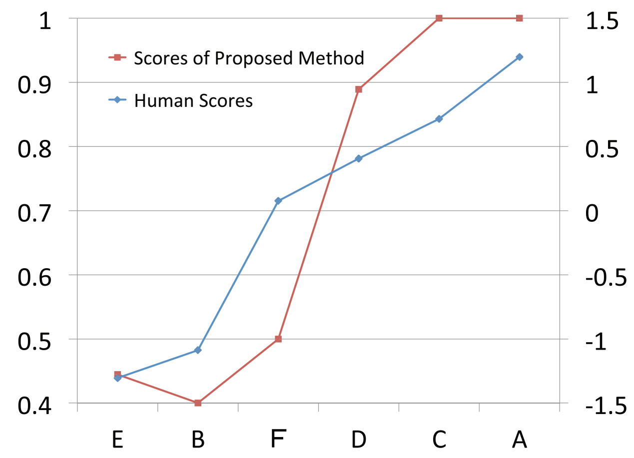


Figure 7: Human and Computed Scores for Restaurants under Aspect “Price” (Experiment 1)

Figure 8: Human and Computed Scores for Ra-men Restaurants under Aspect “Quality of Noodles” (Experiment 2)

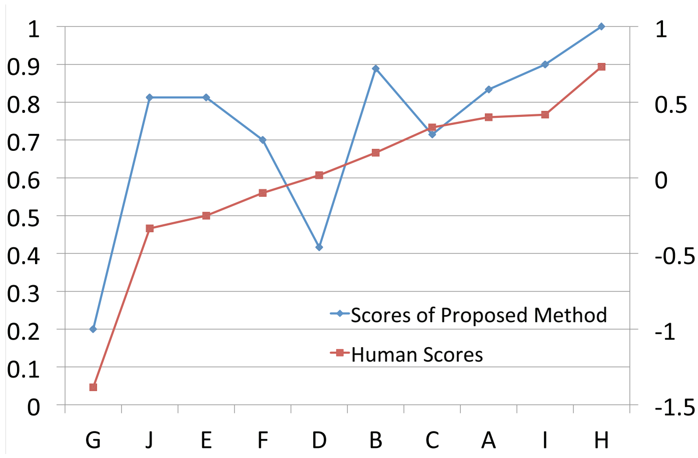


Figure 9: Human and Computed Scores for Ra-men Restaurants under Aspect “Taste of Soup” (Experiment 2)

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 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.” Note that these two aspects would be still so simple and would not be as complicated as usual practical descriptions. However, in this paper, we use these two aspects to perform a first-step investigation to clarify the basic property of the proposed method.

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 restau-

	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)

rants 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 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.

Figures 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.

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

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.

One of the most important future task on the proposed method is to cope with the problem described in Section 3.4, i.e., the problem that likes and dislikes exist in several evaluation words. Although by nature this problem occurs inevitably, we have several choices to avoid the inconvenience that comes from the problem. The easiest solution is to detect the words that include likes and dislikes of people and exclude them from the computation of evaluation score. Another choice, which would be a more challenging solution, would be to classify people into two groups, say, 'like' group and 'dislike' group, and show the scores of both groups to the users who use our method. To propose and evaluate the methods to achieve either of them would be challenging work toward practical use of our method.

We finally note that someone would consider that the cost to create a dictionary for each aspect could be a problem for the practical use of the proposed method. Although the proposed method actually requires considerable human labor, we would note that the laborious task would be performed easily with low cost if we use a tool called crowd sourcing. With this useful tool, the proposed method would be one of realistic methods that works in practice.

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 Sentences 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), pp.111-120 (2008).
[11] Tabelog, http://tabelog.com/ (In Japanese, referred to in 2015).

(Received Nobember 10, 2014)

Satoru Hosokawa received the B.E. and M.E. degrees from Wakayama University in 2011 and 2013, respectively. He is currently working with Yahoo Japan Corporation.

Etsuko Inoue received the B.E., M.E. and Ph.D. degrees from Wakayama University in 2002, 2004 and 2007, respectively. She is an Assistant Professor in Wakayama University from 2007. She is interested in database systems, web applications, data visualization, and so on. She is a member of IPSJ.

