# 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
[**] Graduate School of Information Environment, Tokyo Denki University, 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.[1]

***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 substitution," 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 substitution."

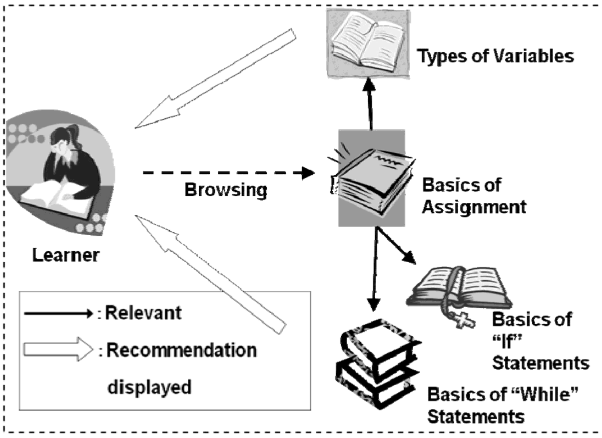



Figure 1: Bidirectional recommendation system.

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.

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. |

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

[1] 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).

materials, so we reexamined the functions requested by learners.

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.

## 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.

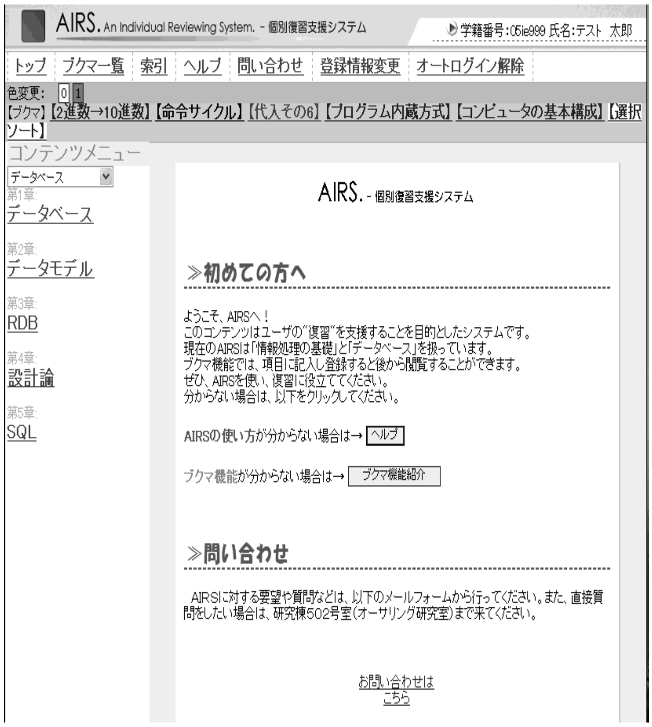



Figure 2: AIRS Japanese top page after login.

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.

**[Database]**

**What is a database?**

**A collection of the necessary data = data cluster**

**or**

**A data box into which data is put**

**A collection of data put together based on certain criteria is a database, see**

**Learner database**

| No. | Name | School register | Grade | Activity group |
|---|---|---|---|---|
| 1 | T | IE | B | Baseball |
| 2 | S | ID | S | Swimming |
| 3 | K | IE | A | Basketball |

**Current level: 1**

**Let's add more detail!**

Figure 3: Example: contents of text.

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 [11]) for that particular learner being automatically extracted.

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).

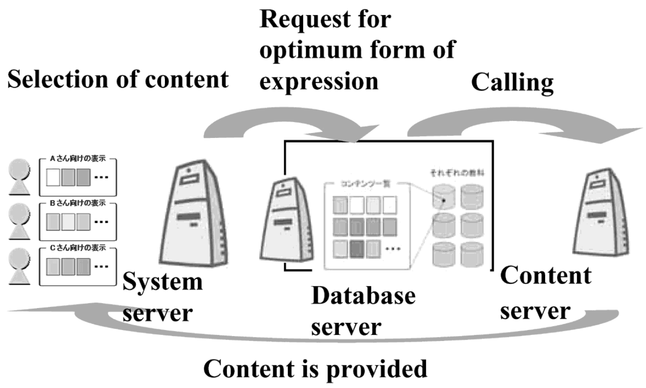



Figure 4: Configuration of AIRS.

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.

