Available online at www.sciencedirect.com
ScienceDirect
Computers & Education 49 (2007) 691–707
COMPUTERS & EDUCATION
www.elsevier.com/locate/compedu

A new approach for constructing the concept map

Shian-Shyong Tseng [a,b,*], Pei-Chi Sue [a,1], Jun-Ming Su [a,2], Jui-Feng Weng [a,1], Wen-Nung Tsai [a,3]

[a] *Department of Computer Science, National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu, Taiwan 300, Taiwan*
[b] *Department of Information Science and Applications, Asia University No. 500, Liufeng Rd., Wufeng Shiang, Taichung, Taiwan 413, Taiwan*

Received 24 August 2005; received in revised form 31 October 2005; accepted 7 November 2005

Abstract

In recent years, e-learning system has become more and more popular and many adaptive learning environments have been proposed to offer learners customized courses in accordance with their aptitudes and learning results. For achieving the adaptive learning, a predefined concept map of a course is often used to provide adaptive learning guidance for learners. However, it is difficult and time consuming to create the concept map of a course. Thus, how to automatically create a concept map of a course becomes an interesting issue. In this paper, we propose a Two-Phase Concept Map Construction (TP-CMC) approach to automatically construct the concept map by learners' historical testing records. Phase 1 is used to preprocess the testing records; i.e., transform the numeric grade data, refine the testing records, and mine the association rules from input data. Phase 2 is used to transform the mined association rules into prerequisite relationships among learning concepts for creating the concept map. Therefore, in Phase 1, we apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic data, apply Education Theory to further refine it, and apply Data Mining approach to find its grade fuzzy association rules. Then, in Phase 2, based upon our observation in real learning situation, we use multiple rule

* Corresponding author. Tel.: +886 3 5712121 56662; fax: +886 3 5721490.
E-mail addresses: sstseng@cis.nctu.edu.tw (S.-S. Tseng), gis91510@cis.nctu.edu.tw (P.-C. Sue), jmsu@csie.nctu.edu.tw (J.-M. Su), roy@cis.nctu.edu.tw (J.-F. Weng), tsaiwn@csie.nctu.edu.tw (W.-N. Tsai).
[1] Tel.: +886 3 5712121 56658; fax: +886 3 5721490.
[2] Tel.: +886 3 5712121 56658; fax: +886 3 5724176.
[3] Tel.: +886 3 5712121 31882; fax: +886 3 5724176.

0360-1315/$ - see front matter
doi:10.1016/j.compedu.2005.11.020

types to further analyze the mined rules and then propose a heuristic algorithm to automatically construct the concept map. Finally, the Redundancy and Circularity of the concept map constructed are also discussed. Moreover, we also develop a prototype system of TP-CMC and then use the real testing records of students in junior high school to evaluate the results. The experimental results show that our proposed approach is workable.

Keywords: Adaptive learning environments; Concept map; Data mining; Testing records

1. Introduction

With vigorous development of the Internet, e-learning system has become more and more popular. Therefore, in the last 5 years, many adaptive learning and testing systems have been proposed to offer learners customized courses in accordance with their aptitudes and learning results (Appleby, Samuels, & Jones, 1997; Carchiolo, Longheu, & Malgeri, 2002; Chang, Liu, & Chen, 1998; Frosini, Lazzerini, & Marcelloni, 1998; Gamboa, 2001; Hsu, Tu, & Hwang, 1998; Hwang, 2003; Hwang, Hsiao, & Tseng, 2003; Triantafllou, Pomportsis, & Demetriadis, 2003; Tsai, Tseng, & Lin, 2001). For achieving the adaptive learning, a predefined concept map of a course, which provides teachers for further analyzing and refining the teaching strategies, is often used to generate adaptive learning guidance. However, it is difficult and time consuming to create the concept map of a course. Thus, how to automatically create a correct concept map of a course becomes an interesting issue.

Therefore, in this paper, we propose a Two-Phase Concept Map Construction (TP-CMC) algorithm to automatically construct a concept map of a course by historical testing records. In the first phase, we apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic, apply Education Theory (Item Analysis for Norm-Referencing) to further refine it, and apply Data Mining approach to find its grade fuzzy association rules. The mined grade fuzzy association rules include four rule types, L–L, L–H, H–L, and H–H, which denote the casual relations between learning concepts of quizzes. For example, if a rule type is $Q_1 \cdot L \rightarrow Q_2 \cdot L$ which means that learners get low grade on quiz Q_1 implies that they may also get low grade on quiz Q_2. We call this rule type is L–L type. The previous articles use single rule type, e.g. L–L type, to analyze the testing data, which may decrease the quality of concept map (Hsu et al., 1998; Hwang et al., 2003; Tsai et al., 2001). Therefore, in the second phase, based upon our observation in real learning situation, we use multiple rule types to further analyze the mined rules and then propose a heuristic algorithm to automatically construct the concept map according to analysis results, which can be used to develop adaptive learning system and refine the learning strategies of learners.

The main contributions of this paper are:

(1) Apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic data, Education Theory (Item Analysis for Norm-Referencing) to further refine it, and Data Mining approach to find its grade fuzzy association rules.

(2) Analyze the mined association rules to generate related prerequisite relationships among concept sets of test item based on our observation in real learning situation.
(3) Propose a heuristic algorithm to automatically construct the concept map of a course.

2. Related work

Novak (1998) proposed Concept Map to organize or represent the knowledge as a network consisting of nodes (points/vertices) as concepts and links (arcs/edges) as the relations among concepts. Thus, a wide variety of different forms of concept maps have been proposed and applied in various domains (Bruillard & Baron, 2000; Gaines & Shaw, 1995; Gordon, 2000). In the adaptive learning environment, the Concept Map can be used to demonstrate how the learning status of a concept can possibly be influenced by learning status of other concepts and give learners adaptive learning guidance.

Thus, Appleby et al. (1997) proposed an approach to create the potential links among skills in math domain . The direction of a link is determined by a combination of educational judgment, the relative difficulty of skills, and the relative values of cross-frequencies. Moreover, a harder skill should not be linked forwards to an easier skill. As shown in Table 1, $f_{\bar{A}B}$ represents the amount of learners with wrong answers of skill A and right answers of skill B. If $f_{A\bar{B}} > f_{\bar{A}B}$, a skill A could be linked to a harder skill B, but backward link is not permitted.

