Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 33 (2007) 6–22
Expert Systems with Applications
www.elsevier.com/locate/eswa

Mining learner profile utilizing association rule for web-based learning diagnosis

Chih-Ming Chen *, Ying-Ling Hsieh, Shih-Hsun Hsu

Graduate Institute of Library, Information and Archival Studies, National Chengchi University, No. 64, Sec. 2, Zhinan Rd., Taipei 116, R.O.C.

Abstract

With the rapid growth of computer and Internet technologies, e-learning has become a major trend in the computer assisted teaching and learning fields. Most past researches for web-based learning focused on the issues of adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of student's solutions. These systems commonly neglect to consider whether learner can understand the learning courseware and generate misconception or not. To neglect learner's learning misconception will lead to obviously reducing learning performance, thus generating learning difficult. In order to discover common learning misconceptions of learners, this study employs the association rule to mine the learner profile for diagnosing learners' common learning misconceptions during learning processes. In this paper, the association rules that occurring misconception A implies occurring misconception B can be discovered utilizing the proposed association rule learning diagnosis approach. Meanwhile, this study applies the discovered association rules of the common learning misconceptions to tune courseware structure through modifying the difficulty parameters of courseware in the courseware database so that learning pathway is appropriately tuned. Besides, this paper also presents a remedy learning approach based on the discovered common learning misconceptions to promote learning performance. Experiment results indicate that applying the proposed learning diagnosis approach can correctly discover learners' common learning misconceptions according to learner profile and help learners to learn more effectively.

Keywords: Web-based learning; Learning misconception diagnosis; Association rule mining; Learner profile

1. Instruction

As numerous web-based tutoring systems were successfully developed, a great quantity of hypermedia in courseware has created cognitive overload and disorientation problems (Berghel, 1997; Borchers, Herlockerm, Konstanand, & Riedl, 1998), such that learners are unable to learn very efficiently. To aid more efficient learning, many powerful personalized/adaptive guidance mechanisms, such as adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of student's solutions, have been proposed in the past researches (Brusilovsky, 1999; Papanikolaou & Grigoriadou, 2002; Tang & Mccalla, 2003; Weber & Specht, 1997). However, although many web-based learning techniques have been proposed to assist web-based learning, few researches have attempted to diagnose students' learning problems for the developed web-based tutoring systems. The learning diagnosis aims to identify learners' misconception and help them to promote the learning performance during learning processes. To help identify general misconceptions that learners might be having in a particular subject is critical and valuable to both teachers and learners in a web-based learning environment. Generally, the discovered learners' misconceptions from the learning behavior can be served as important feedback to the both learners and teachers. In the meanwhile, the web-based tutoring systems can also apply them to perform remedy learning or revise tutoring strategies.

To develop a novel learning diagnosis approach, several studies that have paid attention to the issue of learning

* Corresponding author. Tel.: +866 3 8227106x1519; fax: +866 3 8237408.
E-mail address: cmchen@mail.nhlue.edu.tw (C.-M. Chen).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.04.025

diagnosis for the developed web-based learning systems are summarized and discussed herein. Chang, Liu, and Chen (1998) proposed a testing-based diagnosis system using the student's answers in the test problems to discriminate the student's misconceptions on the subject domain of DC electric circuits, thus helping teachers to tune teaching strategy. Cheng's study (Cheng, Lin, Chen, & Heh, 2005) focused on creating the concept hierarchy by embedding important concepts in a test, then analyzing the results with a hierarchical coding scheme for learning diagnosis. Cheng's study emphasized that the teacher will be able to adjust the teaching and to supply more useful learning materials as necessary by gaining insight into the students' understanding and possible misconceptions. Hwang's study (Hwang, Hsiao, & Tseng, 2003) proposed a computer-assisted approach for teachers to define and analyze concept effect relationships, thus helping them to diagnose student's learning problems. Lo's study (Lo, Wang, & Yeh, 2004) developed an adaptive hierarchical concept level courseware for English prepositions. Based on the formative evaluation results from the proposed test levels, the system diagnoses learner's error types in English prepositions learning and identifies the reasons behind their misconception. Huang et al. (2004) proposed an intelligent learning diagnosis system based on log files that records learners' past online learning behavior to support a web-based thematic learning model for expanding learners' knowledge while surfing in the theme-based learning cyberspace. Jong, Lin, Wu, and Chan (2004) proposed a remedial-instruction decisive (RID) path algorithm based on a conceptual graph to diagnose and analyze student's missing concepts. Their study shows that participants who adopt the diagnostic and remedial learning strategy have better academic performance.

Based on the survey of learning diagnosis strategies mentioned above, this study presents a novel association rule based learning diagnosis approach to support the personalized e-learning system for learning performance promotion. The personalized e-learning system (PELS) based on Item Response Theory (Baker, 1992), which considers both courseware difficulty and learner ability to provide personalized learning paths for learners, was presented in our previous study (Chen, Lee, & Chen, 2005; Chen, Liu, & Chang, 2006). However, this system lacks an intelligent mechanism to identify particular learning misconceptions for learners. Therefore, this study proposes an association rule learning diagnosis approach to analyze learner profile in order to assist PELS to explore learners' learning misconceptions. The proposed system can provide suitable remedy learning courseware to learners to perform enhanced learning according to the discovered learner's misconceptions. Moreover, this study also proposes a mechanism of courseware structure modification to slightly tune the difficulty parameters of courseware according to the discovered learning misconceptions, thus modifying the courseware recommendation sequence for the provided personalized e-learning services. Experimental results show that the personalized e-learning system with the proposed learning diagnosis mechanism can help teachers and learners to identify learning misconceptions based on the proposed association rule learning diagnosis approach, helping learners to learn more effectively in a web-based environment.

2. System architecture

In order to discover the common learning misconceptions, this study proposes a learning diagnosis and remedy learning agent embedded in the proposed PELS, which can perform remedy learning based on learner's common learning misconceptions and modify the difficulty parameters of courseware for courseware structure modification. This section is organized as follows: first an overview of system architecture is presented in Section 2.1. Next, the learning procedure on PELS is explained in Section 2.2. Section 2.3 then describes the system components in detail.

2.1. System architecture

The PELS based on the Item Response Theory, which includes an off-line courseware modeling process, four intelligent agents and four databases, is presented in our previous study (Chen et al., 2005, 2006). The four intelligent agents are the learning interface agent, feedback agent, courseware recommendation agent and courseware management agent, respectively. These four databases include the user account database, user profile database, courseware database and teacher account database. The learner interface agent aims at providing a flexible learning interface for learners to interact with the feedback agent and the courseware recommendation agent. The feedback agent aims at collecting learner explicit feedback information from the learning interface agent and storing it in the user profile database for personalized curriculum sequencing operations. The courseware recommendation agent is in charge of recommending a personalized learning pathway to learner according to learner feedback response and concept relation degrees of courseware (Chen et al., 2005, 2006). Finally, the courseware management agent with authorized account management mechanism provides a responsive courseware management interface, aiding teachers to create new course units, upload courseware to the courseware database and delete or modify courseware from the courseware database.

However, the PELS mainly focuses on performing adaptive learning based on the difficulty parameters of courseware and learner ability for individual learner, it lacks the learning diagnosis and remedy learning mechanisms to support affectively learning. In this paper, the features of the PELS system are extended to include the test agent, testing item database, learning diagnosis and remedy learning agent in order to perform learner's misconception diagnosis for promoting learners' learning performance. In the extended PELS system, the test agent is used to

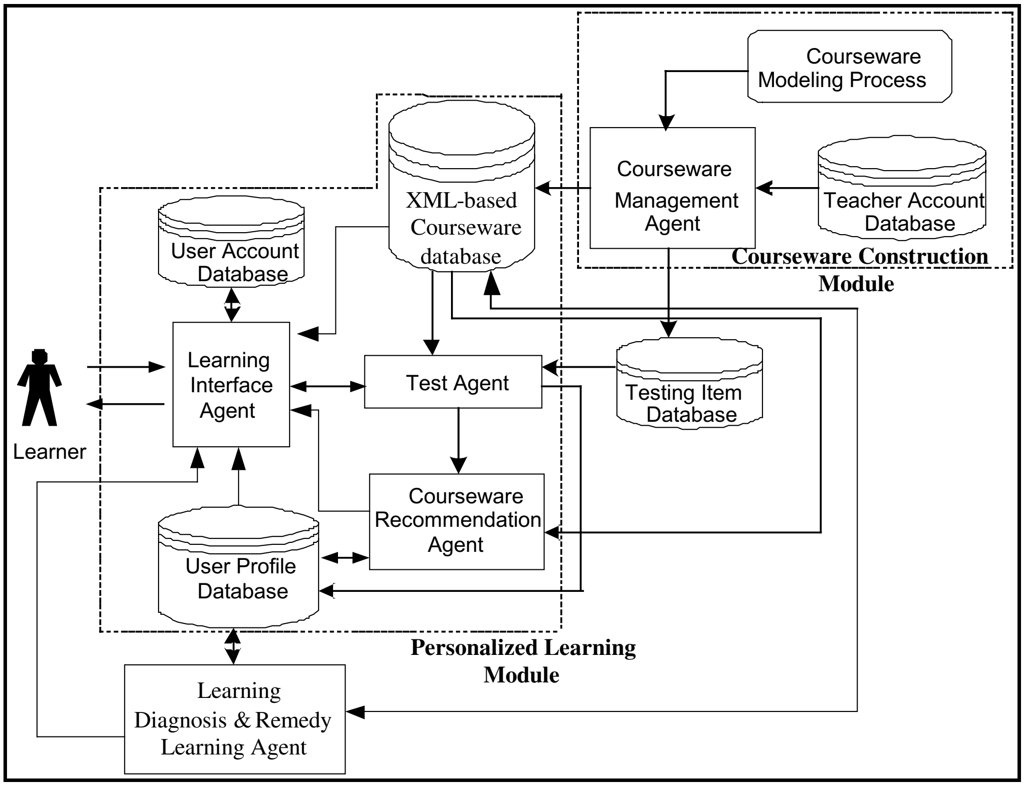


Fig. 1. The system architecture.

replace the feedback agent for collecting learner explicit feedback information and storing it in the user profile database for learning misconception diagnosis. In this study, to collect learner explicit feedback information, the test agent will ask learner to reply one randomly selected testing question related to the learned courseware from the testing item database. The reply is then sent to the courseware recommendation agent and use to determine learners' new abilities, and suggest appropriate course materials to learners. The system architecture is shown in Fig. 1.

