Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I,
IMECS 2014, March 12 - 14, 2014, Hong Kong

Serendipitous Recommendation Method based on WWW Access Log

Toshiharu Nakashima, Kenji Nakamura, and Shigeru Oyanagi

Abstract—**Recommendation is widely used in the various EC site to help a user search desired items. Serendipity is recognized as an important dimension for user satisfaction. Various methods for serendipitous recommendation have been proposed, however most of them utilize content information. Hence, they depend on the content information, which makes it difficult to apply to other fields.**

This study proposes a serendipitous recommendation method without using content information. It uses WWW access log, and divide a session into groups of items based on the co-occurrency. Serendipity is generated by matching between groups.

The proposed method is evaluated by large scale access log on EC site. Two experiment on serendipity and precision are carried out. The result shows that the proposed method can provide serendipity without decreasing precision. The proposed mechanism is also set up on a real EC site, and the feasibility is discussed.

Index Terms—**recommendation, serendipity, WWW access log, accuracy,**

I. Introduction

As the explosion of information on the Web, user's effort to search desired information is rapidly growing. Recommendation has been actively studied in order to provide user's favorite information efficiently [1]-[9]. Typical recommendation methods can be classified into two types, one is based on content information [1],[2], and the other is based on user information[3]-[6].

Recommendation method based on the content information utilizes the characteristics of each content beforehand, and recommend items which are similar to the user's interest by using content information. The precision of recommendation by this method depends on the amount and quality of content information. The merit of this method is that it works well even if user's information cannot be prepared.

Collaborative filtering is a tipical recommendation method based on the user information[3]-[6]. It recommends items which are calculated by other users information with the similar preference. The precision of collaborative filtering is high when sufficient amount of user information is available.

Collaborative filtering can be classified into two classes, one to use explicit information generated by the user[3][4], and the other to use implicit information automatically gathered by user's behavior[5][6]. The merit of the former method is that the recommended items are directly related to user's input. Hence, the user can improve the quality of recommendation by adjusting user information. The merit of the latter method is that the user need not prepare his information. In general, this method uses implicit information obtained from the WWW. Recently, collaborative filtering method using implicit information such as access log[6] is actively studied.

Nowadays, recommendation systems are widely used in the EC site. Various types of recommendation methods are used as mentioned above, each of which has its merit and demerit. However, there are common problems in these recommendation methods, one of which is called serendipity. Serendipitous recommendation helps the user find a surprisingly interesting item he might not have otherwise discovered. Hence, serendipity is recognized as an important dimension for user satisfaction. In contrast, conventional recommendation methods tend to recommend items that are already known to the user. The range of recommended items tends to be limited in a narrow area. These problems may cause the decrease of user satisfaction.

In order to solve this problem, recommendation method considering the serendipity has been actively studied. The existing method of serendipitous recommendation depends on the content information which are carefully prepared. This approach lacks for applicability because of dependency to content information.

In this study, we propose a method of serendipitous recommendation without using the content information. The proposed method only uses access log, which enables to apply to wide range of EC sites. This study also discuss the precision of the serendipitous recommendation.

This article is organized as follows. Section two describes related works, focused on serendipitous recommendation. Section three explains the purpose of this study, and explains the algorithm of the proposed method. Section four evaluates the proposed method with respect to precision and serendipity by the access log of EC site. Section five evaluates feasibility of the proposed system by setting the proposed method on a real EC site. Section six concludes this work.

II. Related Works

The recommendation method considering serendipity presents different information from the result that can be easily expected. Swearingen proposed a serendipitous recommendation method by increasing the diversity of the recommended item [7]. It removes similar items from the recommended item list, and attempts to increase diversity in the book recommendation. It also adds recommended items of low similarity to the recommended item list, thereby avoiding the uniformity of item in the recommendation list. Ziegler reported that the method improves the diversity of

Manuscript received December 23, 2013; revised January 22, 2014.