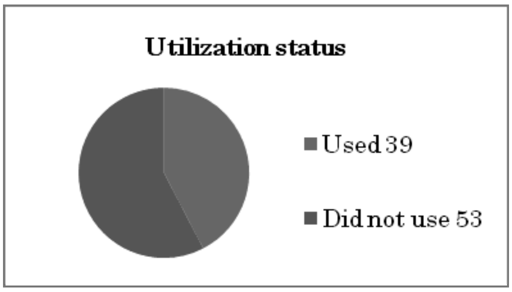



Figure 5: Usage achievements of AIRS.

## 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 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. 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.

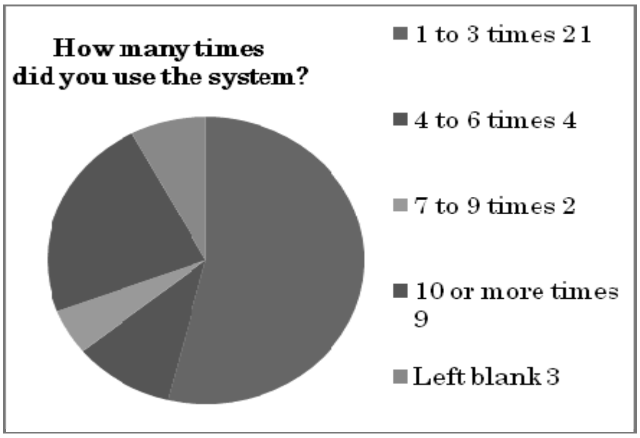
Firstly, 64% of the respondents said that the system was useful as shown in Figure 6 (b).

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

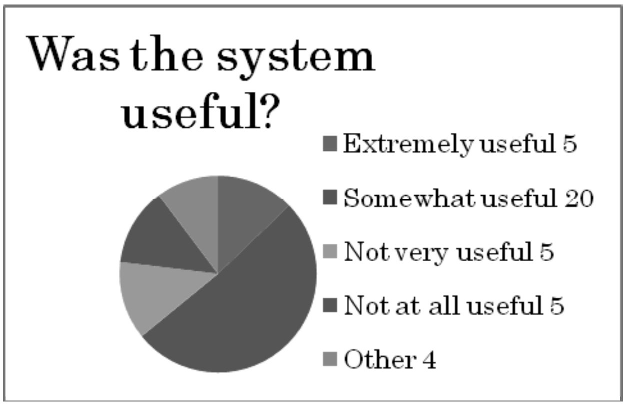
Thirdly, 58% of the respondents said that efficiency improved as shown in Figure 6 (e).

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).

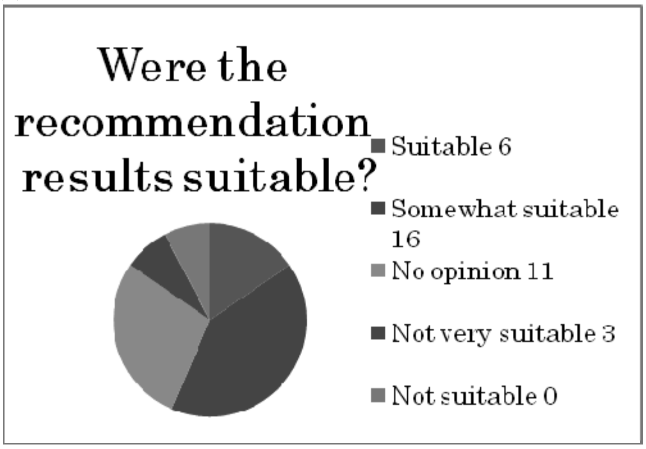
However, 74% of the respondents said that the recommendation function is not easy to operate as shown in Figure 6 (g).



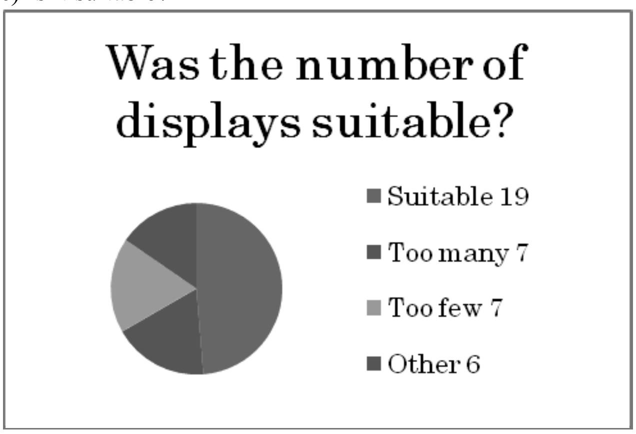

(a) What is the frequency of use?