2.2. Learning procedure on PELS

To explain how to gather the useful learner profile for the learning misconception diagnosis, this section describes briefly the learning procedure on PELS. Fig. 2 shows the entire layout of the learning interface. As a learner logins this system, he/she can choose a course unit that he/she feels interested for learning. In the left frame, system shows the course categories, course units and the list of all courseware in the courseware database using a hierarchical tree topology structure. Currently, courses created by teachers using the course management interface, can be categorized as titles of "Mathematics", "Physics", etc. Moreover, a course can be further divided into several course units by analyzing teaching content. Furthermore, a course unit involves many relevant course materials that convey similar concepts, but such course materials are associated with different levels of difficulty. Additionally, course material organized on Web pages with flash animation and synchronous voice comments is the course element in the proposed system. While a learner clicks a courseware for learning, the content of selected courseware will be exhibited in the upper-right window. Besides, the feedback interface is arranged in the bottom-right window. The proposed system can get learner's feedback response from the interface of test agent through learner replies one randomly selected testing question related to the conveyed learning content.

The answers of testing questions help system to get the learner's comprehension percentage for recommending appropriate courseware to the learner as well as collect the testing questions with wrong answer for the common leaning misconception diagnosis. System passes the feedback response to the courseware recommendation agent to infer the learner's ability using the Item Response Theory (Baker, 1992) detailed in our previous study (Chen et al., 2005, 2006). After a learner presses the button of "submit", this system will reveal a list of the recommended courseware based on his current ability. Fig. 3 shows an example of courseware recommendation based on learner ability after learner gives corresponding feedback response, and the recommended courseware ranked by the order of their information values. The title indicates the subject of the courseware; the recommendation denotes the information value of the recommended courseware; and the description gives a brief description for the corresponding courseware. The length of bar line in the column of recommendation indicates the information value of the

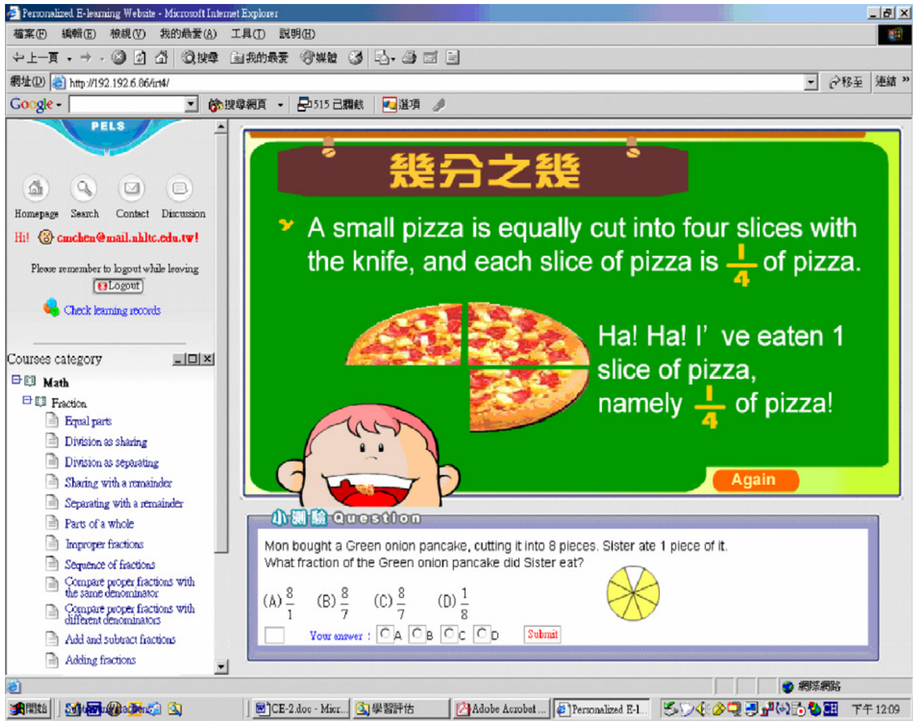


Fig. 2. The learning interface for learners.

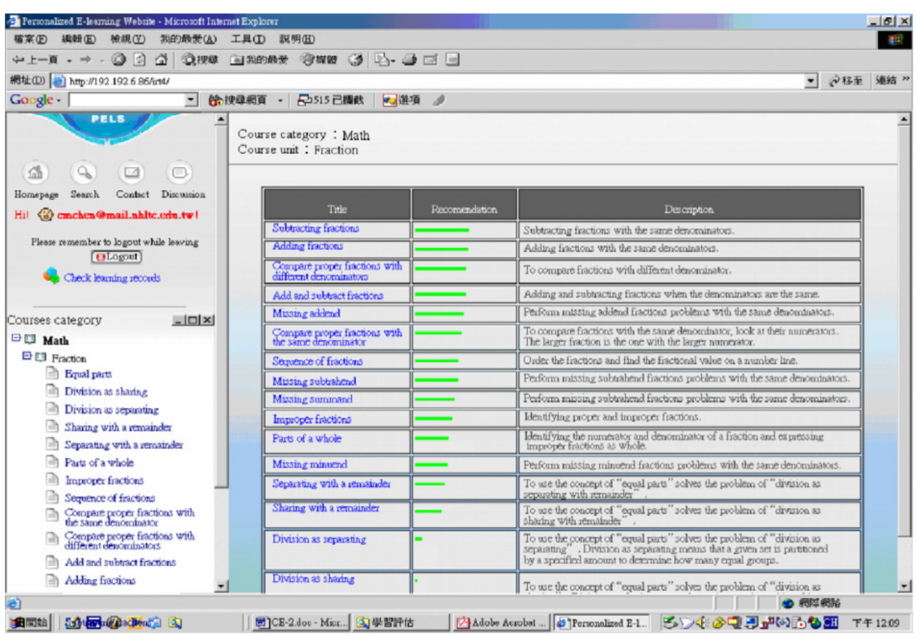


Fig. 3. An example of courseware recommendation ranked by the order of information values.

corresponding courseware. The longer bar line implies a more suitable courseware for learner. On the contrary, the shorter bar line implies an unsuitable courseware for learner. After the learner selects the next courseware according to the suggestion of the courseware recommendation agent for further learning, the learner can continue to learn the selected courseware. The PELS will continue to run the learning cycle until the evaluated learner ability satisfies the stop criterion. Next, the learning procedure will enter the final posttest stage to perform a summative assessment through replying 17 randomly selected testing questions. The results of final posttest will be used to verify the rule patterns of the proposed learning misconception diagnosis approach. Fig. 4 reveals the user interface for the final posttest. Of course, each learner can also browse the individual learning records in the learned course unit,