T.Nakashima is with the Guraduated School of Information Science and Engineering, Ritsumeikan University 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan (e-mail:is029080@ed.ritsumei.ac.jp)

K.Nakamura is with the Faculty of Information Management, Osaka University of Economics 2-2-8 Osumi, Higasiyodogawa, Osaka, 533-8533 Japan (e-mail: k-nakamu@osaka-ue.ac.jp)

S.Oyanagi is with the College of Information Science and Engineering, Ritsumeikan University 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan (e-mail: oyanagi@cs.ritsumei.ac.jp)

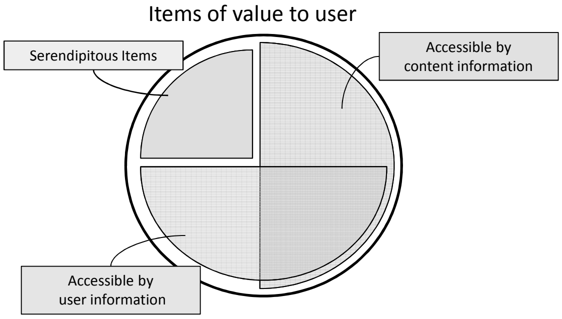

Fig. 1. Classification of recommended item

recommendation item list although recall and precision rate is reduced [8].

Yoshimi proposed serendipitous recommendation using both the content information and usage history of the user[9]. It is applied to movie recommendation, and improves the serendipity by recommending unexpected movies. It extracts features of a movie by classifying movies conceptually, and selects unexpected movie by associating with the click history. Thereby, it is reported that it has succeeded in the conceptual connection between click history of the user and movie information, to increase the user satisfaction of the recommendation item. Compared with the conventional recommendation methods that take into account only to enhance recall and precision, these studies have succeeded in improving the serendipity. However, these methods require to calculate relevance of the items by using the content information. Therefore, recommended items will depend on the content information, which restricts the recommendation only in a predictable range of content.

In order to solve the dependency on the content information, we focus on the recommendation method using access log. In general, the access log is the result of the user selection, hence finding serendipity is difficult. However, it can be considered that a certain number of users exist to perform a serendipitous access in a large amount of access log. In this paper, we propose a method that extracts the user who made the serendipitous access from the access log, and utilizes his behavior for serendipitous recommendation.

III. Proposed Method

A. The Aim of This Study

This study aims to propose a method to solve the problem of conventional serendipitous recommendation method. Namely, dependency of a recommendation item to content information should be removed.

Fig.1 shows the classification of recommended items in terms of the accessibility from typical recommendation methods. Recommended items can be classified into the following four groups, a group of items that can be accessible by content information, a group of items that can be accessible by user information, a group of items that can be accessible by both information, and a group of items that cannot be accessible by both information. Conventional method for serendipitous recommendation takes the approach by content information. However, it requires high quality of large scaled content information and causes dependency to content information.

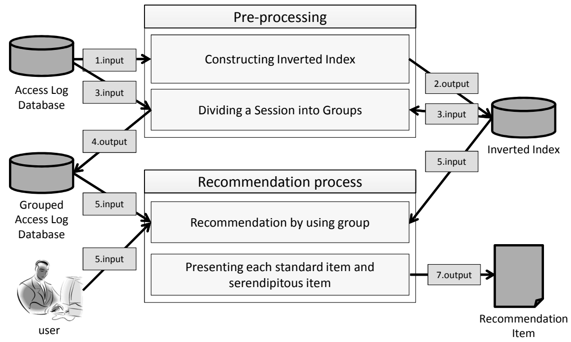

Fig. 2. Flow of Proposed System

Session ID	Item ID
1	A B C D B
2	B C D E
3	C D E
4	A C E
5	B D

Item ID	Session ID
A	1 4
B	1 2 5
C	1 2 3 4
D	1 2 3 5
E	2 3 4

Fig. 3. Example of Constructing Inverted Index

The purpose of this work is to propose a recommendation method that does not rely on content information. The proposed method is based on access log, and aims to extend the range of recommendation items to the serendipitous area.

B. Flow of Proposed Method

Fig.2 shows the flow of the proposed method. The proposed method can be roughly divided into two phases of the pre-processing and recommendation process.

The pre-processing is performed to create the inverted index. This is a data structure that is converted from session-based data into item-based data. By using this data structure, it is possible to efficiently access information on the specified item. In the pre-processing phase, each session is divided into groups of items based on the co-occurrence frequency of items.

In the recommendation process, recommendation items are generated using the divided item groups. A group is searched that is similar to the access log of the target user (hereinafter referred to as an active session), and the recommendation items are selected from items contained in the same session as the group.