Hsu et al. (1998) also proposed a conceptual map-based notation, called Concept Effect Relationships (CER), to model the learning effect relationships among concepts. In brief, for two concepts, C_i and C_j, if C_i is the prerequisite for efficiently learning the more complex and higher level concept C_j, then a CER $C_i \rightarrow C_j$ exists. A single concept may have multiple prerequisite concepts, and can also be a prerequisite concept of multiple concepts. Thus, based upon CER, the learning guidance of necessary concepts to enhance their learning performance can be derived by analyzing the test results of students. Later, based upon statistical prediction and approach of Hsu et al. (1998), a CER Builder was proposed by Hwang et al. (2003). Firstly, CER Builder finds the test item that most students failed to answer correctly and then collects the other test items failed to answer by the same students. Thus, CER Builder can use the information to determine the relationships among the test items. Though the CER Builder is easy to understand, only using single rule type is not enough to analyze the prerequisite relationship among concepts of test items, which may decrease the quality of concept map.

Tsai et al. (2001) proposed a Two-Phase Fuzzy Mining and Learning Algorithm. In the first phase, Look Ahead *F*uzzy *M*ining Association Rule *Alg*orithm (LFMAlg) was proposed to find the embedded association rules from the historical learning records of students. In the second phase, the AQR algorithm was applied to find the misconcept map indicating the missing concepts during students learning. The obtained misconcept map as recommendation can be fed

Table 1
Relative skills frequency

	A is right	A is wrong
B is right	f_{AB}	$f_{\bar{A}B}$
B is wrong	$f_{A\bar{B}}$	$f_{\bar{A}\bar{B}}$

back to teachers for remedy learning of students. However, because the creating misconcept map, which is not a complete concept map of a course, only represents the missing learning concepts, its usefulness and flexibility are decreased. In addition, their approaches generate many noisy rules and only use single rule type to analyze the prerequisite relationship among learning concepts.

3. Two-phase concept map construction (TP-CMC)

In TP-CMC, the Test item-Concept Mapping Table records the related learning concepts of each test item. As shown in Table 2, five quizzes contain these related learning concepts *A*, *B*, *C*, *D* and *E*, where "1" indicates the quiz contains this concept, and "0" indicates not. Moreover, a concept set of quiz *i* is denoted as CS_{Qi}, e.g., $CS_{Q5} = \{B, D, E\}$. The main idea of our approach is to extract the prerequisite relationships among concepts of test items and construct the concept map. Based upon assumptions, for each record of learners, each test item has a grade.

As shown in Fig. 1, our Concept Map Construction includes two phases: *Grade Fuzzy Association Rule Mining Process Phase* and *Concept Map Constructing Process Phase*. The first phase applies fuzzy theory, education theory, and data mining approach to find four fuzzy grade association rule types, L–L, L–H, H–H, H–L, among test items. The second phase further analyzes the mined rules based upon our observation in real learning situation. Even based upon our assumptions, constructing a correct concept map is still a hard issue. Accordingly, we propose a heuristic algorithm which can help construct the concept map.

3.1. Grade fuzzy association rule mining process

In (Tsai et al., 2001), the Look Ahead Fuzzy Association Rule Miming Algorithm (LFMAlg) has been used to find the associated relationship information embedded in the testing records of learners. In this phase, we propose an anomaly diagnosis process to improve LFMAlg and reduce the input data before the mining process.

3.1.1. Grade fuzzification

Firstly, because the numeric testing data are hard to analyze by association rule mining approach, we apply Fuzzy Set Theory to transform these into symbolic. Thus, after the fuzzification, the grade on each test item will be labeled as high (H), middle (M), and low (L) degree, which can be used as an objective judgment of learner's performance.

Table 2
Test item–concept mapping table

	A	B	C	D	E
Q_1	0	0	0	1	0
Q_2	1	0	1	0	0
Q_3	1	0	0	0	0
Q_4	0	1	1	0	0
Q_5	0	1	0	1	1

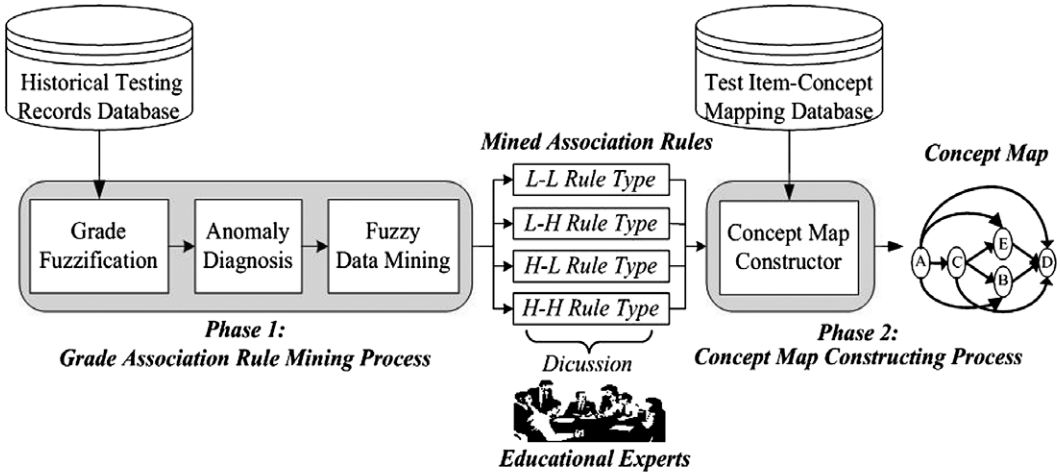

Fig. 1. The flowchart of two-phase concept map construction (TP-CMC).

3.1.2. Anomaly diagnosis

Based upon Item Analysis for Norm-Referencing of Educational Theory (Popham, 1999), the discrimination of item can tell us how good a test item is, i.e., item with high degree of discrimination denotes that the item is well designed. If the discrimination of the test item is too low (most students get high score or low score), this item as redundant data will have no contribution to construct the concept map. For decreasing the redundancy of test data, we propose a fuzzy item analysis, called Anomaly Diagnosis, to refine the test data.

3.1.3. Fuzzy data mining

Then, we can apply LFMAlg (Tsai et al., 2001) to find the grade fuzzy association rules of test items from the historical testing data. In this paper, we analyze the prerequisite relationships among learning concepts of quizzes according to 4 association rule types, L–L, L–H, H–L, H–H, generated from Large 2 Itemset. $Q_i \cdot L$ notation denotes that the *i*th question (Q) was tagged with low (L) degree, e.g., $Q_2 \cdot L \rightarrow Q_3 \cdot L$ means that learners get low grade on Q_2 implies that they may also get low grade on Q_3.