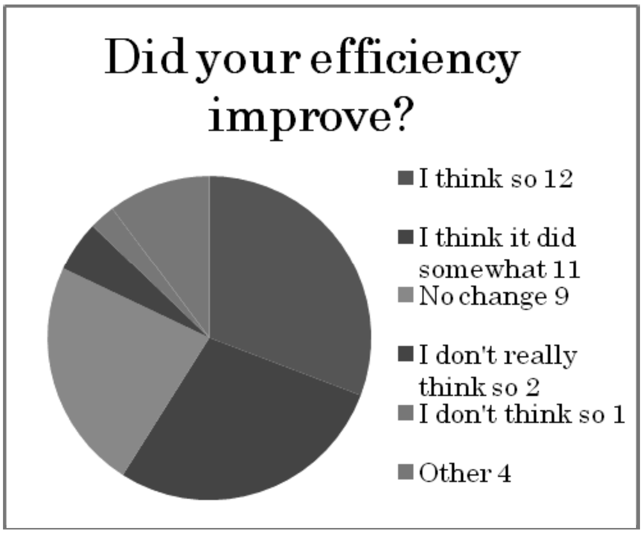

(b) Is it useful?



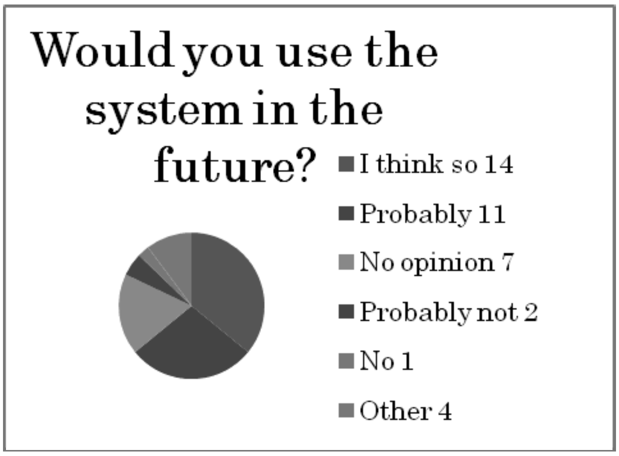

(c) Is it suitable?



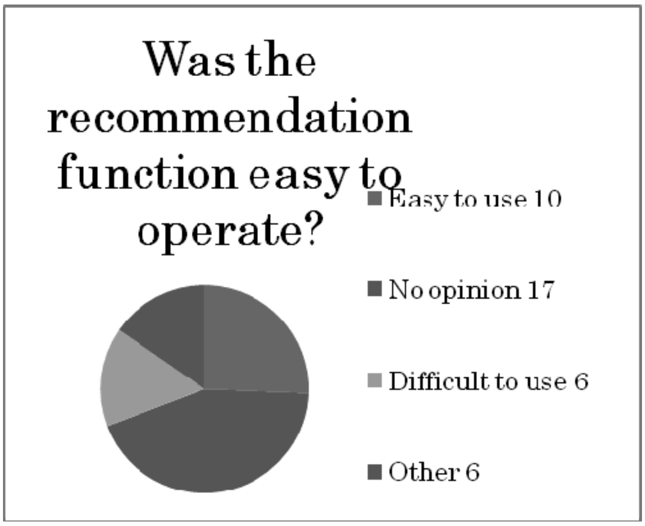

(d) What is the number of displays?




(e) Is it efficient?




(f) Will you use it in the future?




(g) Is it easy to operate?

Figure 6: User evaluation results of bidirectional recommendation system.

## 4 COLLABORATIVE LEARNING RECOMMENDATION SYSTEM

For the learning history data accumulated in AIRS, we focused on the number of times that the learners had browsed the various subject materials, suspecting that frequent browsing indicated that the material had not been sufficiently understood from the lecture, and that there was a strong possibility of this being an area of weakness for the learner.

However, if subject material that the learner had browsed numerous times was recommended as “non-favorite subject material”, without any additional input or modification, there was a strong possibility that the learner was already aware of this material, having browsed it multiple times, and that this would not improve learning efficacy. There were also concerns that the recommendations would be concentrated too heavily on the same subject material, creating a convergence in the recommendation results.