C. Constructing Inverted Index

Fig.3 shows an example of conversion of the access log into an inverted index. Access log data is a session-based data structure which includes items in a session. It is converted into the inverted index which shows the sessions that the item is included. By this conversion, it is possible to obtain session data that are related to a specific item efficiently.

D. Dividing a Session into Groups

A session is divided into groups by using the co-occurrence frequency of each item in the session. An example of session division is shown in Fig.4, where a session (A, B, C, D, E) is divided. At first, the number of occurrences of each item and co-occurrence of items are calculated first.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

	A	B	C	D	E
A	32	15	16	2	1
B	15	46	18	3	2
C	16	18	22	1	2
D	2	3	1	26	19
E	1	2	2	19	29

⇒ Group (A,B,C) (D, E)

Fig. 4. Example of Split Session

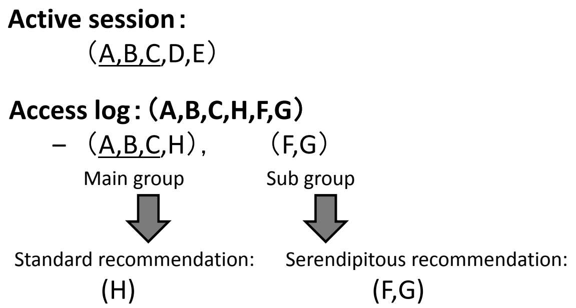

Fig. 5. Example of Recommendation

Then, the co-occurrence probability is calculated. Each item is classified into low-related and high-related item by a threshold. Finally, the session is divided into groups of high-related items together in the same group. In the example, the session is divided into (A,B,C) and (D,E), where items A,B, and C are mutually co-occurrent. At this time, if all items in the session are mutually co-occurrent, the session is to be constituted by one group.

E. Recommendation by Using Group

The recommendation by using the group is performed as follows. At first, a group is searched which is similar to the active session. Then, recommendation items are obtained from the items contained in the same session as the group. At this time, if recommendation items belonging to the group with the highest similarity are selected, then they correspond to a standard recommendation. If recommendation items that belonging to different groups within the same session are selected, they correspond to a serendipitous recommendation.

An example of the recommendation is shown in Fig.5. Assume that active session is (A,B,C,D,E) and access log is (A,B,C,H,F,G). Assume that the access log is divided into two groups (A,B,C,H), and (F,G). The group (A,B,C,H) is more similar to the active session, so it is called as a main-group. The group (F,G) is less similar, so it is called as a sub-group.

Recommendation of items from the main-group would lead to a standard recommendation because of recommending similar items to the active session. Recommendation of items from the sub-group would lead to a serendipitous recommendation because of recommending items of low degree of similarity. This process is performed on the whole log, and items for serendipitous recommendation are selected by the occurrences in the sub-group.

Furthermore, in order to enhance the serendipity, this process is repeated by using serendipitous items that are obtained from the active session. Fig.6 shows an example of

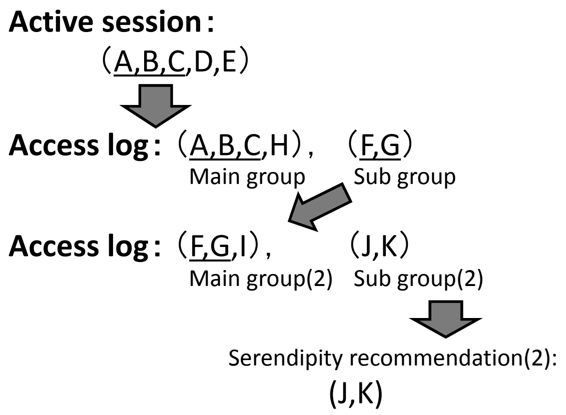

Fig. 6. Examples of Serendipitous Items Extension

serendipitous items extension. Where active session is (A, B, C, D, E), and access log is (A, B, C, H, F, G). The data (F, G) that are obtained as a result of serendipitous recommendation items, the calculation of serendipitous recommendation items is repeated from the associated session (F, G). In the example of Fig.6, access log of (F,G,I,J,K) is selected because the main group (F,G,I) is similar to (F,G), and the items in the sub group (J,K) are selected as serendipitous recommendation. Thus, serendipity of a recommendation is extended by performing the process recursively. The depth of the serendipity can be defined as the depth of recursion, to set the appropriate value according to the data set.