Takuya Yoshihiro received his B.E., M.I. and Ph.D. degrees from Kyoto University in 1998, 2000 and 2003, respectively. He was an assistant professor in Wakayama University from 2003 to 2009. He has been an associate professor in Wakayama University from 2009. He is currently interested in the graph theory, dis- tributed algorithms, computer networks, medial applications, and bioinformatics, and so on. He is a member of IEEE, IEICE, and IPSJ.

Masaru Nakagawa received the B.E, M.E, and Ph.D. degrees from Osaka University in 1970, 1972, and 1990, respectively. He was a researcher in NTT laboratory from 1972 to 1994, and from 1994 he was a Professor in Kinki University. He is a Professor in Wakayama University from 1994. He is interested in Database Design and Computer System Engineering. He is a member of IPSJ, JSAI and JSIK.

A Study of the Synchronization Process for Collaboration between Product Development and Fundamental R&D

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: basic 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.

Figure 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 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

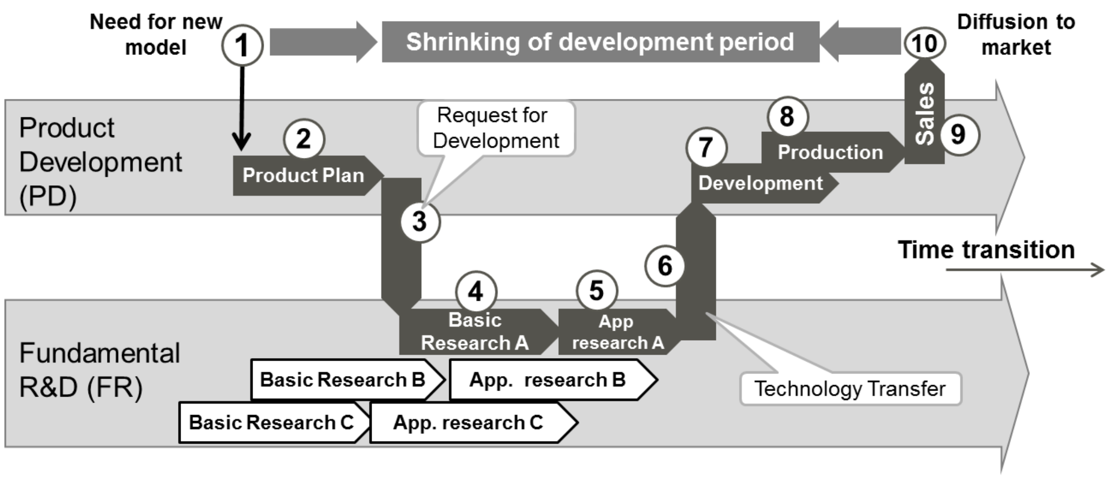


Figure 1: Fundamental R&D and processes of market-driven product development

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).