To resolve this, we focused on collaborative filtering technology [12] in our research. Using this technology, it was possible to identify similar users with similar browsing histories from among the learning history data (see Figure 7).

It was thought that by recommending frequently browsed material from the learning history data of similar users, it would be possible to provide “non-favorite subject material” that the learner did not yet know (see Figure 8).

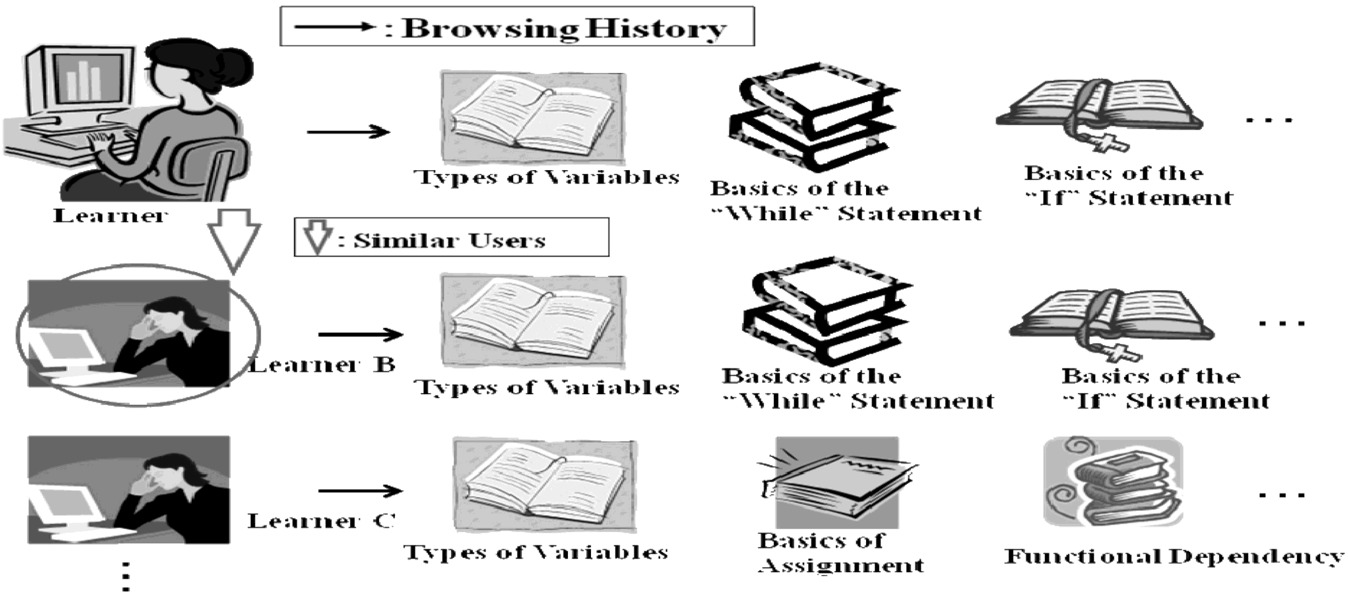



Figure 7: Mining of browsing history data to identify similar users

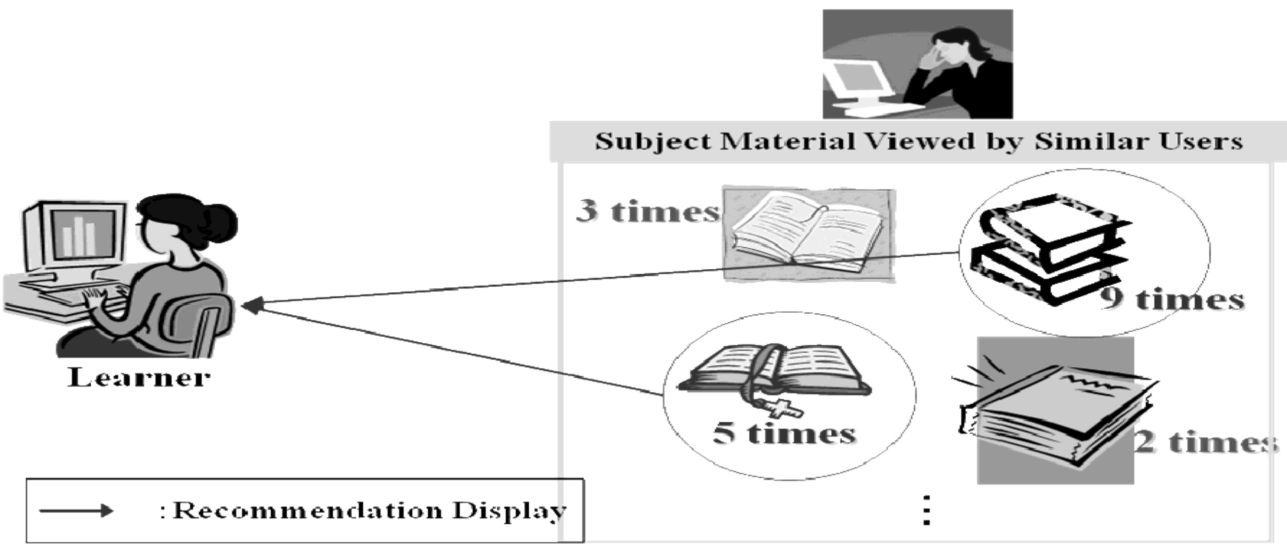



Figure 8: Data mining of non-favorite subject material.

