

Technology for Recommending Optimum Learning Texts Based on Data Mining of Learning Historical Data

Yuji Wada^{*}, Yuuma Hamadume^{**}, Shinichi Dohi^{*}, and Jun Sawamoto^{***}

^{*} Department of Information Environment, Tokyo Denki University, Japan

^{**} Kobe Office, Konami Digital Entertainment Corporation Limited, Japan

^{***} Faculty of Software and Information Science, Iwate Prefectural University, Japan
 yujiwada@sie.dendai.ac.jp, DEN03738@nifty.com, dohi@sie.dendai.ac.jp, and
 sawamoto@iwate-pu.ac.jp

Abstract - We are developing a bidirectional recommendation system that extracts the relationship among digital texts with historical logs, and recommends the optimum texts for learners using data mining methods, such as collaborative filtering. In this paper, we first discuss the bidirectional recommendation and then show results from an evaluation of actual use. Finally, we propose a method for a collaborative learning recommendation system that mines the data of similar users sharing non-favorite subjects using historical logs and user attribute data.

Keywords: e-learning, data mining, recommendation, historical log analysis, collaborative filtering.

1 INTRODUCTION

In recent years, large numbers of institutions of higher learning, businesses and other organizations have been proactively introducing e-learning. That movement has been fostered in part by attention focused on the Web Based Training (WBT) approach [1], leading to the debut of numerous Learning Management Systems (LMS) [1]. Additionally, the proposal of the Sharable Content Object Reference Model, or SCORM [2], which is a global standard, has helped to spur the propagation of e-learning. Opinions are divided, however, as to whether the use of e-learning offers greater advantages to the learner than learning based on paper materials.

To address that question, firstly we implemented a “bidirectional recommendation system” [3] (see Figure 1) developed in our laboratory, in the AIRS “An Individual Reviewing System” [4]. Figure 1 shows a schematic of a bidirectional recommendation. When a learner is browsing the learning text “Basics of Assignment,” it is natural to advance to the next step “Basics of ‘while’ statement” or “Basics of ‘if’ statement.” However, browsing the basic contents “Variable types” again is also natural in learning. In other words, the learning efficiency is expected to improve by recommending not only learning texts frequently shifted from but also frequently shifted to “Basics of Assignment.”

This research is supported by a Grant-in-Aid for Scientific Research C (Subject No. 21500908: Research on Adaptive Recommendation Technology based on Bidirectional Recommendation Technology for E-Learning Texts).

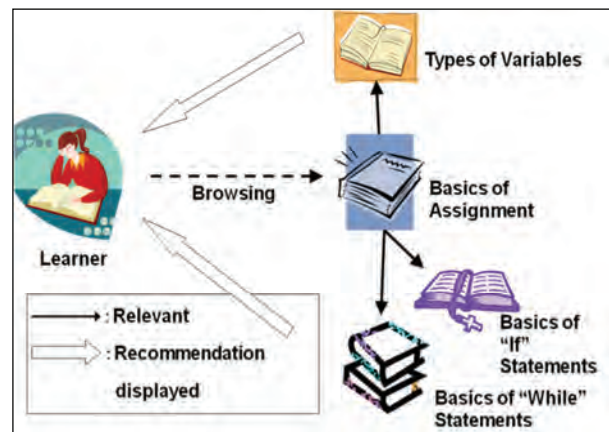


Figure 1: Bidirectional recommendation system.

Secondly, we asked 92 participants of the “Database System” lecture offered by the university to use the system between October 30 and November 5, 2008.

Subsequently, we conducted a survey using questionnaires that examined the actual situation of the user and the learning outcome achieved using the bidirectional recommendation system (see Table 1). In the survey, a number of respondents indicated that they were able to shorten the time spent learning, and the efficacy of learning using the bidirectional recommendation system was confirmed.

Moreover, a recommendation accuracy of 61% resulted from subjective evaluation by users of the appropriateness of the recommendation results (see Table 2). Some respondents indicated, however, that they preferred to browse the lecture materials, so we reexamined the functions requested by learners.

Table 1: User evaluation of bidirectional recommendation system.

Opinions of those indicating that recommendation results were appropriate	Opinions of those indicating that recommendation results were not appropriate
It was easy to figure out what to look at next and less time was required.	I prefer to look at lecture materials and course handouts.

Table 2: Recommendation precision of bidirectional recommendation system.

Percentage of respondents who said that the recommendation results were suitable, or somewhat suitable
Recommendation precision = 61%

We surmised that perhaps the objective of learners using this approach is to thoroughly review using the material used in lectures and deepen their understanding of it, even if it required more time. Based on this, we hypothesized the necessary function to be support information used when reviewing. For example, this could refer to “areas of weakness” that the learner finds harder to understand than the rest of the text.

Based on this, we proposed a “collaborative learning recommendation system” using e-learning, and developed a system designed to improve learning efficacy by recommending “areas of weakness” (hereafter referred to as “non-favorite subject material”).

Finally, we surveyed the state-of-the-art about the recommendation technologies such as collaborative filtering and data mining, as follows. In [5], the design and implementation of a recommender system using social networks was described. In [6], a web content recommendation system based on the similarities is proposed. In [7], collaborative filtering based on C-SVM(Support Vector Machine) was proposed examined. In [8], data mining technologies, such as clustering and sequential pattern mining, for online collaborative learning data are studied. In [9], monitoring online tests, such as learner behavior and test quality, through data visualization are discussed. In [10], an automated learning and skills training system for a database programming environment is presented. In [11], a personalized active recommendation system called COALE is proposed and COALE gives proper awareness at proper timing for each learner to support dynamic course organization aimed at effective and efficient learning. In [12], a hybrid collaborative filtering technique is studied and it is shown to be efficient to make just-in-time recommendation.

In our research, we proposed a method for a collaborative learning recommendation system that mines the data of similar users sharing non-favorite subjects using historical logs and user attribute data. The method for mining non-favorite subject material proposed here is based on the assumption that the more times the content has been browsed, the less skilled the learner is in that subject. In addition, for new learners who do not possess a learning history, we proposed a method which uses attribute data to mine data for learners having similar preferences.