3.2. Concept map constructing process

3.2.1. Concept map constructor

Firstly, the result of analyzing four association rule types, L–L, L–H, H–H, and H–L, are used to construct the prerequisite relationships between concept sets, which are used to define the edge between nodes of concept set and provide teachers with information for further refining the test sheet, of learning concepts of test items. Then, based on the prerequisite relationships of concept set and the Test item-Concept Mapping Table, we propose a Concept Map Constructing (CMC) Algorithm to find the corresponding learning concepts of concept set to construct the concept map according to the join principles of concept-pair.

4. Grade fuzzy association rule mining process

4.1. Grade fuzzification

As described in Section 3.1, we apply fuzzy concept to transform numeric grade data into symbolic, called Grade Fuzzification. Three membership functions of each quiz's grade are shown in Fig. 2. In the fuzzification result, "Low", "Mid" and "High" denote "Low Grade", "Middle Grade" and "High Grade" respectively. $Q_i \cdot L$, $Q_i \cdot M$, and $Q_i \cdot H$ denote the value of LOW fuzzy function, MIDDLE fuzzy function, and HIGH fuzzy function for the quiz i, respectively. By given membership functions, the fuzzification of testing records is described in Example 1.

Example 1. In Fig. 3, assume there are 10 testing records with 5 quizzes of learners and the highest grade on each quiz is 20.

4.2. Anomaly diagnosis

For refining the input testing data, we propose the anomaly diagnosis, called Fuzzy Item Analysis for Norm-Referencing (FIA-NR) by applying Item Analysis for Norm-Referencing of Educational Theory, shown in Fig. 4. A test item will be deleted if it has low discrimination.

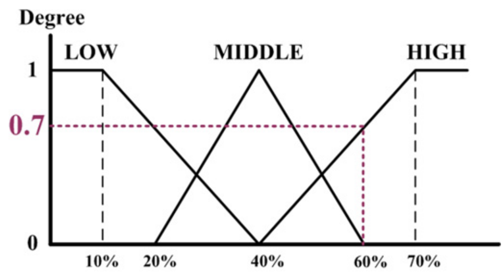

Fig. 2. The given membership functions of each quiz's grade.

Student ID	Q_1	Q_2	Q_3	Q_4	Q_5	Total
1	12	18	20	20	7	77/100
2	12	14	18	3	7	54/100
3	12	16	14	4	7	53/100
4	2	8	12	6	20	48/100
5	2	8	12	2	12	36/100
6	2	10	8	2	20	44/100
7	20	5	5	4	1	35/100
8	10	6	6	1	5	28/100
9	10	5	5	1	5	26/100
10	10	3	4	0	5	21/100

Fuzzification

Student ID	Q_1			Q_2			Q_3			Q_4			Q_5		
	Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High
1	0.0	0.0	0.7	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.2	0.8	0.0
2	0.0	0.0	0.7	0.0	0.0	1.0	0.0	0.0	1.0	0.8	0.0	0.0	0.2	0.8	0.0
3	0.0	0.0	0.7	0.0	0.0	1.0	0.0	0.0	1.0	0.7	0.0	0.0	0.2	0.8	0.0
4	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.7	0.3	0.5	0.0	0.0	0.0	1.0
5	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.7	1.0	0.0	0.0	0.0	0.0	0.7
6	1.0	0.0	0.0	0.0	0.5	0.3	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0
7	0.0	0.0	1.0	0.5	0.3	0.0	0.5	0.3	0.0	0.7	0.0	0.0	1.0	0.0	0.0
8	0.0	0.5	0.3	0.3	0.5	0.0	0.3	0.5	0.0	1.0	0.0	0.0	0.5	0.3	0.0
9	0.0	0.5	0.3	0.5	0.3	0.0	0.5	0.3	0.0	1.0	0.0	0.0	0.5	0.3	0.0
10	0.0	0.5	0.3	0.8	0.0	0.0	0.7	0.0	0.0	1.0	0.0	0.0	0.5	0.3	0.0
Sum	3.0	1.5	4.0	2.1	3.6	3.3	2.0	2.1	4.4	7.5	.5	1.0	3.4	3.3	2.7

Fig. 3. The fuzzification of learners' testing records.

Algorithm: Fuzzy Item Analysis for Norm-Referencing (FIA-NR)
Symbol Definition:
R_{iH} / R_{iL}: The sum of the fuzzy grades (H=1, M=0.5, L=0) on test item i for each learner in the high(H)/Low(L) group.
N_{iH} / N_{iL}: The number of learners in high/low group.
P_{iH} & P_{iL}: the ratios of R_{iH} to N_{iH} and of R_{iL} to N_{iL}, respectively.
Input : Fuzzified testing records of learners
Output : the ***Difficulty index*** (${P_i}$) and the ***Discrimination index*** (${D_i}$) of each test item

Step1: Sort Scores in descending order and divide it into **High**, **Middle**, and **Low** groups, each has 1/3 learners.
Step2: Let $P_{iH} = \frac{R_{iH}}{N_{iH}}$ and $P_{iL} = \frac{R_{iL}}{N_{iL}}$.
Step 3: Compute $P_i = 1 - \frac{P_{iH} + P_{iL}}{2}$ and $D_i = P_{iH} - P_{iL}$, for $i = 1, \ldots, k$.
Step 4: Delete the test items with low Discrimination (<0.5).

Fig. 4. Fuzzy item analysis for norm-referencing (FIA-NR).

Example 2. Table 3 shows the fuzzified testing grades of learners on Q_4 sorted in the descending order of each learner's total score in the test sheet. For example, in Fig. 3, because the result of fuzzification of learner ID 4 is (0.3, 0.5, 0.0), her/his Grade Level can be tagged with *M* by the Max(L, M, H) function.

Then, by applying FIA-NR algorithm, we can get the *Difficulty* and *Discrimination* of every quiz. For example, the $P_{4\mathrm{H}}$ and $P_{4\mathrm{L}}$ of Q_4 are $P_{4\mathrm{H}} = \frac{R_{4\mathrm{H}}}{N_{4\mathrm{H}}} = \frac{\mathrm{H+L+L}}{3} = \frac{1+0+0}{3} = \frac{1}{3}$ and $P_{4\mathrm{L}} = \frac{0}{3} = 0$, respectively. Therefore, its *Difficulty* P_4 and *Discrimination* D_4 are $P_4 = 1 - \frac{P_{4\mathrm{H}} + P_{4\mathrm{L}}}{2} = 1 - \frac{1/3+0}{2} = \frac{5}{6} = 0.83$ and 0.33 respectively. Thus, learners' grade on Q_4 will be deleted because its *Discrimination* is too low to use during the mining process and the construction of the concept map. Accordingly, the test sheet can be redesigned. All evaluated results are shown in Table 4.