IV. Evaluation by Experiment

Two experiments are carried out to evaluate serendipity (Experiment 1) and accuracy (Experiment 2).

In Experiment 1, we define serendipity as a similarity to a conventional recommendation result that does not consider serendipity. Recommended items by the proposed method and those by a conventional method are compared. If the similarity between recommended items by the proposed method and items by the conventional method is low, serendipity of the proposed method is high.

In Experiment 2, we define accuracy as an existence of recommended items in the access log. If there exists a session that includes a serendipitous item in the similar situation, the recommendation of the item is successful, because the log shows the existence of real user who accessed the item in the similar situation. .

An example is shown in Fig.7. Assume a session of (A,B,C,D,E,F,G) where (A,B,C) is an active session and (D,E,F,G) are evaluation items. Assume also that (E,I,J) are recommended items. We search sessions that contain the recommended items (E,I,J) in evaluation items. If we find a session (A,B,F,D,E,I,J) where (A,B,F) is an active session and (D,E,I,J) are evaluation items, (A,B,F) is compared with the original active session (A,B,C).. In this example, (A,B,F) and (A,B,C) are similar because two items are common. It shows the existence of real session (user) that accesses the same recommended items (E,I,J) in the similar active sessions. Therefore, this recommendation is considered to be accurate.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

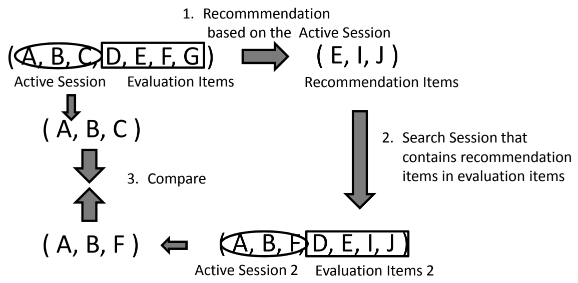

Fig. 7. Example of Successful Recommendation

A. Data for Experiment

Access log of EC site "TECH-JAM (http://www.tech-jam.com.)" for 31 months is used as the experimental data. Tech Jam is the EC site of the largest industry of physics and chemistry equipment sales in Japan, that deals with the product of about 100,000 items. Access log of 15-month period (140,649 sessions) is used for learning data, and of 16-month period (159,715 sessions) is used for evaluation data. In the evaluation data, the beginning three items of a session are handled as an active session, and the following items of the session are handled as the solution of the recommendation.

B. Evaluation of Serendipity

Fig.8 shows the result of the experiment on serendipity. Horizontal axis shows the depth of recursion for serendipity. Four bars correspond to serendipity recommendation of depth 0, 1, 2, and 3 respectively. Here, depth 0 means no consideration on serendipity. Vertical axis shows the ratio of matched items to the conventional recommendation method. It corresponds to the similarity to the conventional method.

The recommendation of depth 0 produces similar items to the conventional method. This is because both methods are doing with a recommendation based on the feature of user behavior. The recommendation of depth 1 produces significantly different items to the conventional method. Ratio of matched items is about 5 percent. In addition, the similarity of the recommendation of depth 3 is 0.05 percent. It shows that the recommended items are completely different from the conventional method. From this result, we can confirm that serendipitous recommendation of depth 1 and 3 can produce items of high serendipity.

In contrast, the similarity of the recommendation of depth 2 is about 20 percent, which is substantially higher than that of depth 1 and 3. It may be explained that serendipitous items from a serendipitous item is a little closer to the original item.

C. Evaluation of Accuracy

Fig.9 shows the result of experiment on accuracy. The horizontal axis represents the number of matched items. The vertical axis represents the ratio of match. Fig.7 shows the example of comparison with each active session. The ratio of match is defined by the possibility of match success.

In this experiment, the length of active session is fixed to three. Hence, the maximum number of matched item is three.