From the above survey results, we found that there are no existential research results which provide a solution on the recommendation technology using attribute data of learners having similar preferences for the new learners who do not possess a learning history. This is our research originality.

2 AIRS AND UTILIZATION STATUS

AIRS is an e-learning system that focuses specifically on review, and was developed starting from fiscal 2004 (see Figure 2).

Figure 2 displays the AIRS Japanese top page after a learner, who is going to review the database contents especially selection function, logs in AIRS. This page is comprised of the book marks (located at the upper side) of the contents available with AIRS, the contents menu (located at the left side) corresponding to the selected book mark, and the help messages for beginners (located at the right side). In figure 2, the contents menu displays database, data model, RDB, design methodology, and SQL. The learner selects the book mark such as selection and sorting before the learner can select the corresponding database contents menu. Then, the learner can proceed to review the database contents.

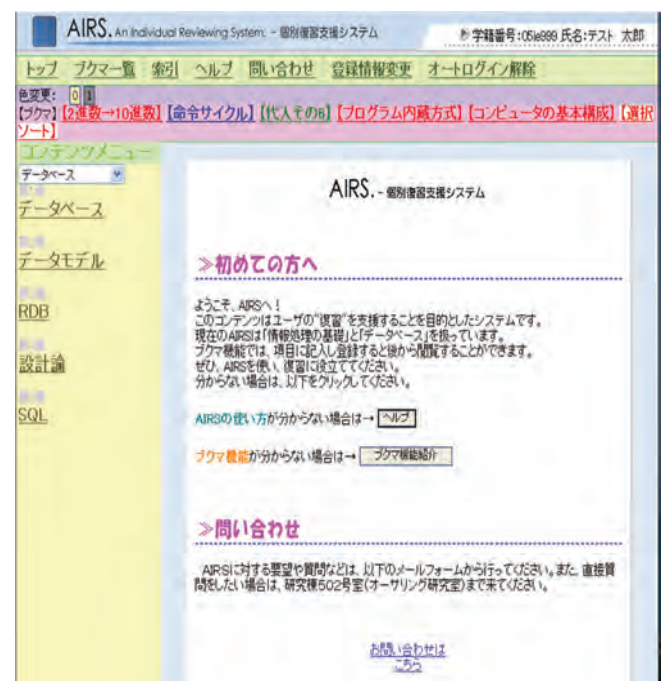


Figure 2: AIRS Japanese top page after login.

By focusing solely on review, the system reduces the possibility that the learner will rely on e-learning instead of sufficiently participating in lectures.

The system is also designed with the aim of improving learning efficacy through the synergistic effect of lecture-based learning and e-learning.

One feature of the system is that it is an e-learning system by learners, for learners. This reason is that system is designed to make learning easier, by having the developers attend lectures corresponding to the teaching content and develop content, to some extent, by anticipating sections that learners would have difficulty understanding (see Figure 3).

Additionally, the system is constructed so that each item being taught is expressed in three different ways (not yet fully implemented), and a function is provided by which learning is tailored to the individual learner, with the appropriate “form of expression” (refer to [13]) for that particular learner being automatically extracted.

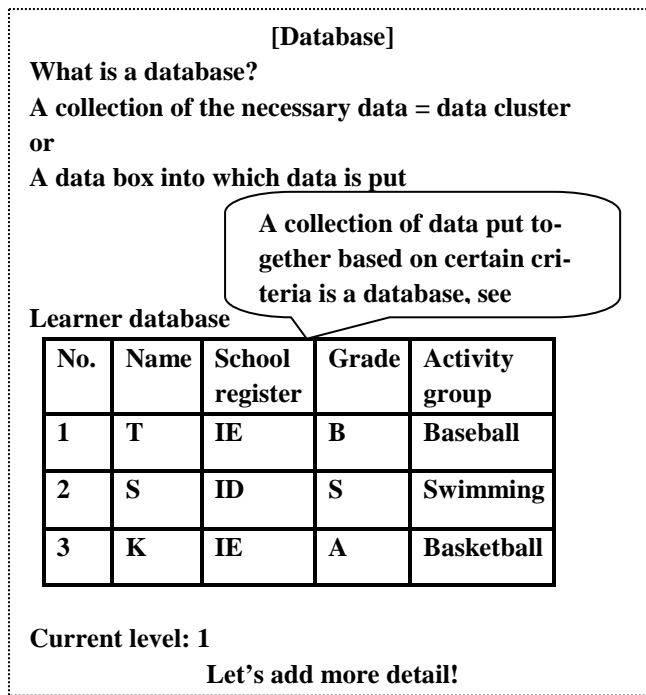


Figure 3: Example: contents of text.

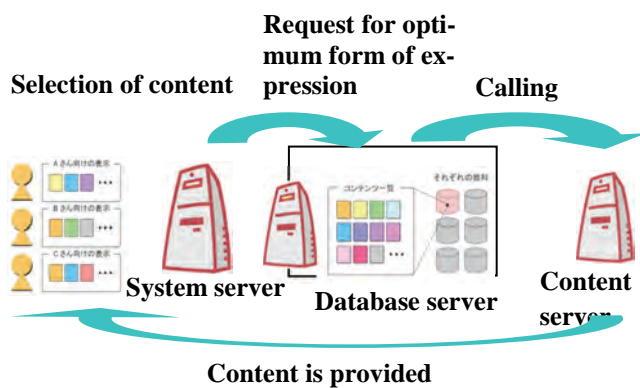


Figure 4: Configuration of AIRS.

The system comprises a database server that runs databases used by functions, such as the one mentioned previously, a content server that makes teaching content available, and a system server that runs AIRS (see Figure 4).

The results of a questionnaire survey, conducted this fiscal year, concerning utilization status are shown in Figure 5. As previously mentioned, the survey targeted 92 participants of the “Database System” lecture offered by the university.

3 BIDIRECTIONAL RECOMMENDATION SYSTEM

The aim of the system is to make it possible for learners to learn efficiently, without having to worry about selecting the content that was expanded through propagation of SCORM.