Table 3
Sorted fuzzified testing grade on Q_4

Group	High			Middle				Low		
Learner ID	1	2	3	4	6	5	7	8	9	10
Total (100)	77	54	53	48	44	36	35	28	26	21
Grade level = Max(L, M, H)	H	L	L	M	L	L	L	L	L	L

Table 4
Difficulty and discrimination degree of each quiz

	Q_1	Q_2	Q_3	Q_4	Q_5
Difficulty (0 to 1)	0.25	0.42	0.42	0.83	0.75
Discrimination (−1 to 1)	0.5	0.83	0.83	0.33	0.5

4.3. Fuzzy data mining

After filtering out these useless quizzes, we can apply Look Ahead Fuzzy Association Rule Mining Algorithm (Tsai et al., 2001) as shown in Fig. 5 to find the fuzzy association rules of test items. In LFMAlg Algorithm, the support value of every itemset x in candidate C_ℓ can be evaluated by the support(x) function, where $x = \{A, B\} \subseteq C_{\ell-1}$, $A \cap B = \phi$. Then, the support(x) = support $(A \cup B) = \sum_1^n \text{Min}(A, B)$, where n is the number of learners. For example, in Fig. 3, support($Q_1 \cdot L$, $Q_3 \cdot H$) = Min(1.0, 0.7) + Min(1.0, 0.7) = 1.4.

Algorithm: LFMAlg Algorithm
Symbol Definition:
α_ℓ : The minimum support threshold in the ℓ -large itemset.
C_ℓ : The ℓ -Candidate itemset.
L_ℓ : The ℓ -large itemset
λ : The minimum confidence threshold.
Input: The test records of learners after Fuzzification and Anomaly Diagnosis. The minimum support threshold α_1 and λ.
Output : The fuzzy association rules of test records of learners.

Step1: Repeatedly execute this step until C_ℓ = NULL.
1.1: Generate and insert the ℓ -itemset into C_ℓ
1.2: $\alpha_\ell = \max(\frac{\alpha_1}{2}, \ \alpha_{\ell-1} - \frac{\alpha_1}{(\ell-1) \times c})$, where ℓ >1 and c is constant.
1.3: $L_\ell = \{ x \mid \text{support(x)} \geq \alpha_\ell, \quad for \quad x \in C_\ell \}$
1.4: $\ell = \ell + 1$
Step2: Generate the association rules according to the given λ in L_ℓ .

Fig. 5. Look ahead Fuzzy Association Rule Mining Algorithm (LFMAlg).

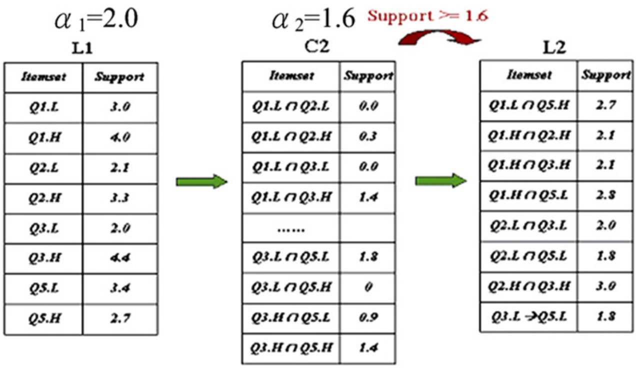

Fig. 6. Mining process of Large 2 Itemset.

Table 5
The mining results ($\text{Conf}_i > 0.8$)

Rule types	Mined rules	Conf_i
L–L	$Q2 \cdot L \rightarrow Q3 \cdot L$	0.95
	$Q3 \cdot L \rightarrow Q2 \cdot L$	1.00
	$Q2 \cdot L \rightarrow Q5 \cdot L$	0.86
	$Q3 \cdot L \rightarrow Q5 \cdot L$	0.90
L–H	$Q1 \cdot L \rightarrow Q5 \cdot H$	0.90
	$Q5 \cdot L \rightarrow Q1 \cdot H$	0.82
H–H	$Q2 \cdot H \rightarrow Q3 \cdot H$	0.91
H–L	$Q5 \cdot H \rightarrow Q1 \cdot L$	1.00

Example 3. For the data shown in Examples 1 and 2, Fig. 6 shows the process of finding the association rules with large 2 itemset by LFMAlg algorithm.

Thus, Table 5 shows the grade fuzzy association rules with minimum confidence 0.8 generated from large 2 itemset into L–L, L–H, H–H, and H–L types. The Conf_i (Confidence) is used to indicate the important degree of *i*th mined association rule. For example, the Confidence (Conf_1) of rule $Q_2 \cdot L \rightarrow Q_3 \cdot L$ can be obtained as follows.

$$Q_2 \cdot L \rightarrow Q_3 \cdot L : \text{Confidence} = \frac{\text{support}(Q_2 \cdot L \cup Q_3 \cdot L)}{\text{support}(Q_2 \cdot L)} = 0.95$$

5. Concept map constructing process

5.1. Concept map constructor

Before constructing the concept map, we can get the prerequisite relationship among concepts of quiz from analyzing four association rule types, L–L, L–H, H–L, and H–H, based upon our observation obtained by interviewing the educational experts, in real learning situation. Therefore, we can conclude the *Heuristic 1*: given two quizzes Q_1 and Q_2, if concepts of Q_1 are the prerequisite of concepts of Q_2, Learner gets low grade on Q_1 implies that s/he may also get low grade on Q_2 or Learner gets high grade on Q_2 implies that her/his grade on Q_1 is high. As shown in Table 6, for each rule type, we use *Heuristic 1* to get its prerequisite relationships among concept

Table 6
Prerequisite relationship of association rule

Rule	W*i*	Prerequisite relationship
$Q_i \cdot L \rightarrow Q_j \cdot L$	1.0	$CS_{Qi} \xrightarrow{\text{pre.}} CS_{Qj}$
$Q_i \cdot L \rightarrow Q_j \cdot H$	0.8	$CS_{Qj} \xrightarrow{\text{pre.}} CS_{Qi}$
$Q_i \cdot H \rightarrow Q_j \cdot H$	1.0	$CS_{Qj} \xrightarrow{\text{pre.}} CS_{Qi}$
$Q_i \cdot H \rightarrow Q_j \cdot L$	0.8	$CS_{Qi} \xrightarrow{\text{pre.}} CS_{Qj}$

sets of quizzes with parameterized possibility weight, which are used to construct the concept map. The definition of the symbols used in Table 6 is described as follows.