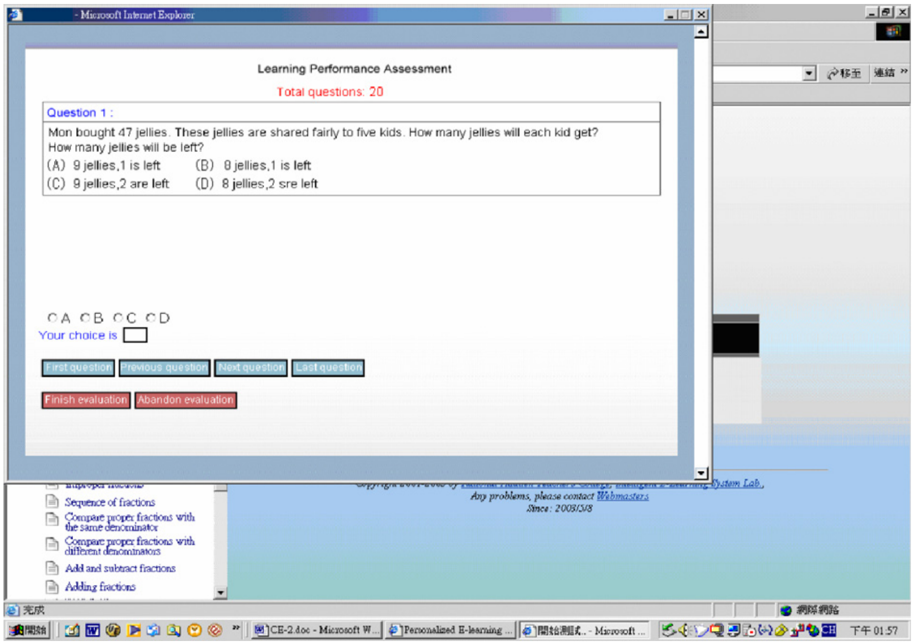
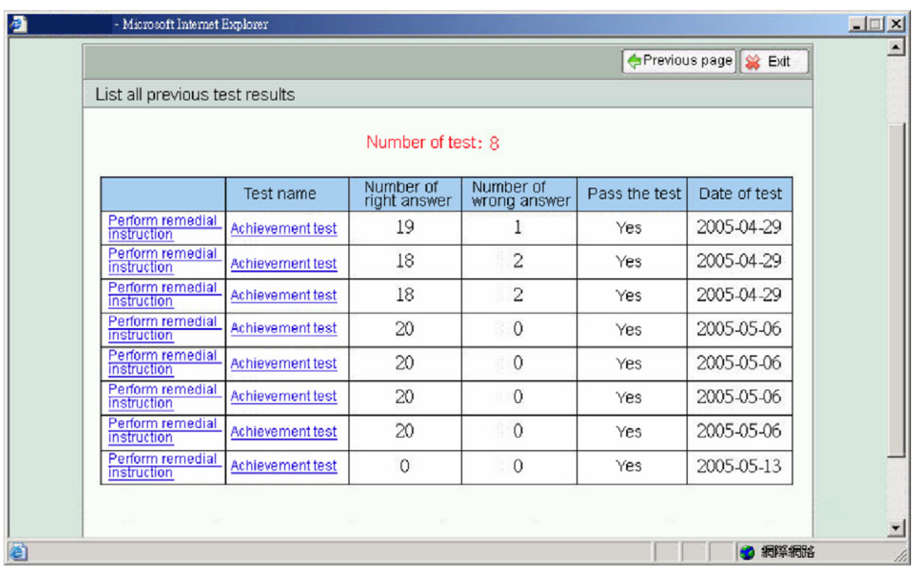


Fig. 4. The user interface of final testing after learning.



	Test name	Number of right answer	Number of wrong answer	Pass the test	Date of test
Perform remedial instruction	Achievement test	19	1	Yes	2005-04-29
Perform remedial instruction	Achievement test	18	2	Yes	2005-04-29
Perform remedial instruction	Achievement test	18	2	Yes	2005-04-29
Perform remedial instruction	Achievement test	20	0	Yes	2005-05-06
Perform remedial instruction	Achievement test	20	0	Yes	2005-05-06
Perform remedial instruction	Achievement test	20	0	Yes	2005-05-06
Perform remedial instruction	Achievement test	20	0	Yes	2005-05-06
Perform remedial instruction	Achievement test	0	0	Yes	2005-05-13

Fig. 5. The user interface of checking learning records of individual learner.

and check whether performing the remedy learning is needed or not. Fig. 5 shows the user interface for checking learning records of individual learner. Additionally, Fig. 6 reveals the user interface for performing the remedy learning and testing. The left window in Fig. 6 exhibits the learning diagnosis results.

2.3. System components

2.3.1. Courseware database

In this study, the courseware is provided in web-based environment. In order to simultaneously satisfy both the normal and remedy learning requirements, the courseware database shown in Fig. 1 is distinguished into two major parts, including the standard courseware database and remedy courseware database, respectively. Fig. 7 illustrates the courses database architecture. The details are described as follows.

(1) *Standard courseware:* In the PELS, the standard courseware database contains course materials for the normal learning process.
(2) *Remedy courseware:* The remedy courseware database contains course materials with easier difficulty level than the course materials in the standard courseware database for the remedy learning process. After

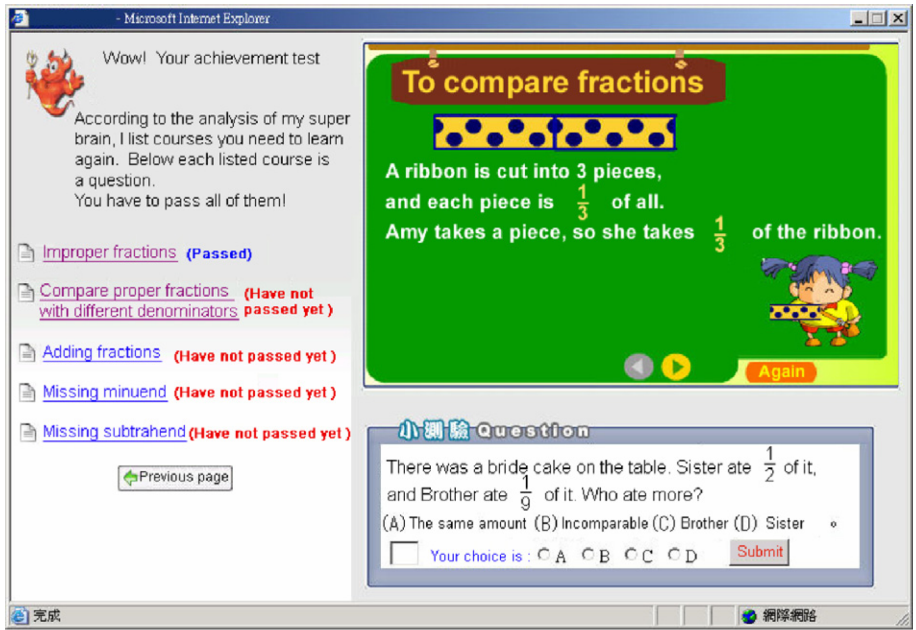


Fig. 6. The user interface of individual learner for performing remedy learning and testing.

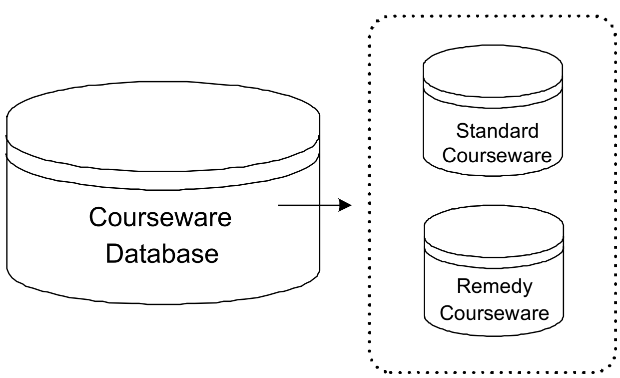


Fig. 7. The courseware database architecture.

a learner finishes the learning process, the system will provide the remedial course materials according to the discovered learner's learning misconceptions. The remedial course materials convey similar learning concepts with the discovered learner's learning misconceptions, but they contain different learning content.

2.3.2. Test agent

To diagnose learners' common learning misconceptions during learning processes; therefore, this study proposes a courseware construction process to build the learning courseware and their corresponding testing items. Fig. 8 shows the detailed procedures to explain how to construct the courseware and their corresponding testing items. Except for courseware construction, the courseware management agent also provides a friendly interface to interact with teachers for testing item construction. After teachers log in this system to add testing items, the added testing items and their corresponding courseware will be automatically built the connected relationships to each other. In other words, to perform the testing item with corresponding courseware can evaluate whether learner can understand the learning courseware or not. Moreover, the test agent can communicate with both the testing item database and courseware database shown in Fig. 8. During the learning process, the learner learns the recommended courseware, and then the test agent immediately chooses the corresponding testing item from the testing item database to learner to evaluate if learner can understand the recommended courseware. The tested results are recorded in the user profile database for diagnosing learner's learning misconceptions. In this study, the "wrong answer" will be served as "misconception" for the proposed association rule learning diagnosis approach.

2.3.3. Learning diagnosis and remedy learning agent

As the mentioned above, the user profile database contains learner's feedback responses for the testing items after learner learnt the recommended courseware. In order to provide the remedial instructions for learners to revise learning misconceptions, the learning diagnosis and remedial learning agent identifies particular leaning misconceptions for learners utilizing the proposed association rule based learning diagnosis approach. Fig. 9 illustrates the detailed operation procedures for the proposed learning diagnosis and remedy learning mechanisms. The following subsections will describe each operation procedure in detail.

2.3.3.1. Data preparation. Data preprocessing aims at cleaning and transferring the data in the user profile

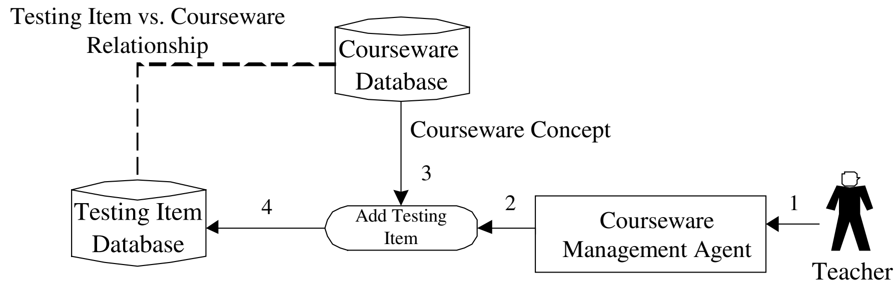


Fig. 8. The courseware construction process for learning courseware and their corresponding testing items.