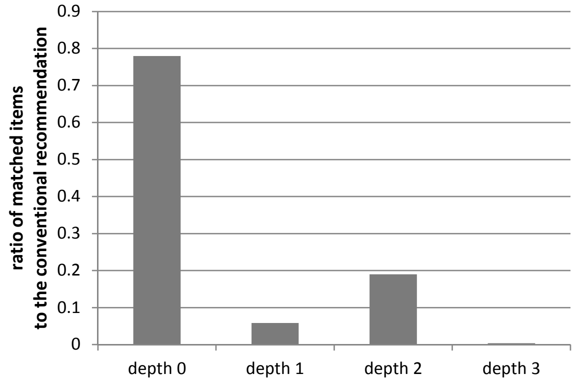

Fig. 8. Result of Experiment on Serendipity

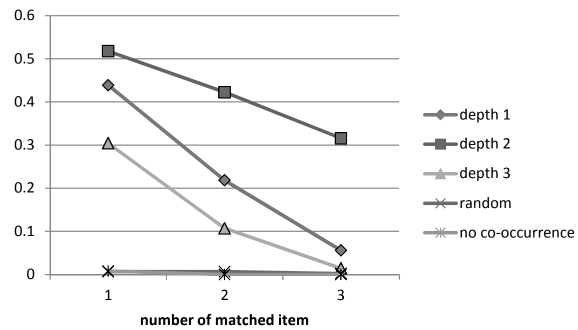

Fig. 9. Result of Experiment on Accuracy

In addition, experiments of recommending random items and recommending items of no co-occurrence are performed in order to compare the feasibility with the proposed serendipitous recommendation. As the result, there are no matched items by these methods. It shows that no log is found from items that are not related each other.

In contrast, in the proposed method, a certain number of related log exists. In particular, serendipitous recommendation of depth 2 can produce 30 percent of perfect match. It is confirmed that the proposed recommendation method produces related items at a fairy high level, which shows that the proposed method works well for certain level of accuracy.

V. Experiment on Real EC-site

We set up the proposed recommendation method on a real EC-site "TECH-JAM", and evaluated the feasibility by observing user's access. This experiment has been carried out about a month from November 1 to November 24 in 2013. The TECH-JAM system already equips a recommendation method based on collaborative filtering using the access log, so the proposed recommendation method is integrated with the existing one. The screen image of the TECH-JAM is given in Fig.10. The recommended items are shown in the righthand side, upper is the items of conventional method, and the lower is the items of the proposed method.

We evaluate the feasibility of the proposed method by comparing the number of access to the recommendation items between the proposed method and existing method. In this experiment, the depth of serendipity is set to 1, and length of active session is set to 1.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

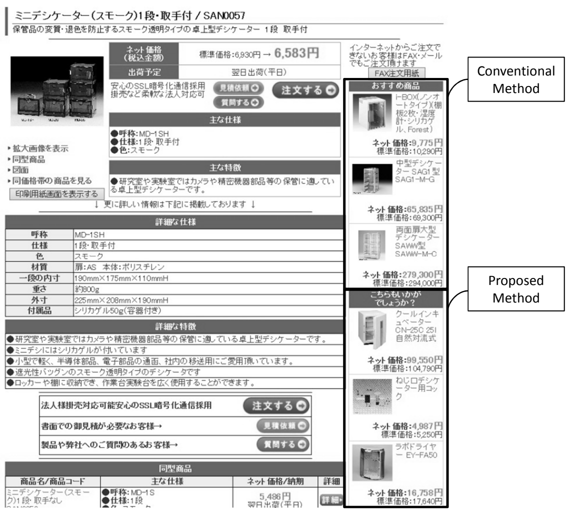

Fig. 10. Screen Image of TECH-JAM

A. Comparison of the Number of Successful Recommendations

Here, we define successful recommendation when the recommended item is really accessed in the session. Fig.11 compares the number of successful recommendations of the proposed method and conventional method. In the conventional method, 918 logs exist that access the recommendation items. In the proposed method, 657 logs exist that access the recommended items. The number of logs that move among items within the site is 11,023. Therefore, success rate is 8.33 percent in the conventional method and 5.96 percent in the proposed method.

This result says that the difference between the proposed method and the conventional method is not so large in recommending items which is interested by users. We have already convinced that the proposed method has high degree of serendipity in Experiment 1. The result shows that the proposed method can provide serendipity without decreasing accuracy.