Moreover, based on the browsing history data of AIRS, it was found that an extremely large number of learners are sequentially browsing the course material in accordance with

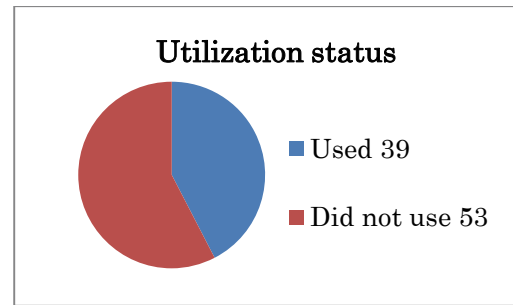


Figure 5: Usage achievements of AIRS.

the flow of the material displayed on the screen. This is not different from review using paper materials and suggests learning efficacy will decrease as the volume of material expands. In actuality, the number of AIRS nodes (not taking the “form of expression” into consideration) has grown to 210 for two subjects. To solve these problems, the bidirectional recommendation system was developed as a means for recommending course material that is strongly relevant to the material currently being read, and thus improving both review efficiency and speed.

An overview of the system is presented here, together with a detailed description of the questionnaire previously described.

3.1 Overview

The system overview is presented in Figure 1. We assume here that the learner is browsing the material for the course called “Basics of Assignment” under “Fundamentals of Information Processing”. At this point, the learner would naturally shift to the next steps, “The Basics of the ‘While’ Statement” and “The Basics of the ‘If’ Statement”.

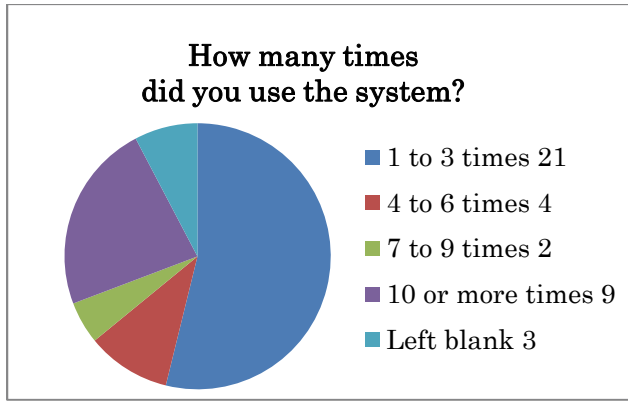
However, during the review process, it would not be unnatural for the learner to go back and re-read “Types of Variables”, which is part of the basic content. In other words, learners could review the material, if the system, instead of recommending only material to which many learners shift after reading “Basics of Assignment” at the same time, recommends material that many learners read before moving to “Basics of Assignment”. Looking back over material is a fundamental part of the review process, and we could expect an improvement in learning efficiency. This is why bidirectional recommendations are necessary, and is a feature of the bidirectional recommendation system.

3.2 Evaluation

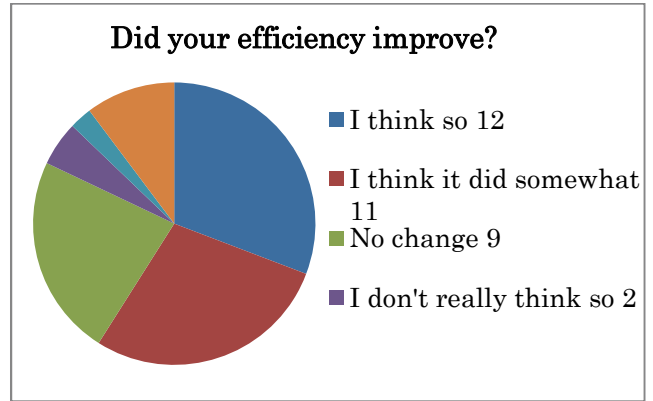
Users of the bidirectional recommendation system filled out questionnaires regarding the number of times they used the system, the recommendation results, learning efficiency, whether or not they would like to use the system in the future, operability and other questions. The results are shown in Figure 6.

The targeted users and the organizations conducting the survey are the same as those for the questionnaire survey previously described.

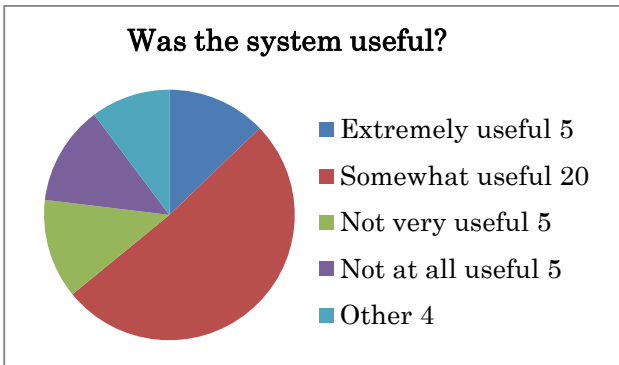
The results indicated a large number of learners used the system infrequently because they had problems logging in.



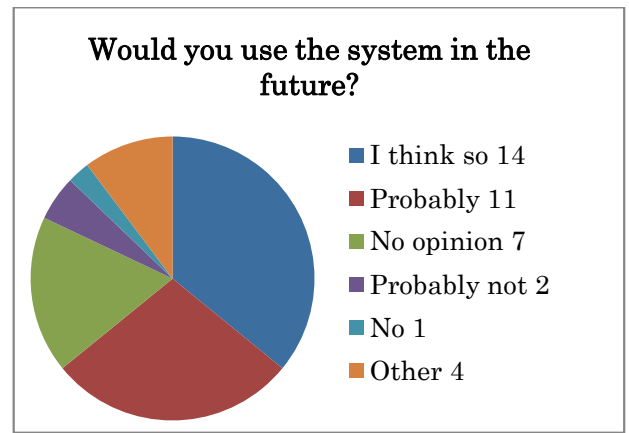
(a) What is the frequency of use?