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

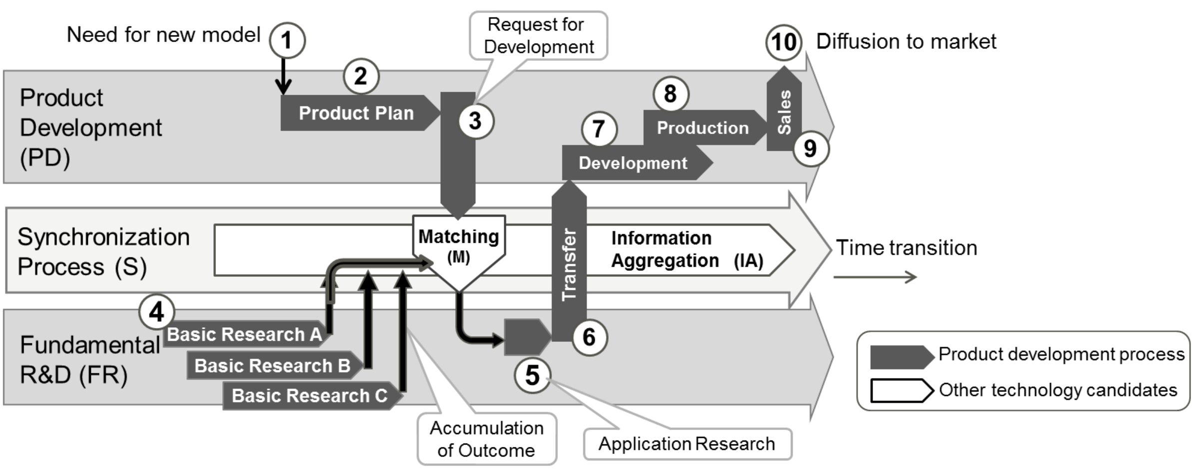


Figure 2: Core concept of synchronization process

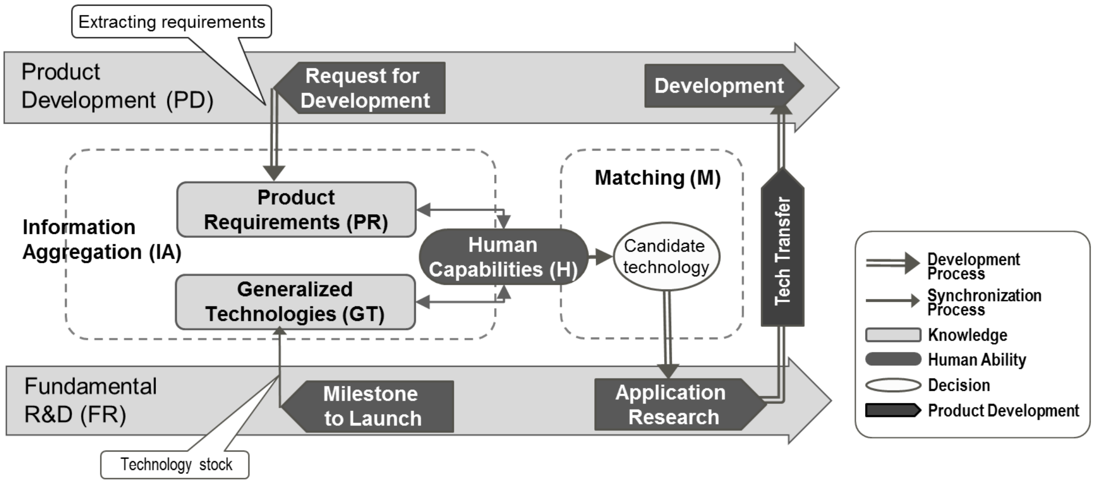


Figure 3: Internal components of synchronization process

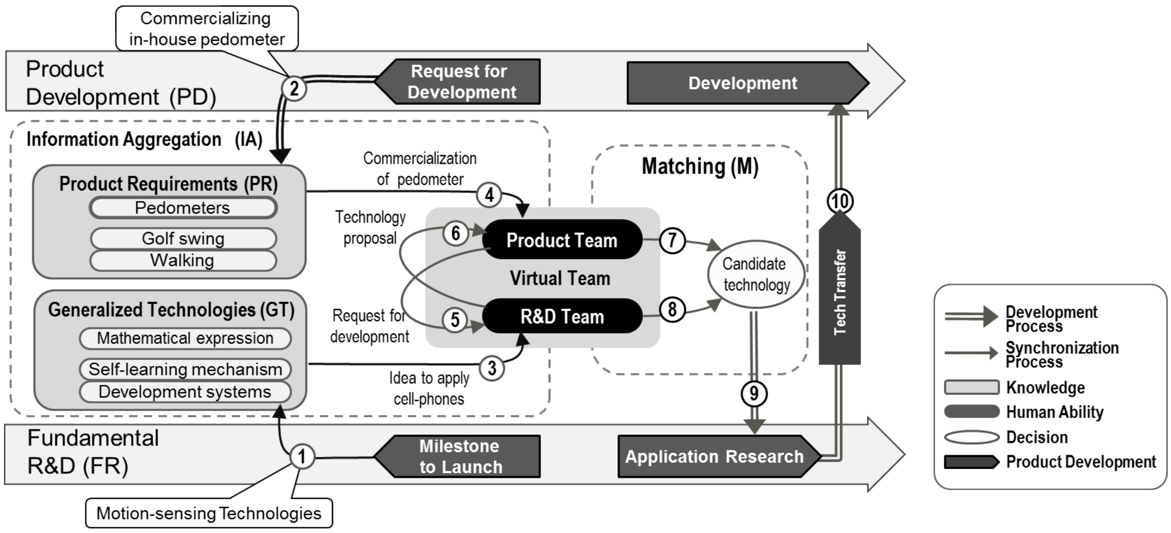


Figure 4: Synchronization process: Motion-sensing technology