Symbol definition:
CS_{Qi} indicate concept set of quiz i
W_i indicate the possibility of the possible scenario of the rule

In this paper, association rules generated from Large 2 Itemset are firstly used to analyze the prerequisite relationships between learning concepts of quizzes. Therefore, by looking up Table 6, we can obtain the prerequisite relationships of concept set of quizzes with the possibility weight (W_i) for each mined rule in Table 5. The possibility W_i is a heuristic parameter of CMC algorithm because it can be modified according to different domains and learners' background. Moreover, the related explanations of the analysis in Table 6 are shown in Table 7. Table 8 shows the result of transforming association rules in Table 5 by analyzing the prerequisite relationships in Table 6.

For example, in Fig. 7, the mined rules, $Q1 \cdot L \rightarrow Q2 \cdot H$ and $Q1 \cdot H \rightarrow Q2 \cdot L$, can be transformed into corresponding prerequisite relationship of concept set, resulting in a confused relation as a cycle between concept sets, called circularity. That is to say, concepts of Q_1 and concepts of Q_2 are prerequisite of each other, which is a conflict in our analysis. Therefore, during creating the concept map, we have to detect whether a cycle exists or not, e.g., $CS_{Q1} \rightarrow CS_{Q2} \rightarrow CS_{Q1}$.

Because each concept set may contain one or more learning concepts, we further define a principle of joining two concept sets and then generate corresponding concept-pair, (C_i, C_j), that is, if $CS_{Q1} = \{\cup_1^n a_i\}$ and $CS_{Q2} = \{\cup_1^m b_j\}$, *the set of concept-pair is* CS_{Q1} JOIN $CS_{Q2} = \{\cup_1^k (a_i, b_j)\}$,

Table 7
The explanations of rule types

Rule	Description of learning scenario
$L \rightarrow L$	If the association rule $Q_i \cdot L \rightarrow Q_j \cdot L$ is mined, it means that the CS_{Qi} is the prerequisite of CS_{Qj}, represented as $CS_{Qi} \overset{pre.}{\rightarrow} CS_{Qj}$. That is why getting low grade on Q_i might imply getting low grade on Q_j.
$H \rightarrow H$	If the association rule $Q_i \cdot H \rightarrow Q_j \cdot H$ is mined, it means that the CS_{Qi} is the prerequisite of CS_{Qj}.
$L \rightarrow H$	If the association rule $Q_i \cdot L \rightarrow Q_j \cdot H$ is mined, it means that the CS_{Qj} is the prerequisite of CS_{Qi} because CS_{Qi} may be not learned well resulting from CS_{Qj}.
$H \rightarrow L$	If the association rule $Q_i \cdot H \rightarrow Q_j \cdot L$ is mined, it means that the CS_{Qi} is the prerequisite of CS_{Qj}.

Table 8
Result by analyzing the prerequisite relationships in Table 6

Rule type	Association rules of quiz	Prerequisite relationship of concept set	$Conf_i$	W_i
L–L	$Q2 \cdot L \rightarrow Q3 \cdot L$	$CS_{Q2} \overset{pre.}{\rightarrow} CS_{Q3}$	0.95	1.0
	$Q3 \cdot L \rightarrow Q2 \cdot L$	$CS_{Q3} \overset{pre.}{\rightarrow} CS_{Q2}$	1.00	1.0
	$Q2 \cdot L \rightarrow Q5 \cdot L$	$CS_{Q2} \overset{pre.}{\rightarrow} CS_{Q5}$	0.86	1.0
	$Q3 \cdot L \rightarrow Q5 \cdot L$	$CS_{Q3} \overset{pre.}{\rightarrow} CS_{Q5}$	0.90	1.0
L–H	$Q1 \cdot L \rightarrow Q5 \cdot H$	$CS_{Q5} \overset{pre.}{\rightarrow} CS_{Q1}$	0.90	0.8
	$Q5 \cdot L \rightarrow Q1 \cdot H$	$CS_{Q1} \overset{pre.}{\rightarrow} CS_{Q5}$	0.82	0.8
H–H	$Q2 \cdot H \rightarrow Q3 \cdot H$	$CS_{Q2} \overset{pre.}{\rightarrow} CS_{Q3}$	0.91	1.0
H–L	$Q5 \cdot H \rightarrow Q1 \cdot L$	$CS_{Q5} \overset{pre.}{\rightarrow} CS_{Q1}$	1.00	0.8

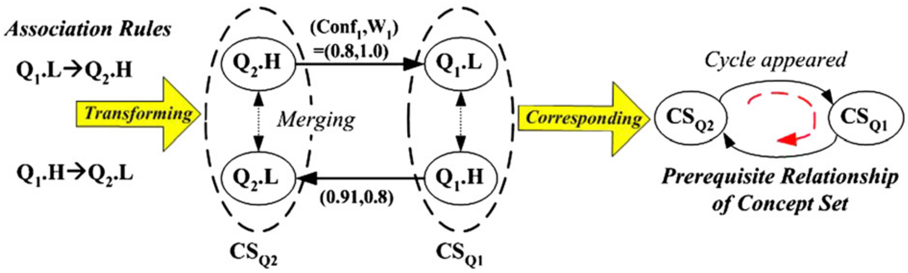

Fig. 7. The transforming of association rules.

where $a_i \neq b_j$ and $k \leqq n \times m$. For example, if $CS_{Q1} = \{a_1, a_2\}$ and $CS_{Q2} = \{b_1, b_2\}$, CS_{Q1} JOIN $CS_{Q2} = \{(a_1, b_1), (a_1, b_2), (a_2, b_1)\}$, where $a_2 = b_2$ is deleted. The related definition used in creating the concept map is given as follows:

Concept Map CM = (V, E), where

- $V = \{C_i |$ the node is unique for each $i\}$
- $E = \{\overrightarrow{C_iC_j} \mid i \neq j\}$

The node, C_i, denotes the learning concept and the edge, $\overrightarrow{C_iC_j}$, which connects C_i and C_j, denotes that C_i is the prerequisite of C_j. The $\overrightarrow{C_iC_j}$ has an *Influence Weight*, IW_k, denotes the degree

Algorithm : Concept Map Constructing (CMC) Algorithm
Input : Association rules of quiz with its grade level
Output : Concept Map

Step 1 : ***CM*** = {*Null*}.
Step 2 : For each association rule
  2.1: Insert CS_{Qi} for Q_i.L or Q_i.H and CS_{Qj} for Q_j.L or Q_j.H.
  2.2: Create a new *prerequisite edge* with confidence (*Conf$_i$*) and weight (*W$_i$*) of rule to connect the CS_{Qi} and CS_{Qj} according to Table 6.
Step 3 : In *CM*, repeatedly execute this step until no cycle is found.
  3.1: Do *Cycle Detection Process*.
  3.2: If a cycle is found
   Then Remove the edge with the lowest confidence.
Step 4 : Delete all independent nodes without connected edges.
Step 5 : For each node CS_{Qi} in *CM*
  { Insert node's corresponding learning concepts according to ***Test item–Concept Table*** }
Step 6 : For each edge $\overline{CS_{Qi}CS_{Qj}}$ in *CM*, perform the following substeps.
  Step 6.1: Use **JOIN** principle to join two connected concept sets for generating the concept-pair, (C_i, C_j) which replaces the original concept sets.
  Step 6.2: Re-compute the Influence Weight IW_k of the *kth* concept-pair (C_i,C_j) according to weight (W_k) and confidence (*Conf$_k$*) of $\overline{CS_{Qi}CS_{Qj}}$.
Step 4: Output the ***CM***.