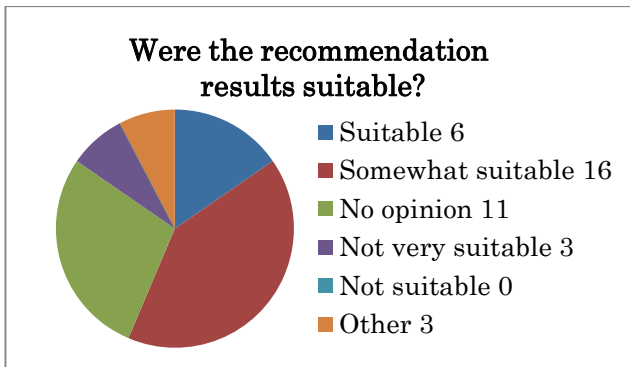
(e) Is it efficient?



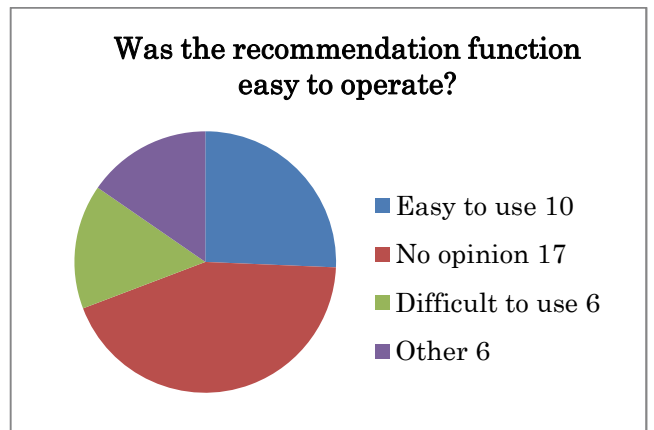
(b) Is it useful?



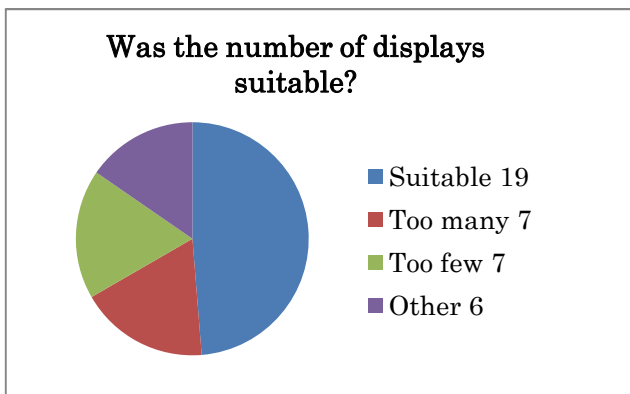
(f) Will you use it in the future?



(c) Is it suitable?



(g) Is it easy to operate?



(d) What is the number of displays?

Figure 6: User evaluation results of bidirectional recommendation system.

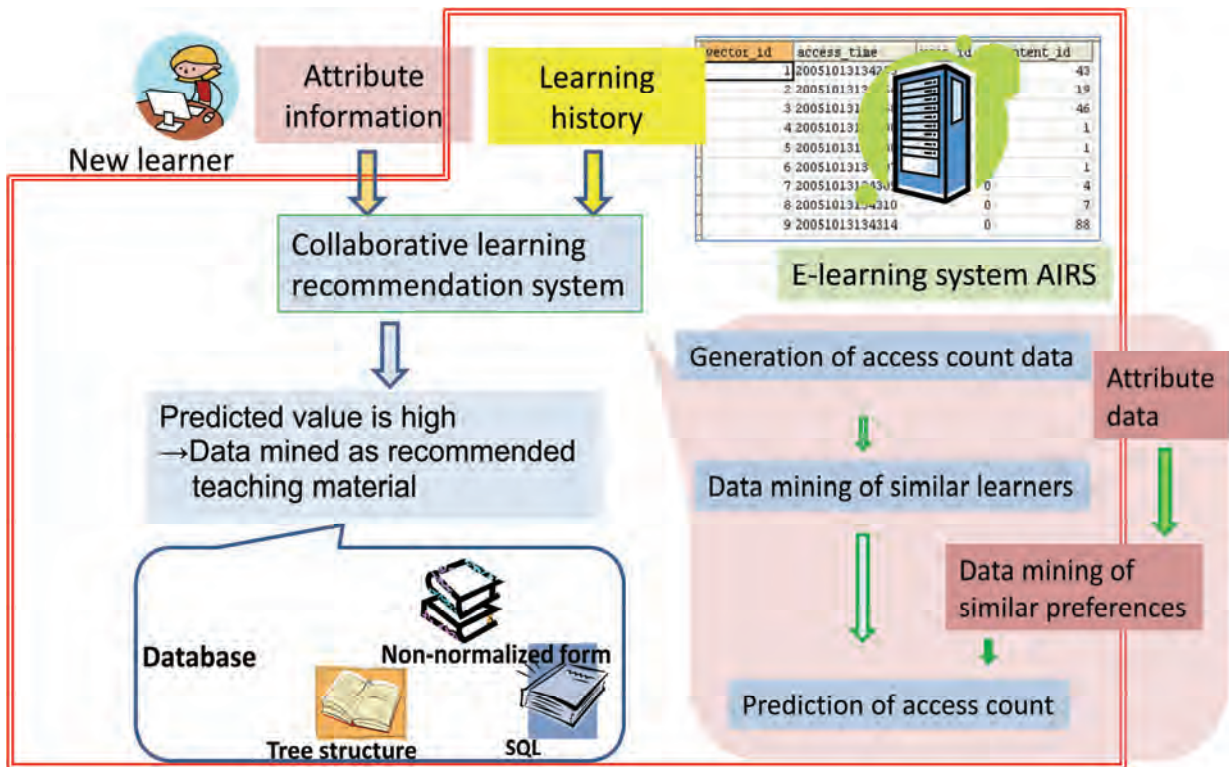


Figure 7: Overall system flow.

Even taking that into consideration, the majority of users obtained favorable results using the system for only one week, and the system can be expected to improve learning efficacy.

Firstly, 64% of the respondents said that the system was useful as shown in Figure 6 (b).

Secondly, 66% of the respondents said that the recommendation results were suitable as shown in Figure 6 (c).