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 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

Figure 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).

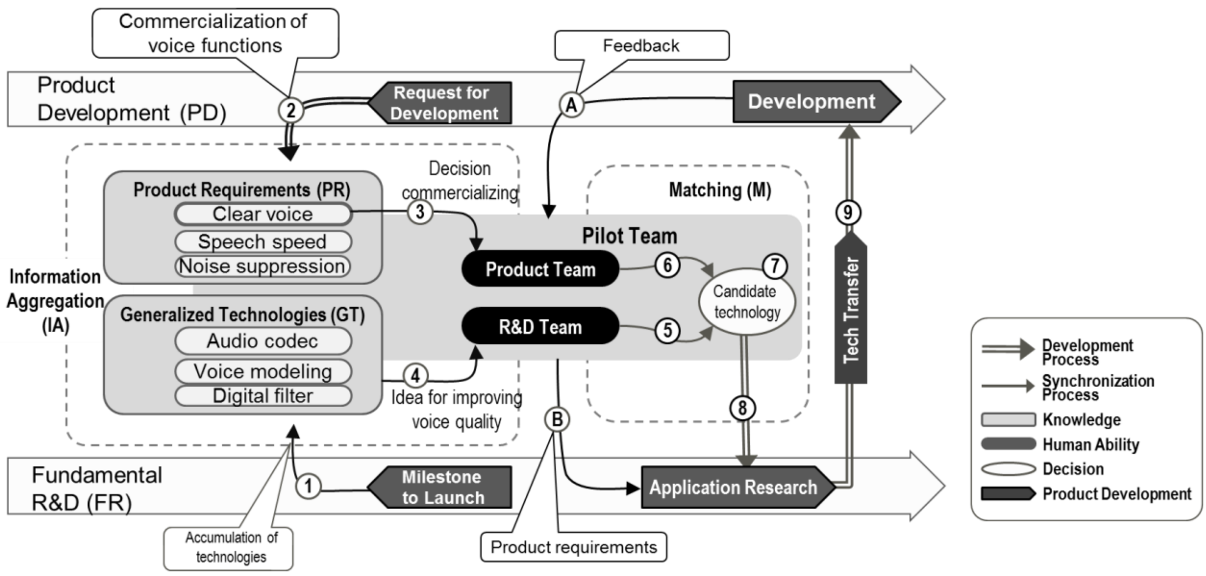


Figure 5: Synchronization Process: Voice-processing technologies

(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.
(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