Fig. 8. Concept Map Constructing (CMC) Algorithm.

of relationship between learning concepts. The formulation of IW_k is $((k-1)\times IW_{k-1}+W_k\times Conf_k)/k$, $1\leqslant k\leqslant n$, where n is the amount of $\overrightarrow{C_iC_j}$.

The proposed Concept Map Constructing (CMC) algorithm is shown in Fig. 8.

For the CMC algorithm shown in Fig. 8, the main purpose of *Cycle Detection Process* is to detect the unreasonable prerequisite relationship as a cycle among concept sets. It should be noted that the prerequisite relationship in the concept set map also fulfills the indicator $f_{1\bar{2}} > f_{\bar{1}2}$ in Table 9, which is an extension of Appleby et al. (1997) after cycle detection. The indicator denotes that if concepts of Q_1 are prerequisite of concepts of Q_2, it is reasonable that $f_{1\bar{2}} > f_{\bar{1}2}$, where $f_{1\bar{2}} = \text{Count}(Q_1\cdot H\cap Q_2\cdot L)$ and $f_{\bar{1}2} = \text{Count}(Q_1\cdot L\cap Q_2\cdot H)$. In addition, the Influence Weight, IW_k, denotes the degree how the learning status of concept C_i influences C_j. Therefore, the number of $\overrightarrow{C_iC_j}$ will enhance the value of Influence Weight. In the formulation of influence weight, the W_i denotes the possibility of the learning scenario of the association rule in our analysis. Thus, the educational experts can assign different value of W_i to the algorithm according to different domains and learner's backgrounds.

For the association rules given in Table 8, the process of CMC algorithm is shown in Fig. 9. In Fig. 9b, the edges drawn as dash line have the lowest confidences in cycles will be deleted in *Cycle*

Table 9
Relative quizzes frequency

	(Q_1) Higher	(Q_1) Lower
(Q_2) Higher	f_{12}	$f_{\bar{1}2}$
(Q_2) Lower	$f_{1\bar{2}}$	$f_{\bar{1}\bar{2}}$

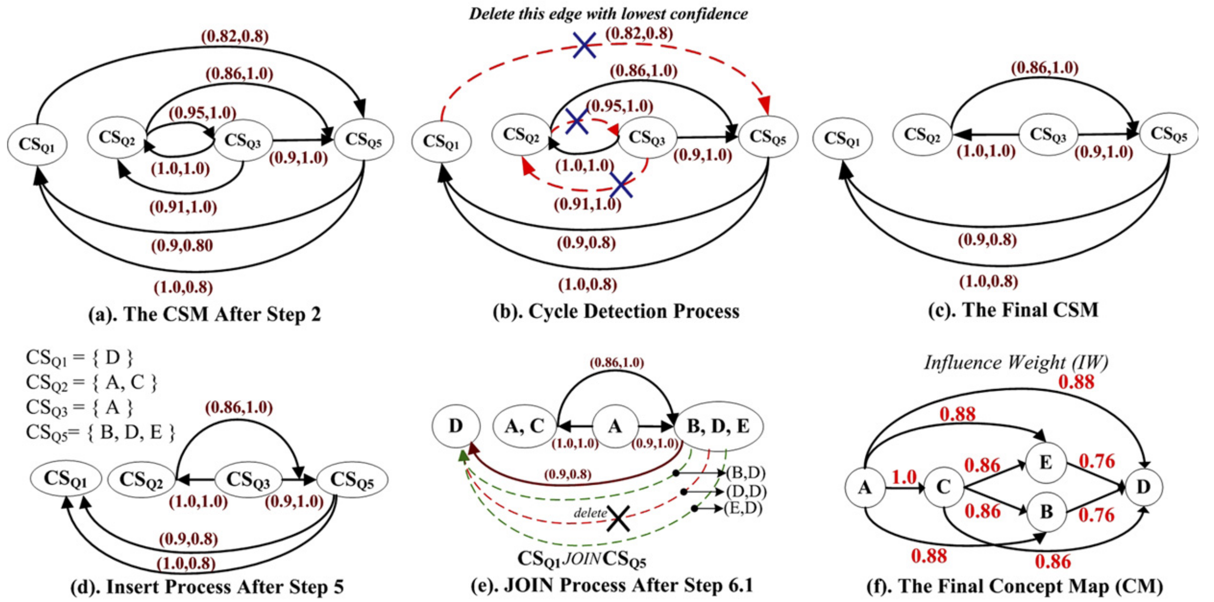

Fig. 9. The process of Concept Map Constructing Algorithm.

Table 10
The result of computing the influence weight of concept-pair (B, D) in Fig. 9f

Rule	Prerequisite relationship	$Conf_i$	W_i	IW_i
$Q_1 \cdot L \to Q_5 \cdot H$	$CS_{Q5} \to CS_{Q1}$	0.90	0.8	$W_1 \times Conf_1 = 0.9 * 0.80 \cong 0.72$
$Q_5 \cdot H \to Q_1 \cdot L$	$CS_{Q5} \to CS_{Q1}$	1.00	0.8	$\frac{(2-1)\times IW_1 + W_2 \times Conf_2}{n} = \frac{(1)\times 0.72+(0.8)\times 1.00}{2} \cong 0.76$

Detection Process. Moreover, Table 10 shows the example of computing the Influence Weight of Concept-Pair (B, E) in Fig. 9f. Because the Concept-Pair (B, E) has two edges between CS_{Q5} and CS_{Q1}, we have to compute the Influence Weight twice.