# 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 7), 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 9.

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 8).

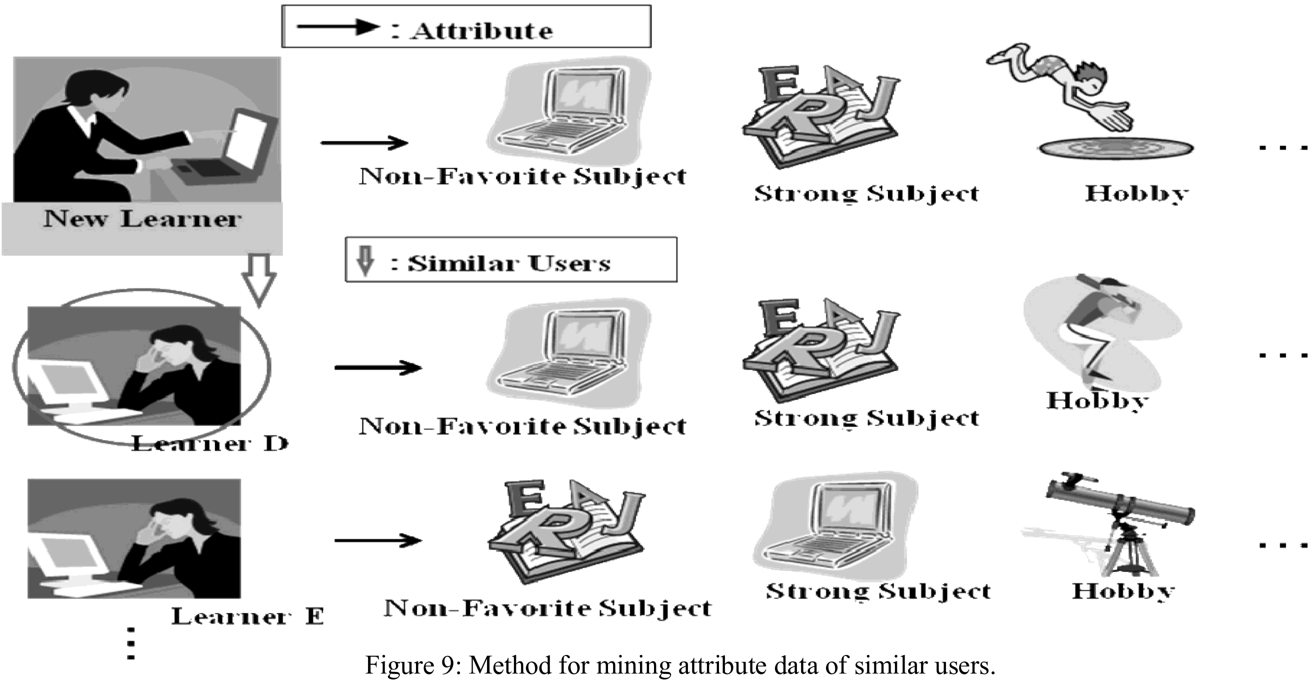



Figure 9: Method for mining attribute data of similar users.

## 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 10.

(Step 1) First, one attribute data combination is created. For the time being, this is called the “first attribute group”.

(Step 2) Similar users are identified, referring to this first attribute group (see Figure 9). These are “similar users based on the first attribute group”.