Figure 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

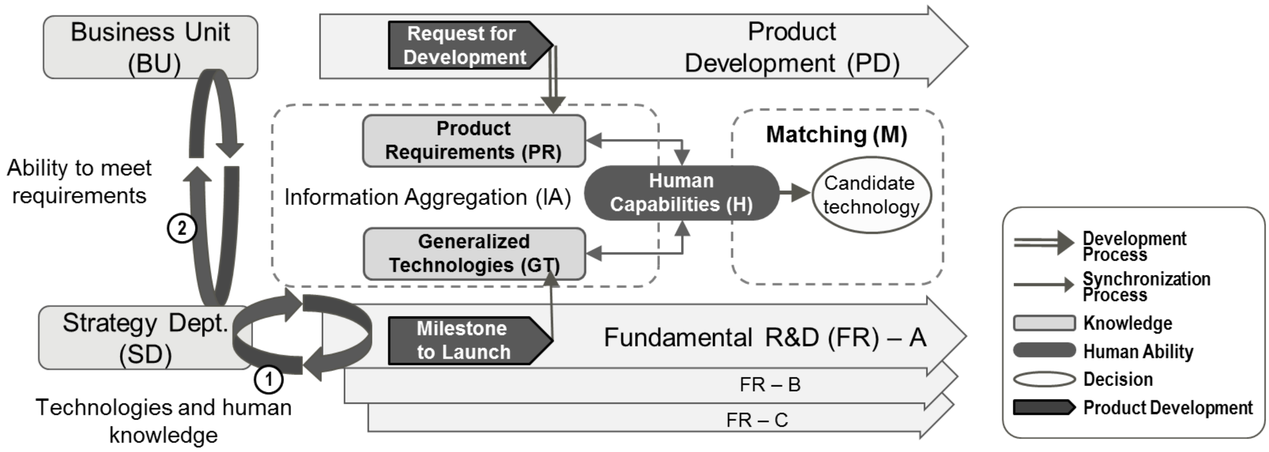


Figure 6: Synchronization process: Operation Model

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 (fundamental R&D and strategy dept. in Fig. 6). 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.

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] L. Branscomb, P.E. Auerswald, "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, Vol. 64, No. 1, pp. 137-146 (1986).

[5] B. Godin, “The Linear model of innovation the historical construction of an analytical framework,” Science, Technology & Human Values, Vol. 31, No. 6, pp. 639-667 (2006).
[6] S. J. Kline, N. Rosenberg, “An Overview of Innovation,” The positive sum strategy: Harnessing technology for economic growth, pp. 275-305 (1986).
[7] S. J. Kline, “Innovation is not a linear process,” Research management, Vol. 28, No. 4, pp. 36-45 (1985).
[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, Vol. 30, No. 1, pp. 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, pp. 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, pp. 397- 401 (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, pp.43-50 (2011).
[13] K. Hayashida, et al., “Development Concept and Functions of Raku-Raku PHONE,” FUJITSU, Vol. 61, No. 2, pp. 184-191 (2010).
[14] K. Chujo, et al., “Human Centric Engine and Its Evolution toward Web Services,” FUJITSU Scientific & Technical Journal, Vol. 49, No. 2, pp. 153-159 (2013).
[15] Fujitsu Introduces Miniature Humanoid Robot, HOAP-1, Fujitsu press release on 10 September (2001), URL: http://pr.fujitsu.com/en/news/2001/09/10.html (accessed March 2nd, 2015).
[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 (accessed March 2nd, 2015).
[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 (accessed March 2nd, 2015).
[18] Cloud Service Launched for Wandant Dog Pedometer, Fujitsu press release, November 27 (2012), URL: http://www.fujitsu.com/global/about/resources/news/press-releases/2012/1127-01.html (accessed March 2nd, 2015).
[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).

(Received November 12, 2014)

Tatsuo Tomita

Tatsuo Tomita is Chairman and Director of Fujitsu Laboratories Ltd. based in Japan, as part of the Fujitsu Laboratories Group with R&D sites in Japan, China, the U.S., and the U.K. Mr. Tomita was appointed Chairman of Fujitsu Laboratories Ltd. in 2014, after serving as President of Fujitsu Laboratories Ltd. from 2010, after an extensive career at Fujitsu Limited spanning over three decades in which he held various executive and senior management positions in Computer and Mobile Phone-related groups, including appointments as Member of the Board and Corporate Senior Executive Vice President in charge of Fujitsu’s Product Business Group; Corporate Senior Vice President and President of the System Products Business Group; and Corporate Vice President and President of the Mobile Phones Business Unit. Mr. Tomita received his Bachelor of Science (B.S.) in science from the University of Tokyo in 1972. His industry-wide activities in Japan include being a member of the Sub-Committee on Planning, Committee on Industrial Technology of the Keidanren of Japan, committee member of the Council on Competitiveness-Nippon (COCN), and member of the Steering Board for the Tsukuba Innovation Arena (TIA) of Japan.