Thirdly, 58% of the respondents said that efficiency improved as shown in Figure 6 (e).

Moreover, 64% of the respondents said that they would use the system in the future, including those who thought the contents were easy to understand and those who would use it if errors were corrected, as shown in Figure 6 (f).

However, 74% of the respondents said that the recommendation function is not easy to operate as shown in Figure 6 (g).

4 COLLABORATIVE LEARNING RECOMMENDATION SYSTEM

A characteristic feature of the proposed method is that it identifies learners exhibiting similar areas of weakness and recommends information for overcoming these weaknesses. Figure 7 shows the overall flow. This method functions by recommending teaching material via a collaborative learning recommendation system in which the learning history is used as the basic data. For new learners having no learning history, recommendations are made using the attribute information.

4.1 Learner Access Count Data

First, as a pre-process, the learning history is converted to generate data indicating the number of times each learner accessed the respective teaching material. Table 3 shows an excerpt of this data. The first line indicates the number of accesses by learner number 0 and shows that content1 was accessed 53 times, content4 was accessed 49 times and content7 was accessed 40 times. In addition to providing directly observable values, this data is thought to reflect such characteristics as the learner's interests and level of proficiency. In the case of a system such as AIRS that aims to support learning, the learners' objectives are to prepare for and review lectures, but because the use of such systems thus far has been limited to times prior to examinations, the learner's objective can be considered to be review. In other words, this data is thought to indicate areas in which the level of understanding is inadequate and weaknesses exist.

Table 3: Access count data.

user_id	con1	con2	con3	con4	con5	con6	con7
0	53	0	0	49	0	0	40
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	2	0	0	1	0	0	2
4	9	0	0	7	0	0	5
5	5	0	11	1	0	0	0
6	4	0	0	0	0	0	1
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	9	0	0	12	0	0	1
10	0	0	0	0	0	0	0

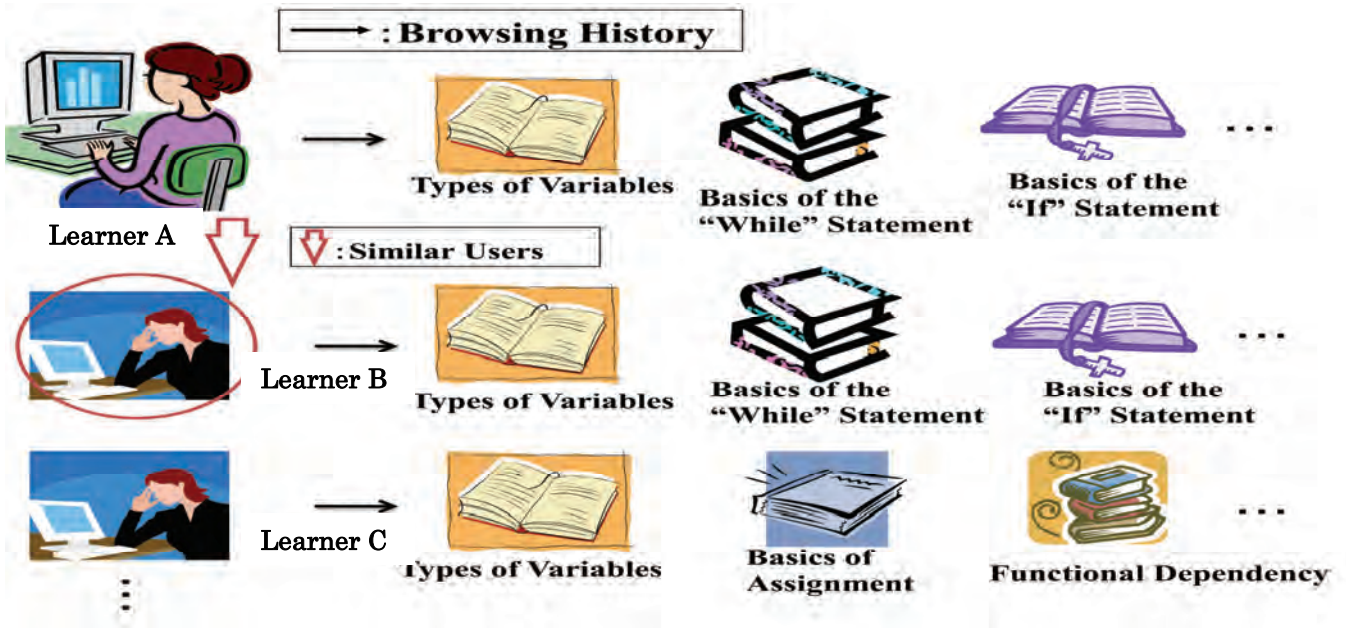


Figure 8: Mining of browsing history data to identify similar users

4.2 Data Mining of Similar Learners

The access count data described in Table 3 is compared for all learners to mine data for learners having similar weaknesses, or in other words, similar learners. Figure 8 illustrates the concept of similar learner data mining. Here, Learner A, as well as Learner B and Learner C, has a learning history. At this time, Learner A accessed the teaching materials of “Types of variables,” “Basics of the ‘while’ statement” and “Basics of the ‘if’ statement,” Learner B similarly accessed the same three materials and Learner C accessed the materials of “Types of variables,” “Basics of assignment” and “Functional dependence.” In this case, because the teaching materials accessed by Learner A and Learner B are similar, there is a high degree of resemblance between Learner A and Learner B, and therefore they are similar learners.

A correlation coefficient algorithm, implemented, for example, in collaborative filtering [14], is used in the actual mining of data for similar learners. The form of the computational equation applied in this research is shown in equation (1) below.

$$r_{ab} = \frac{\sum_{i=1}^T (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^T (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^T (b_i - \bar{b})^2}} \quad (1)$$

The above r_{ab} is the resemblance between Learner A and Learner B. This r is the first letter of the word resemblance. Values of r range from 1.0 to -1.0 , and values approaching 1.0 indicate greater resemblance, values approaching -1.0 indicate less resemblance, and the value 0 indicates that there is no relationship. The numerator indicates covariance, and the denominator indicates the product of the standard deviations.