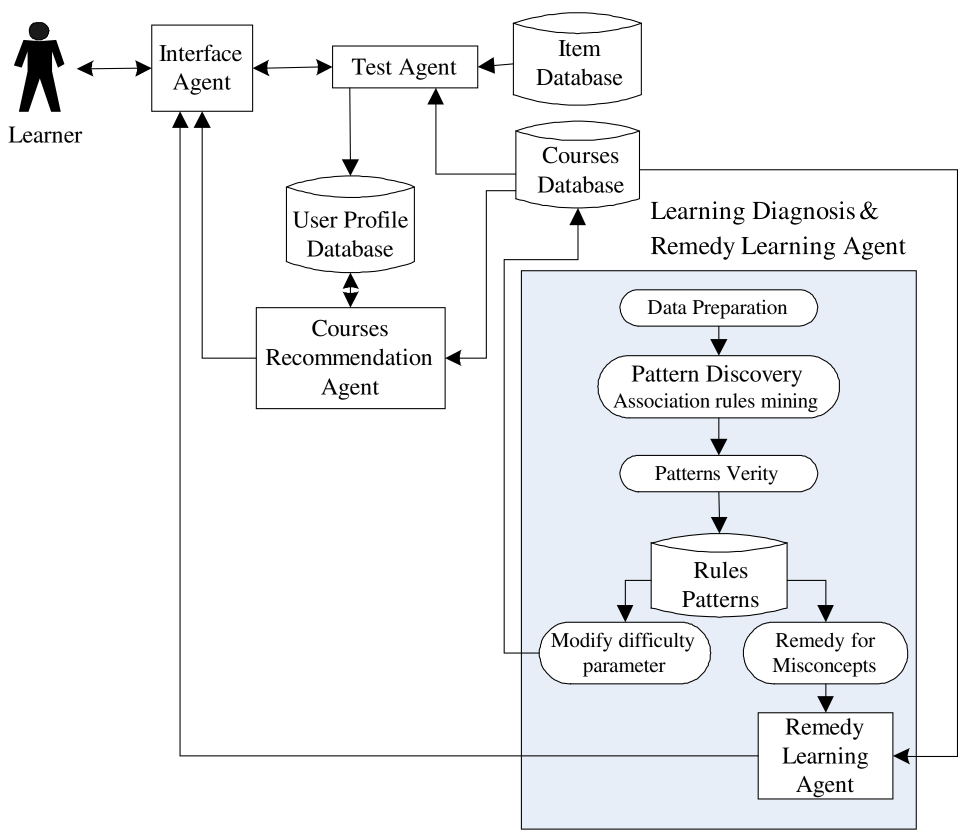


Fig. 9. The detailed operation procedures for the learning diagnosis and remedial learning agent.

database into suitable format for association rule mining. The details of operation procedure are described as follows:

Step 1: Select the needed database fields as the features, including e-mail, unit ID, courseware ID and test result. Table 1 illustrates the detailed database fields in this study.

Step 2: Count the total number of learned pages for each learner's account.

Step 3: Delete the learner's account that the total number of learned pages is less than the assigned threshold value.

Step 4: Transform data into the simplified representation notations for association rule mining. Table 2 gives the list of the simplified representation notations.

Table 1
The needed features for learning misconception diagnosis in the user profile database

Email	Learner account
Unit_id	Course unit
Content_id	The serial number of courseware
Answer responses of testing items	The on-line testing results of learner

Table 2
The transferred representation notations in the user profile database for association rule mining

○	Learner's correct testing item response
×	Learner's incorrect testing item response
?	This courseware has not been learned

2.3.3.2. Pattern discovery. In this study, pattern discovery aims to discover implicit learning misconceptions from a large amount of learner profile records. Among data mining techniques, the association rule is a well-known data mining technique, which can discover implicit relationships among item sets in a transaction database. Hence, in order to discover learning misconceptions based on the incorrect testing item responses and their corresponding courseware, Apriori algorithm (Agrawal & Srikant, 1994) is employed to perform the data mining task herein. Let $I = I_1, I_2, \ldots, I_m$ be a set of m distinct attributes, T be transaction that contains a set of items such that $T \subseteq I$, D be a database with different transaction records. An association rule is an implication in the form of $X \Rightarrow Y$, where X, $Y \subset I$ are sets of items called itemsets, and $X \cap Y = \varnothing$. The X is called antecedent while Y is called consequent, the rule means X implies Y.

There are two important basic measures for association rules, support(s) and confidence(c) (Agrawal, Imielinski, & Swami, 1993; Dunham Margaret, 2003), which can be defined as follows. The support (s) of an association rule is the ratio of the records that contain $X \cup Y$ to the total number of records in the database, and formulated as follows:

$$\mathrm{Support}(X \Rightarrow Y, T) = \mathrm{Support}(X \cup Y) \tag{1}$$

For a given number of records, confidence (c) is the ratio (in percent) of the number of records that contain $(X \cup Y)$ to the number of records that contain X, and formulated as follows:

$$\mathrm{Conf}(X \Rightarrow Y, T) = \frac{\mathrm{Supp}(X \cup Y)}{\mathrm{Supp}(X)} \tag{2}$$

Association rule mining aims to find out association rules that satisfy the pre-defined minimum support and confidence values from a given database. Besides, to insure that the discovered rules are interesting and accuracy, the certainty factor proposed in Delgado, Marán, Sánchez, and Vila (2003) is employed to select very strong rules for the proposed learning diagnosis approach. The certainty factor can be defined as follows:

$$\mathrm{CF}(X \Rightarrow Y) = \begin{cases} \frac{\mathrm{Conf}(X \Rightarrow Y) - \mathrm{Supp}(Y)}{1 - \mathrm{Supp}(Y)}, & \text{if } \mathrm{Conf}(X \Rightarrow Y) > \mathrm{Supp}(Y) \\ \frac{\mathrm{Conf}(X \Rightarrow Y) - \mathrm{Supp}(Y)}{\mathrm{Supp}(Y)}, & \text{if } \mathrm{Conf}(X \Rightarrow Y) \leqslant \mathrm{Supp}(Y) \end{cases} \tag{3}$$

where assuming that if $\mathrm{Supp}(Y) = 1$ then $\mathrm{CF}(X \Rightarrow Y) = 1$, and if $\mathrm{Supp}(Y) = 0$ then $\mathrm{CF}(X \Rightarrow Y) = -1$.

The certainty factor takes values in $[-1, 1]$. It is positive when the dependence is positive, 0 when there is independence, and a negative value when the dependence is negative. In sum, the proposed learning diagnosis approach using mining association rule is typically a three-step process, and described as follows:

Step 1: Find all sets of items, which occur with a frequency that is greater than or equal to the user-specified threshold support(s).

Step 2: Generate the desired rules using the large itemsets, which have user-specified threshold confidence(c).

Step 3: Verify the discovered rules using the certainty factor to judge whether the discovered rules are very strong rules or not.

The Apriori algorithm can be briefly listed as follows:

```
Input: Database, D, of transactions; minimum support threshold, min_sup
Output: L, frequent itemsets in D.

L1 = find_frequent_1 - itemsets(D);
for (k = 2; L(k-1) ≠ 0; k++) {
  Ck = apriori_gen(L(k-1), min_sup);
  for each transaction t ∈ D {//scan D for counts
    Ct = subset(Ck, t); //get the subsets of t that are candidates
    for each candidate c ∈ Ct
      c.count++;
  }
  Lk = {c ∈ Ck | c.count ⩾ min_sup}
}
Return L = ∪k Lk;
```

D	Transaction database
k-itemset	An itemset containing k items
C_k	Set of candidate k-itemsets (potentially frequent itemsets)
L_k	Set of frequent k-itemsets (k-itemsets with minimum support)
$\cup_k L_k$	Set of generated itemsets

Moreover, an example is illustrated in Fig. 10 to show the detailed process of Apriori algorithm for learning misconception diagnosis. This example includes a user profile database D, which contains 10 testing records with wrong answer in the database. The threshold of minimum support value is set to be 5. This example indicates that Apriori algorithm how to find large item sets for mining common misconception rules. According to the obtained large itemsets, the association rules which pass the user-specific confidence threshold can be generated.

2.3.3.3. Pattern verification. To evaluate the validity for the discovered learning misconceptions is an important task. In our study, after the learner finishes the learning process for the recommended courseware, learner will be guided to perform an overall posttest for the learned course unit. The test agent also gathers the answer sheet of the leaner for the verification of the discovered learning misconceptions. Here, we also use the association rule to mine the answer sheet in the posttest, then getting the discovered association rule set of R_2. Assume the discovered

Minimum support count=5

Sample Database	
TID	Item ID
T1	C1,C2
T2	C4,C5
T3	C1,C2,C4
T4	C2,C3
T5	C2,C4,C5
T6	C1,C2,C4,C5
T7	C1,C2,C3,C4
T8	C2,C3,C4
T9	C1,C5
T10	C1,C2,C4

C_1	
Item set	Supp
C1	6
C2	8
C3	3
C4	7
C5	4

L_1	
Item set	Supp
C1	6
C2	8
C4	7

C_2	
Item set	Supp
{C1,C2}	5
{C1,C4}	4
{C2,C4}	6

L_2	
Item set	Supp
{C1,C2}	5
{C2,C4}	6

C_3	
Item set	Supp
{C1,C2,C4}	4

The process is stopped since no itemsets in C_3 satisfy the assigned minimum support

Fig. 10. An example for the detailed process of Apriori algorithm.

association rule set from the on-line testing during learning process as set R_1. This study considers that the association rule set of R_1 belongs to learner's short memory, but the association rule set of R_2 from an overall posttest belongs to learner's long memory. To execute remedy learning more efficiently, the proposed method adopts the intersection of sets R_1 and R_2 to obtain the final diagnosis misconceptions and stores them into the rule pattern database for remedy learning and courseware structure modification.