Yoichiro Igarashi

Yoichiro Igarashi is a researcher in charge of Management of Technology (MOT) in the R&D Strategy and Planning Office, Fujitsu Laboratories, Ltd. He received his B.E. and M.E. in electric engineering from Tokyo University of Science in 1993 and 1995, respectively. He joined Fujitsu Limited in 1995 as an engineer of telecommunication systems and moved to Fujitsu Laboratories Ltd. in 1998. He also received a master’s degree in Management of Technology (MOT) from the Graduate School of Innovation Studies, Tokyo University of Science, in 2012. He is a researcher at the Research Center for Practical Wisdom, Fujitsu Research Institute.

Masao Yamasawa .
Masao Yamasawa received his B.E. and M.E. from Tokyo Institute of Technology, Japan in 1968 and 1970, respectively. In 1973, he received a Ph. D. in Electrical and Electronics Engineering from the same Institute. He joined Fujitsu Laboratories Limited as a researcher of ultra-high frequency digital processing and analog-to-digital conversion LSI's. In 1990, he moved Fujitsu to join development of the second generation (i.e. digital) mobile communication system, including the i-mode phones. In 2003, he served as Group President of the Ubiquitous System Research Center of Fujitsu Laboratories Ltd. After leading development of RFID and application systems, he returned Fujitsu Mobile Phone Business Group to promote mobile technology standardization and industry-university cooperation. He left Fujitsu in 2014 but remain an advisor to the Mobile Phone BG. He is a member of IPSJ.

Kaoru Chujo
Kaoru Chujo was appointed Vice President of Advanced Technologies, Ubiquitous Business Strategy, of Fujitsu Limited in 2013. She started her career at Fujitsu Limited as an engineer of signal processing technologies. She led R&D projects involving network technologies at Fujitsu Laboratories of America (FLA) during 2000-2003. She is currently in charge of developing advanced technologies for mobile phones including cameras, multimedia, and smartphones in the mobile phone business unit.

Masayuki Kato
Masayuki Kato was appointed Head of the R&D Strategy and Planning Office of Fujitsu Laboratories in 2010. He joined Fujitsu Laboratories Ltd. in Japan in 1983, and was previously a researcher in fields including holographic optical elements, material engineering for optical and photonic devices, and electronics packaging technologies related to semiconductors for servers and supercomputers. Mr. Kato received his B.S. and M.S. degrees in Applied Physics from the Tokyo Institute of Technology in Tokyo, Japan in 1981 and 1983, respectively. He is a member of Institute of Electronics, Information and Communication Engineers (IEICE), Japan Society of Applied Physics (JSAP) and Japan Institute of Electronics Packaging.

Ichiro Iida
Dr. Ichiro Iida received B.S., M.S. and Ph.D. degrees in electronics engineering from the University of Tokyo in 1978, 1980 and 1983, respectively. He joined Fujitsu Laboratories Ltd in 1983. He has been engaged in the research and development on computer network and distributed software and he is now the fellow in Fujitsu labs. Ltd., responsible for the human centric computing research group.
His research interests include the Internet architecture, Web technologies and mobile ubiquitous computing.
He is a member of IPSJ and IEICE japan.

Hiroshi Mineno
H. Mineno received his B.E. and M.E. degrees from Shizuoka University, Japan in 1997 and 1999, respectively. In 2006, he received his Ph.D. degree in information science and electrical engineering from Kyushu University, Japan. Between 1999 and 2002, he was a researcher in the NTT Service Integration Laboratories. In 2002, he joined the Department of Computer Science of Shizuoka University as an Assistant Professor. He is currently an Associate Professor. His research interests include sensor networks as well as heterogeneous network convergence. He is also a member of ACM, IEICE, IPSJ, and the Informatics Society.

Submission Guidance

About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

Computer supported cooperative work and groupware
Intelligent transport system
Distributed Computing
Multi-media communication
Information systems
Mobile computing
Ubiquitous computing

Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word™

Sample document http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

Copyright

Publisher

Address: Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan
E-mail: secretariat@infsoc.org

CONTENTS