The above a_i indicates the number of times Learner A has accessed the i^{th} teaching material and \bar{a} indicates the average number of accesses per teaching material. As well, the above b_i indicates the number of times Learner B has accessed the i^{th} teaching material and \bar{b} indicates the average number of accesses per teaching material. T indicates the total number of teaching materials. The learners found to have a high degree of resemblance according to this method are mined as similar learners.

4.3 Data Mining of Recommended Teaching Material

Here, assuming that the history of similar learners contains a history of overcoming weaknesses, teaching material that has been accessed many times by similar learners is mined as teaching material that helps to overcome these learners’ weaknesses. Figure 9 shows the method of mining recommended teaching material. In this example, similar learners accessed “Types of variables” three times, “Basics of the ‘while’ statement” nine times, “Basics of the ‘if’ statement” five times and “Basics of arrays” two times. If recommended teaching materials are mined from the historical data, “Basics of the ‘while’ statement” and “Basics of the ‘if’ statement,” which were the most frequently accessed, will be recommended to learners.

The actual computation for mining recommended teaching materials is performed by a method that calculates the predicted value used with collaborative filtering. The data for this calculation is based on the resemblance among similar learners and the learner vector of similar learners, and the value calculated is the predicted value of the access count. In this study, the predicted value is computed using equation (2) below.

$$P_{a,1} = \bar{a} + \frac{\sum_{U_i \in \text{User}} (C_{i,1} - \bar{C}_i) r_{ai}}{\sum_{U_i \in \text{User}} |r_{ai}|} \quad (2)$$

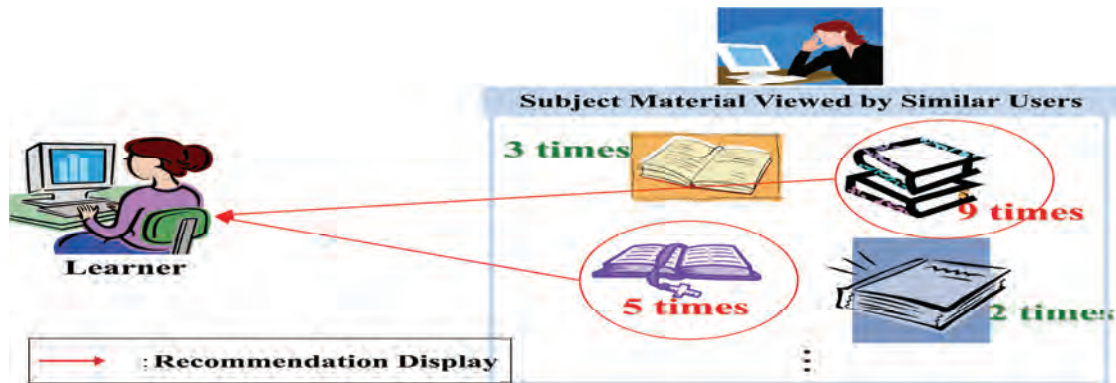


Figure 9: Data mining of non-favorite subject material.

Here, $P_{a,1}$ is the predicted value (predicted access count) necessary for Learner A to overcome a weakness in teaching material 1. For example, in the case where $P_{a,1}$ has a value of 3.4, Learner A is thought to be able to overcome his or her weakness in teaching material 1 by accessing that content approximately three times. Also, \bar{a} is the average number of times that a group of learners having some correlation with the Learner A (i.e., learners having a non-zero resemblance value) accessed the teaching material; in other words, the average of the average access count. "User" is a group of learners having some correlation with Learner A, and the i^{th} learner within this group is denoted as U_i . \bar{C}_i indicates the number of times that the i^{th} learner accessed teaching material 1, \bar{C}_i is the average access count and indicates the average number of times that the i^{th} learner reviewed any single teaching material, and r_{ai} indicates the resemblance between Learner A and the i^{th} learner. In this manner, teaching material for which the predicted access count is a large value is mined as recommended teaching material.

5 MINING DATA OF SIMILAR USERS

5.1 Attribute Data and Similar Users

Information not obtained by AIRS includes the learner attribute data, such as age, gender, hobbies and preferences.

In the present study, the strong subjects, non-favorite subjects, average learning time, hobbies and preferences, number of AIRS logins, usage time and other parameters of the user were additionally defined as learner attribute data. The purpose of acquiring this attribute data was to provide detailed recommendations even if the user was new to the system.

When browsing histories of similar users are mined (see Figure 8), new users are unable to find similar users because they have no learning history data, and recommendation accuracy drops sharply as a result. When all users have the same attribute data, it becomes possible to mine data for new users and similar users as well. The method for mining attribute data of similar users is shown in Figure 10.

The collaborative filtering method was used for mining the data of similar users, and mining of non-favorite subject material was done as described in section 4 (see Figure 9).

5.2 Effective Attribute Data Group

In the present study, mining all of the attribute data would not be useful in identifying similar users. Therefore, it was considered important to identify "attribute data groups" that were useful or effective, consisting of combinations of several attribute data elements.

A method proposed for identifying these attribute data groups is shown in Figure 11.

- (Step 1) First, one attribute data combination is created by an e-learning system administrator (such as AIRS administrator). For the time being, this is called the "first attribute group".
- (Step 2) Similarity between users is calculated using the equation (1) as mentioned before. Similar users are identified measuring the value of the similarity as described in section 4.2, referring to this first attribute group (see Figure 10). These are "similar users based on the first attribute group".
- (Step 3) The "similar users based on the first attribute group" (for example, Learner B and D in Figure 11) identified at step 2 are compared to the "similar users based on browsing histories" (for example, Learner B and D in Figure 11) identified using the method described in section 4 (see Figure 8). This means whether the name of the former similar user is equal to that of the latter similar user or not. And, the percentage of matches is calculated as the match rate.
- (Step 4) The learner who will serve as the reference is substituted for Learner B, and the match rate is calculated by repeating steps 2 and 3. In the same way, the match rates for subsequent learners (e.g., Learner C, Learner D) are determined until match rates have been determined for all of the users. The total of the match rates for all users is then divided by the number of users (n) to find the mean match rate, and that value is used as the "effective index of the first attribute group".