2.3.3.4. Rule patterns for remedy learning and courseware structure modification. In this study, the discovered rule patterns for learner's common learning misconceptions are applied to modify the difficulty parameters of courseware in the courseware database. This result will lead to the change of the courseware recommendation sequence for the proposed personalized e-learning system. Moreover, the discovered learning misconceptions are also used to perform remedy learning to promote learning performance for individual learner. The experimental results are demonstrated and discussed in the next section.

3. Experiments

The PELS was published on the web site http://192.192.6.86/irt4 to provide personalized e-learning services and enable the performance of the proposed e-learning system in recommending personalized course material to be evaluated. To verify the quality of rules patterns for learner misconceptions in the proposed personalized e-learning system with learning diagnosis and remedy learning mechanisms, 630 third grade students of Taipei County Jee-May Elementary School who have majored in the course unit of "Fraction" of mathematics of elementary school were invited to participate in this experiment. The experimental environment and results are analyzed and described as follows.

3.1. The designed course materials in the course unit "Fraction"

Currently, under the course category, "Mathematics of elementary school", the proposed system contains one course unit, "Fraction", and includes 34 course materials designed by several mathematical teachers. Among 34 course materials in the course unit "Fraction", a half of course materials are designed to perform the normal learning process, and another half of course materials are used to perform the remedy learning process after learners' common misconceptions are diagnosed. Moreover, each course material has a corresponding difficulty parameter, initially determined by course experts. All designed course materials for the normal learning process and their corresponding difficulty parameters are listed in Table 3.

3.2. The gathered user profile database

Table 4 illustrates a list of data statistics in the user profile database for the learning misconception mining. The total number of learning records in the user profile database generated from 630 learners is 23,721. To filter out those learners who only study few course materials in the course unit, the assigned threshold for the number of learned courseware is set to be 8. This is because they may affect the accuracy of mining common learning misconceptions. Finally, the number of learners that satisfies the assigned threshold of the number of learned courseware is 607. Besides, the testing item database contains 393 testing questions, which can be randomly selected by PELS system to evaluate learners' if understand the learned courseware. Moreover, except for the on-line testing records, the number of the final posttest records is 770.

3.3. Common learning misconception diagnosis

Table 5 lists the number of the discovered misconception rules under the assigned threshold values for Apriori asso-

Table 3
The contents of the designed course materials and the difficulty levels of the corresponding course materials in the course unit "Fraction"

Course material	Concept description	The difficulty level of course material
Equal parts	To understand the meaning of "equal parts" is to divide a unit into n equal parts	−1.8
Division as sharing	To use the concept of "equal parts" solves the problem of "division as sharing". Division as sharing means that a given set is partitioned into a specified number of groups to determine how many partitions are in each equal group	−1.5
Division as separating	To use the concept of "equal parts" solves the problem of "division as separating". Division as separating means that a given set is partitioned by a specified amount to determine how many equal groups	−1
Sharing with a remainder	To use the concept of "equal parts" solves the problem of "division as sharing with remainder"	−0.1
Separating with a remainder	To use the concept of "equal parts" solves the problem of "division as separating with remainder"	0
Parts of a whole	Identifying the numerator and denominator of a fraction and expressing improper fractions as whole	0.1
Improper fractions	Identifying proper and improper fractions	0.2
Sequence of fractions	Order the fractions and find the fractional value on a number line	0.4
Compare proper fractions with the same denominator	To compare fractions with the same denominator, look at their numerators. The larger fraction is the one with the larger numerator	0.5
Compare proper fractions with different denominators	To compare fractions with different denominator	0.7
Add and subtract fractions	Adding and subtracting fractions when the denominators are the same	1.2
Adding fractions	Adding fractions with the same denominators	0.8
Subtracting fractions	Subtracting fractions with the same denominators	1
Missing addend	Perform missing addend fractions problems with the same denominators	1.3
Missing subtrahend	Perform missing subtrahend fractions problems with the same denominators	1.5
Missing summand	Perform missing subtrahend fractions problems with the same denominators	1.6
Missing minuend	Perform missing minuend fractions problems with the same denominators	1.8

Table 4
Data statistics in the user profile database

Data items	The number of records
The number of learners who participates in this experiment	630
The number of course materials in the course unit "Fraction" for the normal learning process	17
The number of course materials in the course unit "Fraction" for the remedy learning process	17
The total number of learning records in the user profile database	23,721
The assigned threshold for the number of learned courseware	8
The number of learners that satisfies the assigned threshold for the number of learned courseware	607
The number of testing items in the testing item database	393
The number of final posttest records (the times of the final posttest for each learner may be over one time)	770

ciation rule mining approach. The rule sets R_1 and R_2 contains 259 and 21 rules under the assigning threshold values of minimum support, confidence, and certainty factor, respectively. The intersection of the rule sets R_1 and R_2 includes 12 strong rules for the learners' common misconceptions. Table 6 illustrates 12 common learning misconception rules in the course unit "Fraction" discovered by the proposed association rule learning diagnosis approach. We find that the most of the discovered misconception rules satisfy the learning hierarchy of concept in the course unit "Fraction". For example, the rule 1 indicates that occurring the misconception "Division as sharing" implies that the misconception "Parts of a whole" will also occur. This rule shows the previous concept "Division as sharing" is pre-requisite knowledge of the latter concept "Parts of a whole". The rules 3, 4, 5, 10, 11 and 12 are the similar cases with the rule 1. Moreover, the rules 6, 7, 8 and 9 indicate that occurring the misconception "Adding fractions", "Missing subtrahend", "Add and subtract fractions" or "Missing addend" implies that the misconception "Compare proper fractions with different denominators" will also occur. To analysis these occurring misconceptions in detail, we find that they are completely related to the concept of "Proper fractions with the same denominator". The reason could be that most of learners cannot completely understand the concept of "Proper fractions with the same denominator", thus leading to occurring the misconception of "Proper fractions with different denominators". Fig. 11 gives a statistics comparison of the number of learners who give the incorrect answers for the corresponding course materials between the on-line testing and final posttest. This statistics data support the results of common misconception diagnosis because the most of learners cannot pass the testing questions of both the concepts of "Compare proper fractions with different denominators" and "Compare proper fractions with the same denominator" neither

Table 5
The number of the discovered rules under the assigned threshold values

The rule set R_1 (online testing)			The rule set R_2 (posttest)			The intersection of rule sets R_1 and R_2
S	C	The number of rules	S	C	The number of rules	
0.05	0.3	259	0.02	0.3	21	12

Table 6
The discovered learning misconception rules for the course unit "Fraction" (where S, C and CF indicate the values of support, confidence and certainty factor for each discovered rule, respectively)

Rule id	The discovered learning misconception rule	The rule set R_1			The rule set R_2		
		S	C	CF	S	C	CF
1	Division as sharing → Parts of a whole	0.107	0.519	0.357	0.033	0.345	0.110
2	*Adding fractions → Parts of a whole*	*0.071*	*0.346*	*0.126*	*0.022*	*0.371*	*0.146*
3	Division as sharing → Compare proper fractions with different denominators	0.119	0.574	0.064	0.05	0.527	0.17
4	Improper fractions → Compare proper fractions with different denominators	0.209	0.673	0.281	0.064	0.552	0.214
5	Sequence of fractions → Compare proper fractions with different denominators	0.234	0.649	0.228	0.069	0.533	0.181
6	*Adding fractions → Compare proper fractions with different denominators*	*0.134*	*0.654*	*0.240*	*0.031*	*0.514*	*0.148*
7	*Missing subtrahend → Compare proper fractions with different denominators*	*0.079*	*0.547*	*0.003*	*0.052*	*0.577*	*0.258*
8	*Add and subtract fractions → Compare proper fractions with different denominators*	*0.157*	*0.646*	*0.221*	*0.05*	*0.453*	*0.04*
9	*Missing addend → Compare proper fractions with different denominators*	*0.14*	*0.566*	*0.046*	*0.056*	*0.540*	*0.195*
10	Sequence of fractions → Compare proper fractions with the same denominator	0.205	0.569	0.294	0.041	0.32	0.093
11	Subtracting fractions → Missing summand	0.06	0.344	0.129	0.022	0.406	0.208
12	Missing subtrahend → Missing summand	0.06	0.413	0.220	0.029	0.327	0.102

in the on-line testing or final posttest process, thus leading to eight common misconception rules related to the two concepts.