(Step 3)The “similar users based on the first attribute group” identified at step 2 are compared to the “similar users based on browsing histories” identified using the method described in section 4 (see Figure 7), 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) An attribute data group is created at step 1 again, and the process through step 4 is repeated.

This process is repeated and the effective index sequentially increased to find the “most effective attribute data group”.

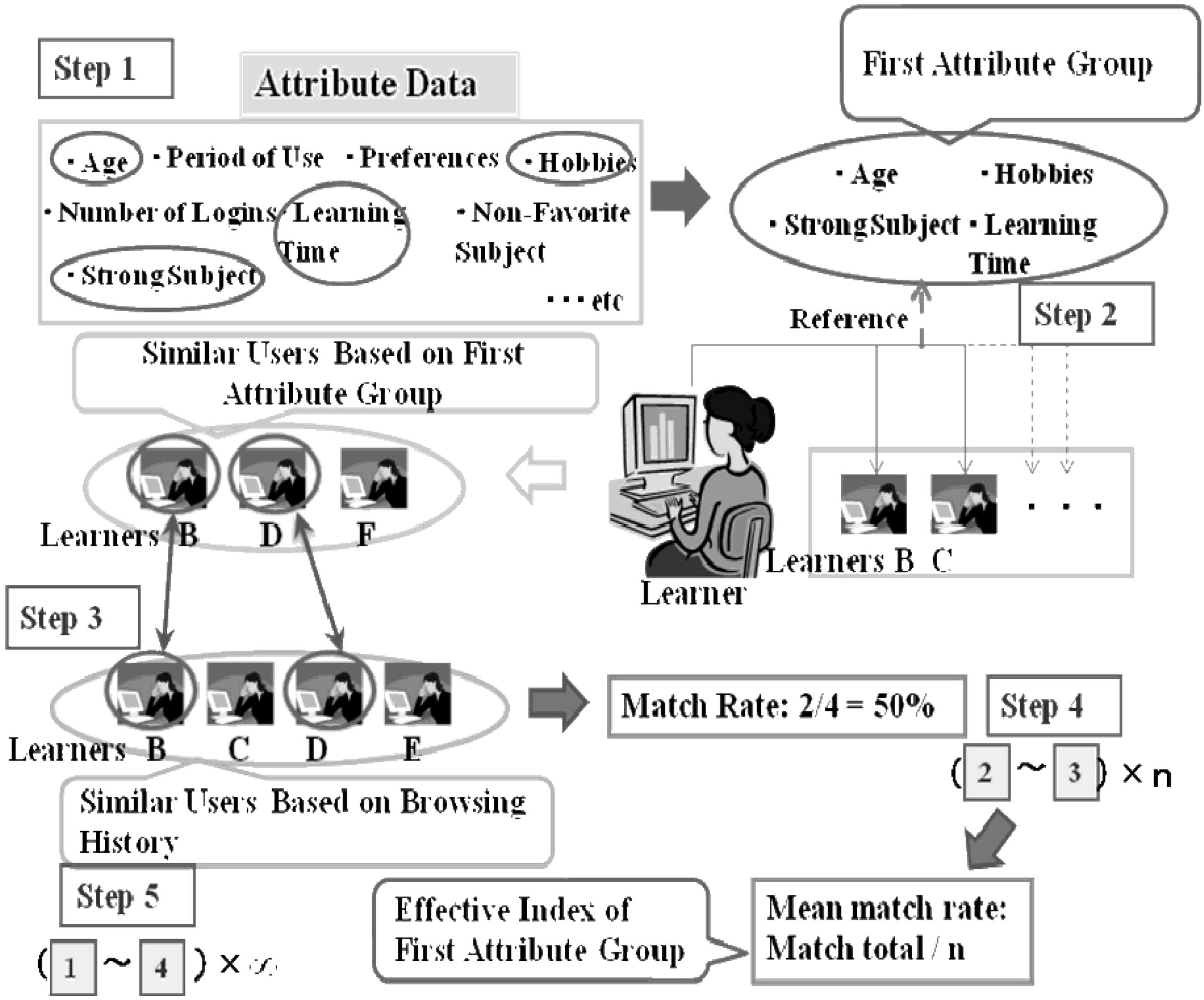



Figure 10: Method for mining useful attribute data.

## 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.Fuccela, 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] 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).

[12] N.Yamazaki, Software Design ISSN 0916-629, Vol... 268 , Gijutsu-Hyohron Co., Ltd, Tokyo Japan(2007-8).