6. Evaluating the redundancy and circularity of concept map

In this paper, creating a concept map without *Redundancy* and *Circularity* is our concern. As shown in Fig. 10, we create three concept maps by using different approaches and evaluate their

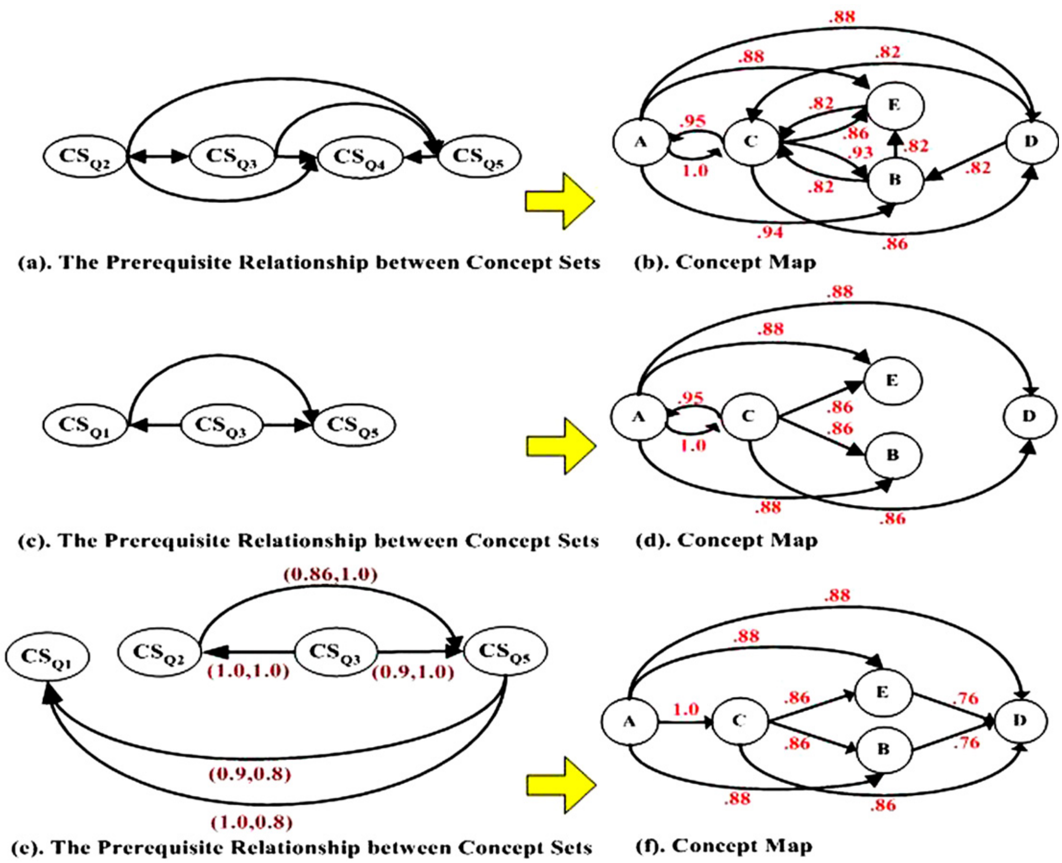

Fig. 10. The (a) and (b) created based upon analyzing L–L rule type only. The (c) and (d) are created based upon Anomaly Diagnosis and analyzing L–L rule type only. The (e) and (f) created by our approach.

difference in terms of *Redundancy* and *Circularity*. Thus, we use three processing steps including *anomaly diagnosis*, the prerequisite relationship based upon analyzing L–L or L–L, L–H, H–L, H–H rule types, and *cycle detection* to create different concept maps. As shown in Fig. 10, the prerequisite relationship between concept sets in Fig. 10a is created based upon analyzing L–L rule type only, and Fig. 10c is created based upon analyzing L–L rule type and anomaly diagnosis we proposed. Then, the concept maps as Fig. 10b and d are transformed according to the **Test Item–Concept Mapping Table**. Fig. 10e and f are created by our proposed approach.

Based upon these results of different approaches, the characteristics of approach are concluded as follows.

- *Non-redundancy*: the *anomaly diagnosis* can filter many useless test items with low discrimination for refining the input data. For example, in Fig. 10a, the Q_4 with low discrimination results in generating many co-prerequisite links as a cycle in Fig. 10b.
- *Non-circularity*: the *cycle detection process* can delete these cycles, e.g., the cycle between *A* and *C* in Fig. 10d, to make the concept map un-ambiguous. Moreover, analyzing association rule with L–L, L–H, H–L, and H–H types can refine the concept map, e.g., the edges $\overrightarrow{ED}$ and $\overrightarrow{BD}$ connect the node D only in Fig. 10f.

7. The experiment of TP-CMC in physics course

In this section, we describe our experiment results of the Two-Phase Concept Map Construction (TP-CMC) approach.

7.1. Experimental results

The participants of experiment are the 104 students of junior high school in Taiwan and the domain of examination is the Physics course. The related statistics of testing results and related concepts of testing paper are shown in Tables 11 and 12.

The prototype system of TP-CMC is developed based on PHP4 web language, MySQL database, and JGraph web graphic tool (JGraph, 2004). As shown in Fig. 11a–c, the concept maps with Discrimination 0.0 and 0.3, and 0.5 are created by TP-CMC approach respectively. As mentioned in Section 4.2, Anomaly Diagnosis process in TP-CMC can refine the test data for decreasing its redundancy. As we see, the concept maps with low discrimination criteria in Fig. 11a and b shows

Table 11
The related statistics of testing results in physics course

Subject	Information
Educational degree	Junior high school
The number of students	104
Average score of exam	61.06
Standard deviation of scores	18.2
The number of test items	50
The number of concepts	17

Table 12
Concepts list of testing paper in physics course

Concept ID	Learning concept
1	Tools and theories for timing
2	Unit of time
3	Isochronism of pendulum
4	Change of position
5	Movements
6	Speed and direction of motion
7	Average and instant speed
8	*X–t* diagram
9	Change of speed and direction
10	Acceleration
11	Uniform acceleration
12	Free fall
13	*V–t* diagram
14	The resultant of forces
15	Balance of forces
16	Torque
17	Balance of rotation