Moreover, the rule 2 indicates that occurring the misconception "Adding fractions" implies that the misconception "Parts of a whole" will also occur. This result is not so logical according to the concept hierarchy of courseware in the course unit "Fraction", but this misconception indeed occurs frequently for most learners in our experiment. This reason could derive from the poor courseware design or testing questions for the two course materials. However, this result also benefits teachers to modify the courseware structure or content. Next, we will describe how to apply the obtained association rules for learner's common mis-

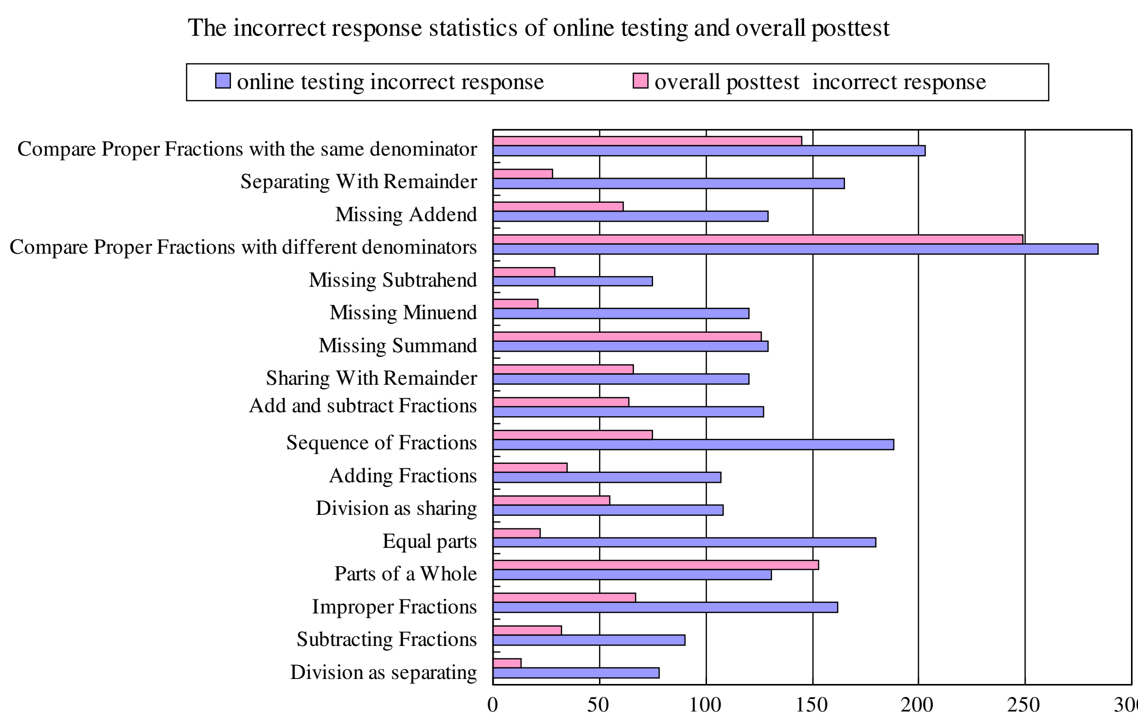


Fig. 11. Statistics comparison of the number of learners who give the incorrect answer for the corresponding course material between the on-line testing and final posttest.

Table 7
An example for modifying difficulty parameters of courseware

Rule Patterns	The original difficulty parameters of courseware in the courseware database	Revise difficulty parameter
$A \rightarrow B$ (this rule stands for the courseware A is easier than the courseware B)	$A = 0.8$, $B = 0.5$ ($A > B$)	Yes
	$A = 0.6$, $B = 0.8$ ($A < B$)	No

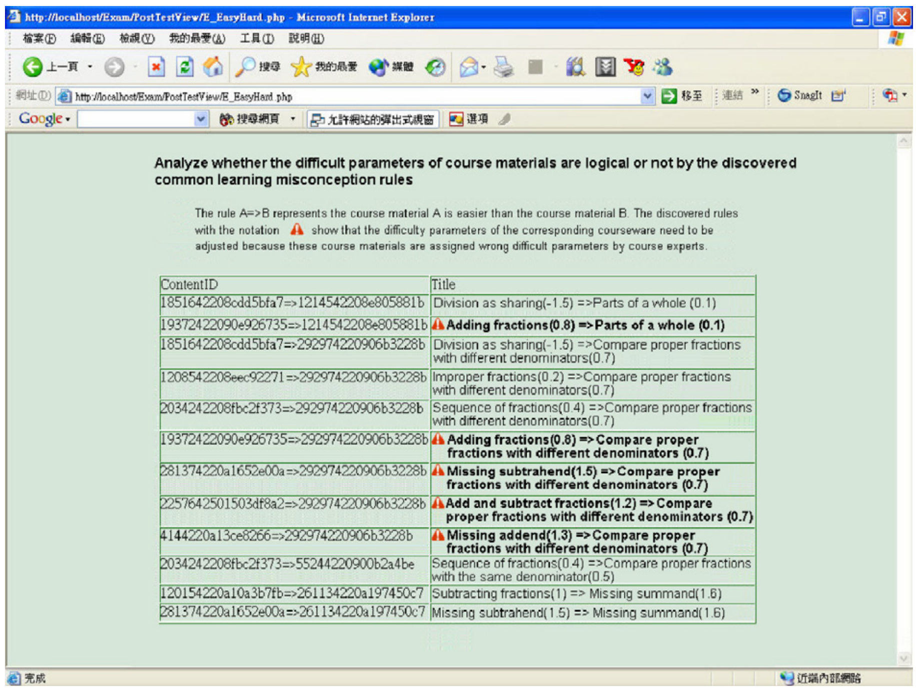


Fig. 12. The teacher interface for the discovered misconception rules that need to be adjusted the difficulty parameters (the notation ⚠ shows that the difficulty levels of the corresponding courseware need to be adjusted).

Table 8
The strategy for the remedy learning

Assume the discovered common misconception set R_1 from the on-line testing during learning processes	The discovered common misconception set R_2 from an overall posttest		Perform remedy learning determined by the set $R_1 \cap R_2$
$A \rightarrow B$	Learner 1	No match	No
$D \rightarrow E$	Learner 2	$D \rightarrow E$	Yes
	....	....	....

conceptions to tune the courseware structure through modifying the difficulty parameters of courseware in the courseware database.

3.4. Modifying the difficulty parameters of courseware

Assume that the proposed method discovered the rule patterns $A \rightarrow B$ for learning misconceptions, the rule represents that occurring misconception A implies that the misconception B will also occur. In other words, this rule indicates that the misconception A is easier than the misconception B. Therefore, assume the original difficulty levels of courseware in the courseware database for both the misconceptions A and B are that A is larger than B. The PELS will suggest that the difficulty parameters of courseware A and B should be modified to generate appropriate courseware recommendation sequence for the proposed personalized e-learning system. Conversely, assume the original difficulty parameters in the courseware database for the misconceptions A and B are that A is smaller than B. The difficulty parameters of courseware A and B will not be changed. Table 7 gives an example to explain the strategy for modifying difficulty parameters of courseware. In Table 6, the discovered misconception rules marked by *italics* represent that their difficulty parameters need to be adjusted. In addition, Fig. 12 shows that the proposed system also provides

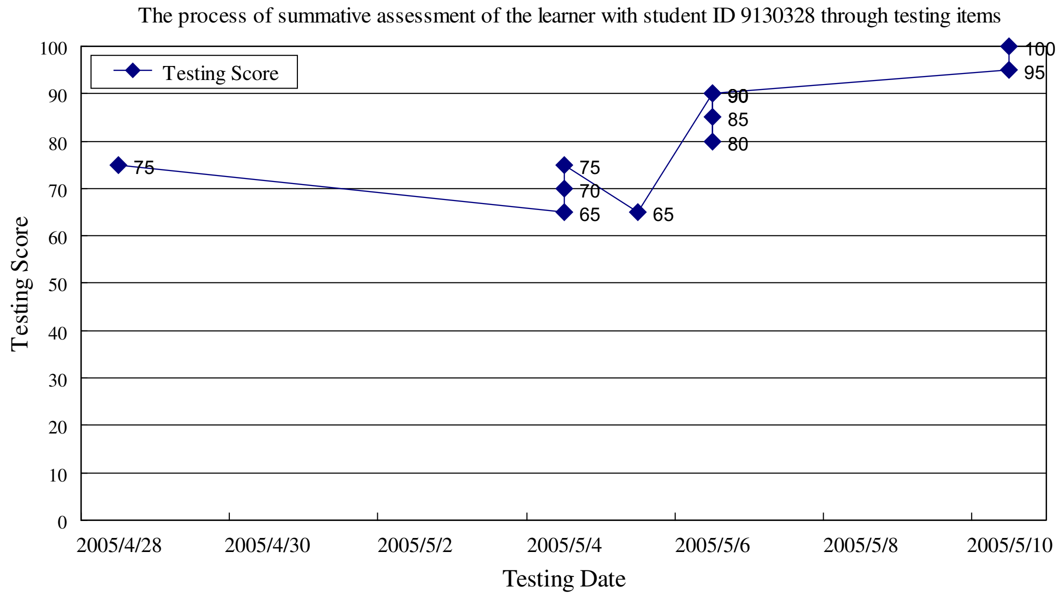


Fig. 13. The learning performance promotion curve of a learner with low learning ability after performing remedy learning.