B. Comparison of the Number of Successful Recommendations for each Item

Fig.12 shows the comparison of the number of successful recommendations for each item.

This graph shows the result of comparison for each item by the larger number of successful access through the conventional method or the proposed method. In 724 items, number of access by the conventional method is larger. In 465 items, number of access by the proposed method is larger. In 30 items, number of access is the same. From this result, there exists 38 percent of items that brings higher degree of satisfaction to the user than the conventional method. It can be seen that the proposed method can widen the area of recommendation items by avoiding too obvious items.

VI. CONCLUSION

In this study, we propose a method for serendipitous recommendation by using only the access log. The major feature of the proposed method is to classify items in a session into groups by co-occurence, and to select appropriate group by the degree of serendipity.

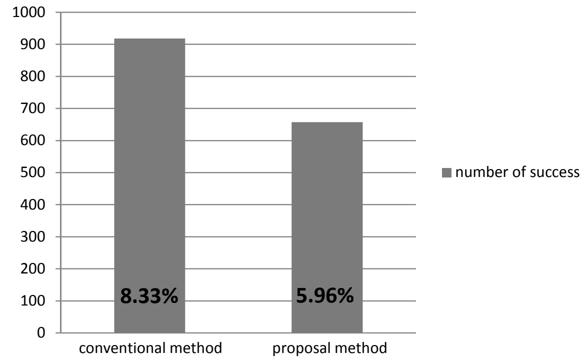

Fig. 11. Comparison of the Number of Successful Recommendatons

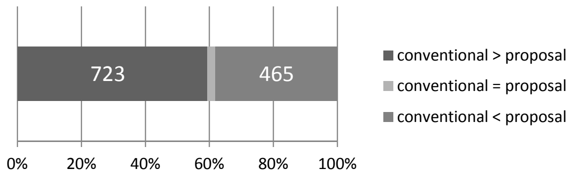

Fig. 12. Comparison of the Number of Successful Recommendation for each Item

The proposed method is evaluated by large scale access log of EC site.

The result shows that the proposed method can provide serendipity without decreasing accuracy. The proposed method is integrated into a real EC site, and demonstrate feasibility.

In the future, we plan to evaluate the effect of the depth of serendipity on the real site. We also plan to analyze the practical feasibility of serendipitous recommendation on the real EC site.

ACKNOWLEDGMENT

We thank to Tech jam Ltd. for providing access log of EC site, and integrating the proposed method into the site.

REFERENCES

[1] J. Alspector, A. Kolcz, and N. Karunanishi: Feature-based and Clique-based User Models for Movie Selection, User Modeling and User-Adapted Interaction, Vol.7, No.4, pp.279-304, 1997.
[2] R. Mooney, and L. Roy: ContentBased Book Recommending Using Learning for Text Categorization, Proceedings of the ACM Conference on Digital Libraries, ACM, pp.195-204, 2000.
[3] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. and Riedl, J.: GroupLens:Applying Collaborative Filtering to Usenet News, Comm. ACM, Vol.40, No.3, pp.77-87(1997).
[4] P. Resnick, N. lacovou, M. Suchak, P. Bergstrom, and J. Riedl: An Open Architecture for Collaborative Filtering of Netnews, CSCW 1994, ACM, pp.175-186, 1994.
[5] S.Middleton, N. Shadbolt, and D. Roure: Capturing Interest Through Inference and Visualization: Ontological User Profiling in Recommender Systems, Proceeding of the Second International Conference on
[6] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa,: Effective Personalization Based on Association Rule Discovery from Web Usage Data, Proceeding of the 3rd International Workshop on Web Information and Data Management, pp.9-15, 2001.
[7] K. Swearingen and R. Sinha: Beyond Algorithms: An HCI Perspective on Recommender Systems, ACM SIGIR Workshop on Recommender Systems, 2001.
[8] C. N. Ziegler, S. M. Mcnee, J. A. Konstan, and G. Lausen: Improving Recommendation Lists Through Topic Diversification, In pro.of World Wide Web Conference, pp.22-32, 2005.
[9] T. Yoshimi, H. Tsuji: A Proposal of the Recommendation System Providing Unexpected Information Using the History of User's Operation, IEICE Tech. Rep., vol. 110, no. 105, AI2010-7, pp. 37-41, June 2010.