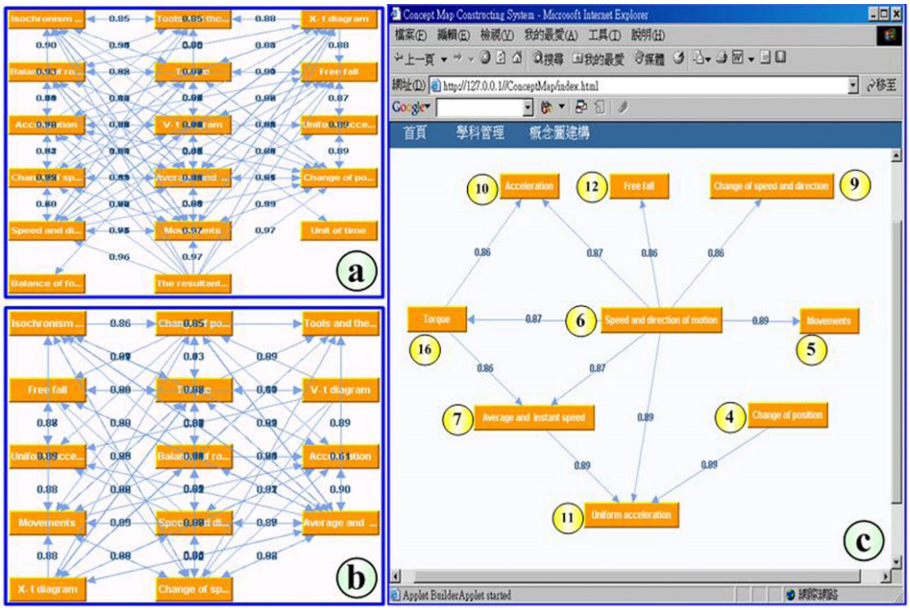

Fig. 11. The concept maps (a)–(c) with Discrimination 0.0, 0.3, and 0.5 are created by TP-CMC approach respectively (support = 50, confidence = 0.85).

that the prerequisite relationships between learning concepts are very disordered and confused. However, with increasing the value of discrimination, the test data can be refined such that the clarity of concept map can be heightened, shown in Fig. 11c. Moreover, the created concept map can provide the embedded learning information of students during learning Physics. For example, the relationship of concept-pair (6, 9) in Fig. 11c represents that if students do not learn concept *6* (*Speed and direction of motion*) well, their learning performance of concept *9* (*Change of speed and direction*) are most likely bad. Therefore, teachers can modify their teaching strategies to enhance students' learning performance of concept 6 for getting high performance of concept 9.

8. Conclusion

The concept map is often used to provide teachers for further analyzing and refining the teaching strategies and to generate adaptive learning guidance in adaptive learning environment. However, creating the concept map of a course is difficult and time consuming. Therefore, in this paper, we propose a Two-Phase Concept Map Construction (TP-CMC) approach to automatically construct a concept map of a course by learners' historical testing records. Phase 1 is used to preprocess the testing records and Phase 2 is used to transform the mined association rules into prerequisite relationships between learning concepts for creating concept map. Thus, in Phase 1, we apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic data, Education Theory (Item Analysis for Norm-Referencing) to further refine it, and Data Mining approach to find its grade fuzzy association rules. In Phase 2, based upon our observation in real learning situation, we use multiple rule types to further analyze the mined association rules and then propose a heuristic algorithm to automatically construct the concept map without Redundancy and Circularity according to analysis results. Thus, the created concept map which can be used to develop adaptive learning system and refine the learning strategies of learners. Moreover, we also develop a prototype system of TP-CMC and then use the real testing records of students in junior high school to evaluate the results. The experimental results show that our proposed approach is workable. In the near future, we will analyze the effect of rules with large-3 itemset for improving the concept map, enhance the TP-CMC system with scalability and flexibility for providing the web service, and do some experiments based upon real learning testing records, too.

Acknowledgement

This research was partially supported by National Science Council of Republic of China under the number of NSC94-2524-S009-001, NSC94-2524-S009-002, and NSC 93-2524-S-009-004-EC3.

References

Appleby, J., Samuels, P., & Jones, T. T. (1997). Diagnosis – a knowledge-based diagnostic test of basic mathematical skills. *Computers and Education, 28*(2), 113–131.

Bruillard, E., & Baron, G. L. (2000). Computer-based concept mapping: a review of a cognitive tools for students. In: *Proceedings of the International Conference on Educational Uses of Communication and Information Technologies, Beijing, China* (pp. 331–338).

Carchiolo, V., Longheu, A., & Malgeri, M. (2002). Adaptive formative paths in a web-based learning environment. *Educational Technology and Society, 5*(4), 64–75.

Chang, K. E., Liu, S. H., & Chen, S. W. (1998). A testing system for diagnosing misconceptions in dc electric circuits. *Computers and Education, 31*(2), 195–210.

Frosini, G., Lazzerini, B., & Marcelloni, F. (1998). Perform automatic exams. *Computers and Education, 31*(3), 281–300.

Gaines, B. R., & Shaw, M. L. G. (1995). Concept maps as hypermedia components. *International Journal of Human–Computer Studies, 43*(3), 323–361.

Gamboa, H. (2001). Designing intelligent tutoring systems: a bayesian approach. In: *Proceedings of the Ana Fred 3rd International Conference on Enterprise Information Systems (ICEIS′2001)* (pp. 452–458).

Gordon, J. L. (2000). Creating knowledge maps by exploiting dependent relationships. *Knowledge Based System, 13*, 71–79.

Hsu, C. S., Tu, S. F., & Hwang, G. J. (1998). A concept inheritance method for learning diagnosis of a network-based testing and evaluation system. In: *Proceedings of the 7th International Conference on Computer-Assisted Instructions* (pp. 602–609).

Hwang, G. J. (2003). A conceptual map model for developing intelligent tutoring system. *Computers and Education, 40*(3), 217–235.

Hwang, G. J., Hsiao, C. L., & Tseng, C. R. (2003). A computer-assisted approach to diagnosing student learning problem in science course. *Journal of Information Science and Engineering, 19*(2), 229–248.

JGraph (2004), Open-Source Graph Component Available for Java, Retrieved April 21, 2004 from http://www.jgraph.com/.

Novak, J. D. (1998). *Learning, creating, and using knowledge: concept maps as facilitative tools in schools and corporations*. Mawah, NJ: Lawrence Erlbaum and Associates.

Popham, W. J. (1999). *Classroom assessment: what teachers need to know*. Pearson Allyn & Bacon, 222–227.

Triantafllou, E., Pomportsis, A., & Demetriadis, S. (2003). The design and the formative evaluation of an adaptive educational system based on cognitive styles. *Computers and Education, 41*(1), 87–103.

Tsai, C. J., Tseng, S. S., & Lin, C. Y. (2001). A two-phase fuzzy mining and learning algorithm for adaptive learning environment. In: *Proceedings of the International Conference on Computational Science (ICCS′01), Lecture Notes in Computer Science (LNCS 2074), Vol. 2* (pp. 429–438). CA, USA.