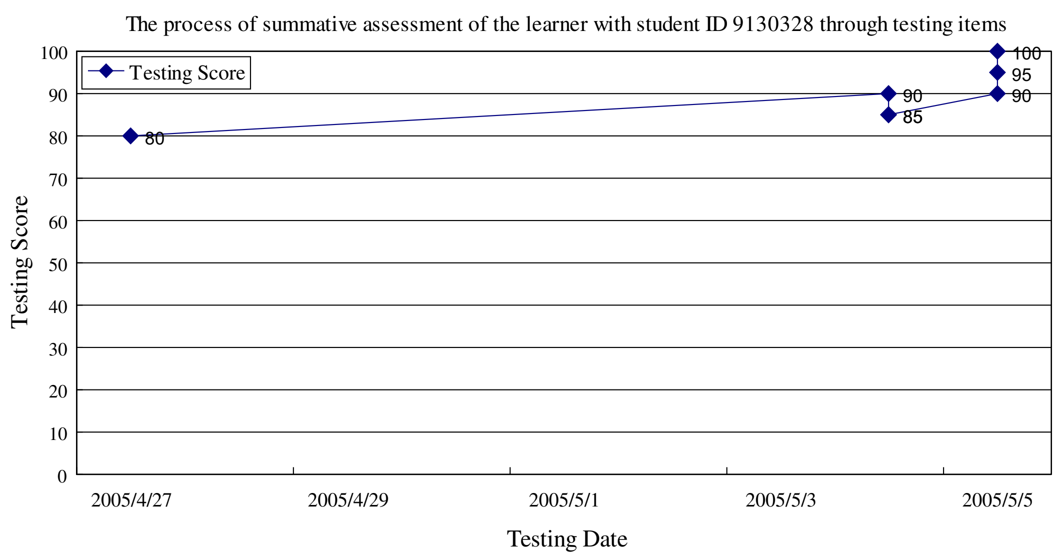


Fig. 14. The learning performance promotion curve of a learner with moderate learning ability after performing remedy learning.

an interface for teachers to remind the discovered misconception rules that need to be adjusted the difficulty parameters.

3.5. Performing remedy learning

According to the discovered rule patterns for common learning misconception diagnosis, the remedy learning agent will compare the discovered common rule pattern set R_1 from the on-line testing during learning processes with the discovered common rule pattern stet R_2 from an overall posttest to determine the remedy learning strategy. Table 8 gives an example to show the strategy for the remedy learning. If the set $R_1 \cap R_2$ is not empty set, then the remedial learning agent will guide the learner to learn the remedial course materials in the remedy courseware database. This is very helpful to learners because the learners' common misconceptions can be enhanced again to promote the learning performance.

To demonstrate the learning performance promotion of learners after performing the remedy learning, Figs. 13–15 show the learning performance promotion curves for the learner with low, moderate and high learning abilities, respectively. We find that performing the remedy learning process is obviously helpful to speed up the learning performance for the observed three learners with different learning abilities because their testing scores gradually progress. Fig. 16 reveals the learning performance promotion curve

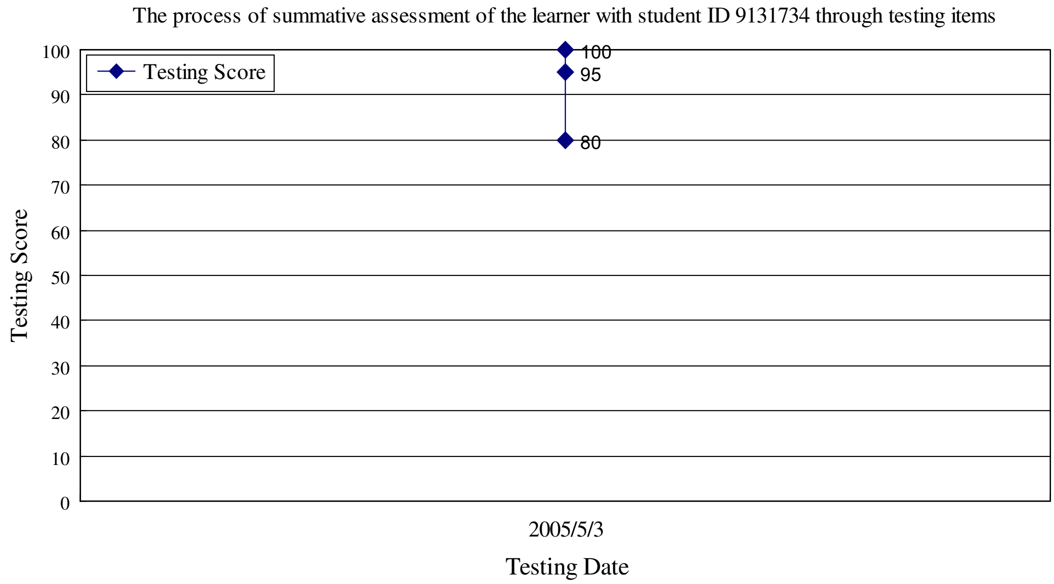


Fig. 15. The learning performance promotion curve of a learner with high learning ability after performing remedy learning.

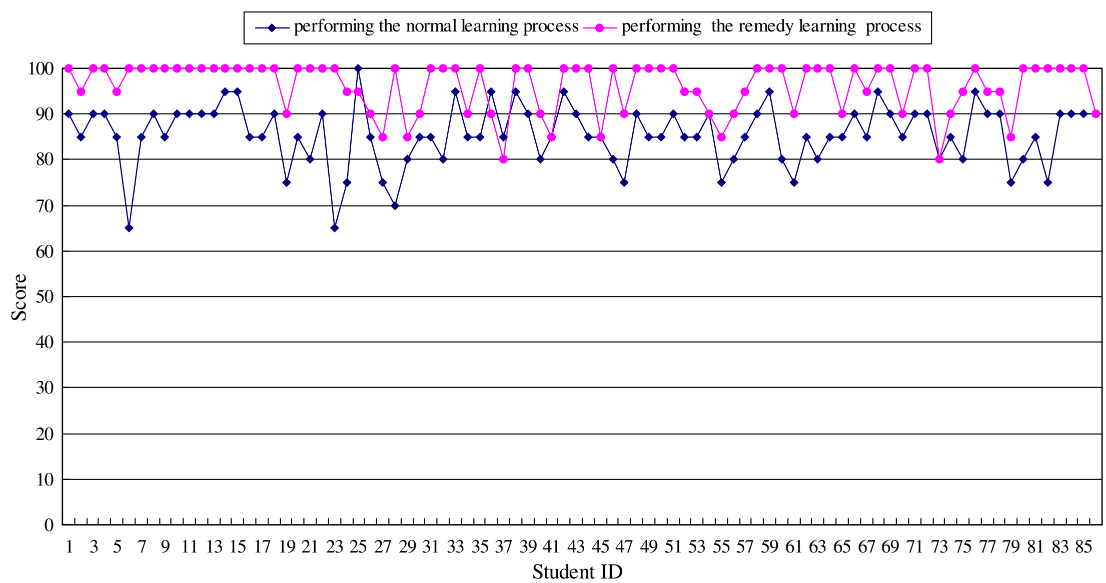


Fig. 16. The learning performance promotion curve for those learners with over second times testing scores.

for 86 learners who perform over second times posttest. In this study, the first posttest score is served as the testing score after performing the normal learning process, and the final posttest score is served as the testing score after

Table 9
The statistics analysis after performing the remedy learning

Learning strategy	Compared item								
	The number of the learners with over second times testing scores	Mean testing score	Standard deviation	Minimum testing score	Maximum testing score	Percentile value			Asymp.Sig. (two-tailed)
						25th	50th	75th	
Only performing the normal learning process	86	85.5233	6.7314	65	100	80	85	90	0.000
Performing the remedy learning except for the normal learning process	86	96.2209	5.5273	80	100	90	100	100	

Table 10
The satisfaction evaluation after learning

Question type	Question number	Question	Percentage (%)				
			Very approved	Approved	No opinion	Disapproved	Very disapproved
The services of software and hardware	1	The personalized e-learning system has a friendly user interface to support the provided e-learning services	59.5	25.7	10.1	3.9	0.8
	2	The personalized e-learning system enhances my impression of learning mathematics due to providing the lively multimedia course materials with flash animation and voice	41.1	34.9	17.9	3.4	2.8
	3	The designed course materials on the personalized e-learning system are very clear and interesting to convey the mathematical concepts	57	20.1	17.0	4.7	1.1
	4	I can completely understand the meaning of the testing questions that appears on the personalized e-learning system	48.6	24.3	20.9	5.6	0.6
	5	I can completely understand the meaning of course materials that appears on the personalized e-learning system	46.1	29.9	17.0	5.3	1.7
	6	I think that the personalized e-learning system can promote my learning interests because I can actively learn mathematics at any time and place	62.3	24.3	7.5	3.6	2.2
	7	I like the learning mode that the PELS gives me a testing question after learning a course material because I can know whether understanding the learned course material or not	47.8	27.4	14.5	7.3	3.1
	8	I like the learning mode that the PELS provides a remedy learning process after finishing a final testing because I can enhance these course materials with common learning misconceptions again	64.2	17.9	13.1	2.2	2.5
	9	I think that the most of remedial course materials recommended by the PELS system are indeed my weak concepts in the learned course unit	44.1	27.4	18.7	3.4	6.4
		Average		*78.1*	*15.2*		*6.7*
Learning performance promotion	10	I think that the PELS can promote my learning confidence specially after I pass the remedial course materials and the corresponding testing questions	57.5	25.4	13.4	2.0	1.7
	21	Compared to the traditional classroom learning, I think that the PELS help me to construct clear mathematical concepts in the learned course unit under the environment of interacting learning	47.5	23.7	22.6	4.2	2.0
	22	I think that my mathematical score has obvious progress due to using the PELS to assist mathematical learning	58.1	24.3	13.4	1.7	2.5
		Average		*78.8*	*16.5*		*4.7*
Interactive design between the PELS system with learners	13	Do not need teachers direct me, I can learn the courseware alone using the PELS system	33.8	22.1	18.7	15.9	9.5
	14	I can learn more efficiently if I can obtain teachers' assistance while I learn the mathematical courseware using PELS system	66.8	20.7	9.5	2.0	1.1
	15	I think that using the PELS to learn mathematics will reduce the communication opportunity with teachers	26.5	15.1	31.0	17.6	9.8