(Step 5) If all combinations of attribute data have been processed, then this step exits. Otherwise, another attribute data group is created at step 1 again, and the process through step 4 is repeated.

The effective index sequentially increased to find the “most effective attribute data group” in the course of repeating the above steps. But, the computing time complexity of this method is $O(n^2)$. As this is not the efficient algorithm, we are planning its improvement.

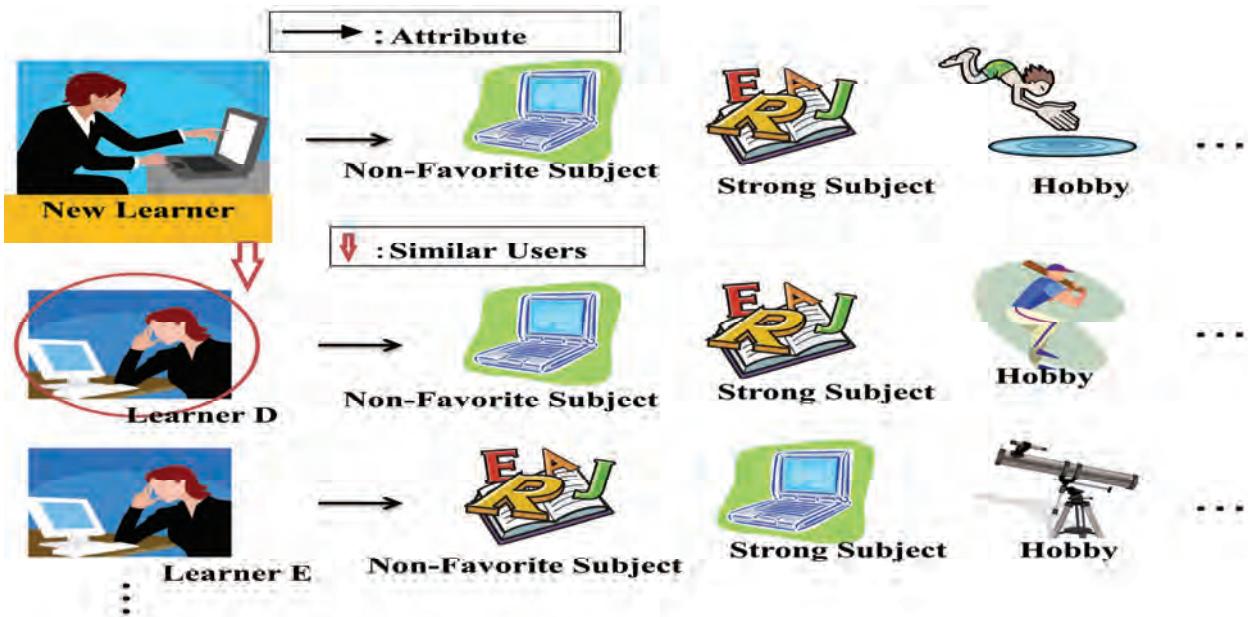


Figure 10: Method for mining attribute data of similar users.

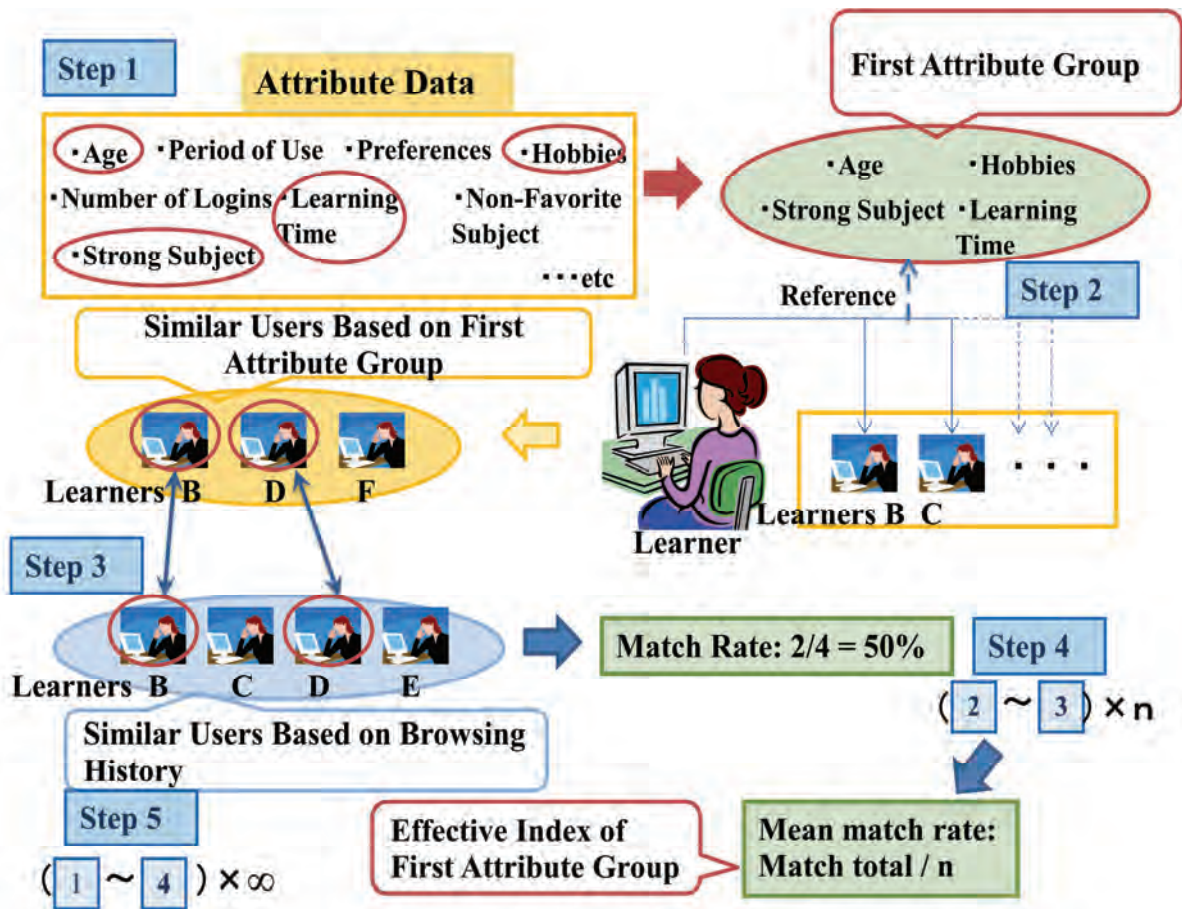


Figure 11: Method for mining useful attribute data.