Table 10 (*continued*)

Question type	Question number	Question	Percentage (%)				
			Very approved	Approved	No opinion	Disapproved	Very disapproved
		Average	*61.7*		*19.7*	*18.6*	
Active learning attitude	11	I feel that using PELS system to learn mathematics is very interesting learning mode	66.2	21.8	9.2	1.7	1.1
	12	I feel that the learning contents of courseware on the PELS system can excite my learning interests	56.1	24.3	14.8	3.1	1.7
	16	I would like to reply the testing question again when I give an incorrect answer for the given testing question	50.3	28.3	14.5	3.4	3.6
	17	I feel that the time passes very quickly when I use the PELS system for the learning of mathematics	42.7	28.2	21.2	4.2	3.6
	18	I feel very exciting while I know to learn mathematics using the PELS system at first time	48.9	27.4	17.0	4.5	2.2
	19	I would like to learn the courseware of mathematics again by the PELS system because using the PELS system to learn mathematics is very convenient	50.6	28.2	15.4	3.6	2.2
		Average	*78.8*		*15.4*	*5.8*	
Using PELS system at available time	20	Do you have computer at home? (Please continue to reply the question 20-1 if your answer is yes)	Yes				No
			84.9% (304 learners)				15.1% (54 learners)
	20-1	Can you use Internet at home? (Please continue to reply the question 20-2 if your answer is yes)	Yes			No	
			84.9% (258 learners)			15.1% (46 learners)	
	20-2	Did you ever use the PELS system to learn the mathematical courseware at home?	Often		Never		
			55% (142 learners)		45% (116 learners)		

performing the remedy learning. From the statistical result shown in Fig. 16, we find that 78 learners promote their testing scores, three learners reduce their testing scores, and five learners keep their testing scores after performing the remedy learning process. Additionally, Table 9 gives the statistics analysis of learning performance for the observed 86 learners after performing the remedy learning. We can find that the mean testing score and standard deviation are, respectively, 85.5233 and 6.7314 before performing the remedy learning process, but the mean testing score is promoted from 85.5233 to 96.2209 and the standard deviation is reduced from 6.7314 to 5.5273 after performing the remedy learning process. Moreover, the index value of Asymp.Sig. is 0.000 to show that the progress of learners' scores is significant. The results of statistics analysis are very encouraging because performing the remedy learning process using the results of common learning misconception diagnosis is indeed helpful for the learning performance promotion.

3.6. Evaluating degree of satisfaction

Finally, we design a feedback form on the Internet to evaluate learners' satisfaction for the personalized e-learning system with the mechanisms of the common learning misconception diagnosis and remedy learning. The feedback form involves 24 questions distinguished five various question types to measure if the provided learning services in the personalized e-learning system satisfy most learners' requirements. The five question types contain the satisfaction degree of software and hardware services, the satisfaction degree of the learning performance promotion, the satisfaction degree of the interactive design between the PELS system with learners, the investigation of the active learning attitude, and if using the PELS system for the mathematical learning at available time. Totally, 358 learners who participated in our experiment logged in our system to fill this feedback form through Internet interface. The evaluating results of satisfaction degree are listed in Table 10. To conveniently observe the evaluating results, the investigation results of "very approved" and "approved" are merged as "approved", and the investigate results of "disapproved" and "very disapproved" are merged as "disapproved".

The evaluating results indicate that the satisfaction degrees of "approved" are over 78% in terms of the software and hardware services, learning performance promotion, and active learning attitude. Moreover, in the investigation item of interactive design between the PELS system with learners, the 55.9% learners replied that he/she could learn the designed mathematical courseware alone using the PELS system under no teachers direct him/her. The 87.5% learners replied that he/she could learn more efficiently if teachers could give assistance while he/she learnt the mathematical courseware using PELS system. The 41.6% learners thought that using the PELS to learn mathematics will reduce the communication opportunity with teachers. Additionally, 142 learners ever used the PELS system to learn the mathematical courseware if he/she can use Internet to learn mathematical courseware at home. This phenomenon shows that the proposed learning mode can be accepted by most of learners.

4. Conclusion

This study presents an association rule learning diagnosis approach for learning misconception diagnosis. The proposed approach can discover learner's misconceptions according to incorrect testing item responses during learning processes. The obtained association rules for learner's common misconceptions can be applied to tune courseware structure through modifying the difficulty parameters of courseware in the courseware database, thus obtaining appropriate courseware recommendation sequence. Moreover, the proposed system also offers a remedy learning strategy based on the discovered learner misconceptions to promote learning performance. Experimental results indicate that the proposed method can precisely discover learner's misconceptions based on the response answers of testing items and help learners to enhance their misconceptions for learning performance promotion.

Acknowledgment

The authors thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC94-2520-S-026-002.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD conference on management of data*, Washington, DC, May (pp. 207–216).

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, & C. Zaniolo (Eds.), *Proceedings of the 20th international conference on very large data bases (VLDB'94)* (pp. 487–499). Los Altos, CA: Morgan Kaufmann.

Baker, Frank B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

Berghel, H. (1997). Cyberspace 2000: dealing with information overload. *Communications of the ACM, 40*(2), 19–24.

Borchers, A., Herlockerm, J., Konstanand, J., & Riedl, J. (1998). Ganging up on information overload. *Computer, 31*(4), 106–108.

Brusilovsky, Peter (1999). Adaptive and intelligent technologies for web-based education. In C. Rollinger & C. Peylo (Eds.). *Special Issue on Intelligent Systems and Teleteaching* (Vol. 4, pp. 19–25). Künstliche Intelligenz.

Chang, Kuo-En, Liu, Sei-Hua, & Chen, Sei-Wang (1998). A testing system for diagnosing misconceptions in DC electric circuits. *Computers & Education, 31*(2), 195–210.

Cheng, Shein-Yung, Lin, Chia-Sheng, Chen, Hsien-Hsun, & Heh, Jia-Sheng (2005). Learning and diagnosis of individual and class conceptual perspectives: an intelligent systems approach using clustering techniques. *Computers & Education, 44*(3), 257–283.

Chen, Chih-Ming, Lee, Hahn-Ming, & Chen, Ya-Hui (2005). Personalized e-learning system using item response theory. *Computers & Education, 44*(3), 237–255.

Chen, Chih-Ming, Liu, Chao-Yu, & Chang, Mei-Hui (2006). Personalized curriculum sequencing utilizing modified item response theory for web-based instruction. *Expert Systems with Applications, 30*(2), 378–396.

Delgado, Miguel, Marán, Nicolís, Sánchez, Daniel, & Vila, María-Amparo (2003). Fuzzy association rules: general model and applications. *IEEE Transactions on Fuzzy Systems, 11*(2), 214–225.

Dunham Margaret, H. (2003). *Data mining introductory and advanced topics*. Englewood Cliffs, NJ: Prentice-Hall/Pearson Education.

Huang, Chenn-Jung, Lium, Ming-Chou, Chu, San-Shine, & Cheng, Chin-Lun (2004). Application of machine learning techniques to web-based intelligent learning diagnosis system. In *Proceedings of fourth international conference on hybrid intelligent systems* (pp. 242–247).

Hwang, Gwo-Jen, Hsiao, Chia-Lin, & Tseng, Judy C. R. (2003). A computer-assisted approach to diagnosing student learning problems in science courses. *Journal of Information Science and Engineering, 19*, 229–248.

Jong, BinShyan, Lin, TsongWuu, Wu, YuLung, & Chan, Teyi (2004). Diagnostic and remedial learning strategy based on conceptual graphs. *Journal of Computer Assisted Learning, 20*(5), 377–386.

Lo, J. J., Wang, H. M., & Yeh, S. W. (2004). Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia. *Computers & Education, 42*(1), 45–63.

Papanikolaou, Kyparisia A., & Grigoriadou, Maria (2002). Towards new forms of knowledge communication: the adaptive dimension of a web-based learning environment. *Computers & Education, 39*, 333–360.

Taipei County Jee-May Elementary School. Web site available at http://www.jmes.tpc.edu.tw/.

Tang, Tiffany Ya, & Mccalla, Gordon (2003). Smart recommendation for evolving e-learning system. In *Proceedings of the 11th international conference on artificial intelligence in education, workshop on technologies for electronic documents for supporting learning*, July 20–24, Sydney, Australia (pp. 699–710).

Weber, Gerhard, & Specht, Marcus (1997). User modeling and adaptive navigation support in WWW-based tutoring systems. In *Proceedings of the sixth international conference on user modeling* (pp. 289–300).