6 CONCLUSION

In this paper, we proposed a method for a collaborative learning recommendation system that mines the data of similar users sharing non-favorite subjects using historical logs and user attribute data.

The method for mining non-favorite subject material proposed here is based on the assumption that the more times the content has been browsed, the less skilled the learner is in that subject.

For this reason, we currently plan to develop a collaborative learning recommendation system and implement it in the AIRS, and to verify the appropriateness of the recommendation results by measuring recommendation precision.

Then, the recommendation precision will be measured using the following data:

(1) Questionnaire results reflecting the subjective view of the student (user),

(2) Information relating to teaching material in which the learner is thought to be weak (as indicated by the course instructor),

(3) Comparison results of learning effectiveness between students who used AIRS with collaborative learning recommendations and students who used AIRS without these recommendations.

Finally, we consider our future work is as follows: we collect new attribute data, we ascertain the usefulness and effectiveness of the attribute data, and we evaluate the recommendation results for new users.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to the teachers, researchers, and students who kindly participated in the experiments on our collaborative learning recommendation system with AIRS.

REFERENCES

- [1] Ministry of Economy, Trade and Industry, "e-Learning White paper 2007/2008," Tokyo Denki University Press (2007).
- [2] R. Wisher, "Sharable Content Object Reference Model (SCORM) 2004 4th Edition Overview Version1.0," Advanced Distributed Learning, pp. 1-18 (31 March 2009).
- [3] Y. Wada, T. Matsuzawa, M. Yamaguchi, and S. Dohi, "Proposal and Verification of Bidirectional Recommendation System for Learning Web Digital Texts," Proceedings of the Second International Conference on the Applications of Digital Information and Web Technologies (ICADIWT2009), pp. 210-215 (August 2009).
- [4] M. Yamaguchi, Y. Takahashi, T. Matsuzawa, T. Takahashi, and S. Dohi, "Development and Evaluation of e-Learning System AIRS with Multi-Expression Contents," IEICE-ET2006-118, Vol. 106, No. 583, pp. 69-74 (2007).
- [5] H. Hotta and M. Hagiwara, "Design and Implementation of Social-network-based Recommender System," IPSJ Transactions on Databases, Vol. 2, No. 1, pp. 46-56 (2009).
- [6] A. Sasaki, T. Miyata, Y. Inazumi, A. Kobayashi, and Y. Sakai, "Web Content Recommendation System Based on Similarities among Contents Cluster of Social Bookmark," IPSJ Transactions on Databases, Vol. 48, No. SIG20, pp. 14-27 (2007).
- [7] K. Oku, S. Nakajima, J. Miyazaki, and S. Uemura, "Context-Aware SVM for Context-Dependent Information Recommendation," DBSJ Letters, Vol. 5, No. 1, pp. 5-8 (2006).
- [8] D. Perea, J. Kay, I. Koprinska, K. Yacef, and O.R. Zaiane, "Clustering and Sequential Pattern Mining of Online Collaborative Learning Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 6, pp. 759-772 (2009).
- [9] G. Costagliola, V. Fuccella, M. Giordano, and G. Polese, "Monitoring Online Tests through Data Visualization," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 6, pp. 773-784 (2009).
- [10] C.Pahl and C.Kenny, "Interactive Correction and Recommendation for Computer Language and Training," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 6, pp. 854-865 (2009).
- [11] N. Furugori, H. Sato, H. Ogata, Y. Ochi, and Y. Yano, "COALE: Collaborative and Adaptive Learning Environment," Proceedings of the Conference on Computer Support for Collaborative Learning, pp. 493-494 (2002).
- [12] T. Tang and G. McCalla, "Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment," International Journal on E-Learning, Vol. 4, Issue 1, pp. 105-129 (2005).
- [13] Y. Takahashi, T. Matsuzawa, M. Yamaguchi, S. Dohi, and Y. Wada, "Recommendation and Delivery of the Optimum Texts for e-Learners," IPSJ SIG Technical Reports 2007-CE-88(22), Vol. 207, No. 12, pp. 157-162 (2007).
- [14] N. Yamazaki, "Software Design," ISSN 0916-629, Vol. 268, Gijutsu-Hyohron Co., Ltd, Tokyo, Japan (2007).

(Received August 26, 2019)

(Revised December 17, 2010)



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 PhD degree in computer science from Shizuoka University of Japan 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 the ACM, the IEEE-CS and the IPSJ.



contents, and recommendation.

Yuuma Hamadume received the B.E. and the M.E. in information environment engineering from Tokyo Denki University of Japan in 2008 and 2010, respectively. He is currently working on a game software development at Konami Digital Entertainment Co., Ltd. His research interests include e-learning system, digital text



System), and improvement of students' motivation in introductory computer programming education.

Shinichi Dohi received the B.E. and the M.E. from Tokyo Denki University in 1982 and 1984, respectively. He is currently an Associate Professor in School of Information Environment, Tokyo Denki University. His research interests include AIRS(An Individual Reviewing



He received his PhD degree from Tokyo Denki University in 2004. His research interests include ubiquitous computing, human-interface system, multi-agent systems, and cooperative problem solving. He is a member of IPSJ, IEEE-CS, ACM.

Jun Sawamoto is currently a Professor of Faculty of Software and Information Science, Iwate Prefectural University, Japan. He received the B.E. and M.E. in mechanical engineering from Kyoto University in 1973 and 1975. He joined Mitsubishi Electric Corporation in 1975.