國立交通大學

資訊科學系

碩 士 論 文

概念圖建構方法之研究

A New Approach for Constructing the Concept Map

研 究 生：蘇培綺

指導教授：曾憲雄 博士

中 華 民 國 九 十 三 年 六 月

概念圖建構方法之研究

A New Approach for Constructing the Concept Map

研 究 生：蘇培綺　　　　Student：Pei-Chi Sue

指導教授：曾憲雄　　　　Advisor：Shian-Shyong Tseng

國 立 交 通 大 學
資 訊 科 學 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Computer and Information Science
College of Electrical Engineering and Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer and Information Science

June 2004

Hsinchu, Taiwan, Republic of China

中華民國九十三年六月

概念圖建構之研究

研究生: 蘇培綺　　　　　　　　　　　　　　　　　　　指導教授: 曾憲雄博士

國立交通大學資訊科學系

摘要

隨著網路學習(e-Learning)的蓬勃與發展，如何在「適性化」理念下發展出智慧型網路學習系統，並依照不同學習能力與背景的學習者來提供適性化的學習路徑已是當今網路學習不可忽視的重要課題。為了更容易達到適性化教學的目的，大部分的網路學習系統會將學生的學習成績加以分析，並依據事先建構好的「課程概念圖」來調整適合學生的學習路徑。 然而，雖然概念圖在適性化的網路學習相關策略設計上非常有用，可是，每當學習一新課程，教育設計者或是相關領域專家就必須進行一段冗長、費時且艱鉅的知識擷取過程才能將此概念圖建立起來。

為了解決這個問題，本文研究提出了「二階段概念圖建構法」，希望達到自動化建立概念圖的目的。在第一階段裡，結合了泛析理論、教育理論、資料探勘等相關技術，藉此找出學生成績間的相關法則 ;在第二階段裡，深入分析成績相關法則所代表的意義，並判斷試題所包含概念之間的先備關係及其可能的相對應的情境解釋。最後，根據之前所做的相關分析，量身訂作了一個「概念圖建構演算法」，藉由此演算法來自動建構出課程概念圖，以利教師或專家進一步分析及應用。

A New Approach for Constructing the Concept Map

Student: Pei-Chi Sue Advisor: Dr. Shian-Shyong Tseng

Department of Computer and Information Science

National Chiao Tung University

Abstract

In recent years, e-learning system has become 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 thesis, 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 form, 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 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 feasible.

誌謝

能夠順利完成這篇論文，首先要感謝的是我的指導教授，曾憲雄博士。曾老師兩年來不論是專業上的知識或是領導處事的技巧，都讓我獲益良多。尤其是在二年級時不斷的給予我鼓勵及耐心的指導，讓我可以在畢業前將論文投至國際會議上發表，深表感激。同時也非常感謝我的論文口試委員，孫春在教授，楊錦潭教授及黃國禎教授，他們給予了我許多寶貴的建議。

接下來要感謝實驗室各位學長，尤其是蘇俊銘學長和翁瑞鋒學長。他們給予我這個題目最初的靈感及想法，並且接受我隨時隨地的詢問及討論，在論文的的修改上給予了我很大的協助，而在課業與待人處世上，也讓我受益匪淺，在此深表感激。另外，也非常感謝同窗伙伴們互相幫助鼓勵的情誼，以及學弟妹們在友情上的支持，讓忙碌的碩士班兩年生涯，能夠愉快而充實的度過。

最重要的，還要感謝我的家人長久以來對我的支持與照顧，在我成長及求學的過程中陪伴我一路走來，是鼓勵我努力向前的最大力量。

最後要感謝的是現在正在閱讀我的論文的您。希望這篇論文能對您有所幫助。有任何問題或建議也都歡迎與我討論，感激不盡。

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

With the vigorous development of computer and WWW technology, computer and web based learning environments are becoming mainstream area of research and development. There are three main paradigms in computer based tutoring systems [15]: Computer Assisted Instruction (CAI) systems, hypermedia systems, and Intelligent Tutoring Systems (ITS).

Computer Assisted Instruction (CAI) systems are usually based on a fixed presentation of didactic material and do not support adaptive tutoring for each individual student. The learning path through the learning material is linear and predefined in the design stage.

Hypermedia systems provide user-driven exploration of the teaching material, where the user has full control over the learning process. However, this is also their drawback, because Hypermedia systems lack expert guidance in the instructional sequence and therefore it is difficult for users to find information and get an overview of the material. Finally, the users may lose focuses on their educational goals.

Intelligent tutoring systems (ITS) are system-driven learning systems, which were developed to adapt the learning speed and level of material for each individual student. The definition of an ITS is based on some kind of knowledge. This "knowledge" includes: (1) domain knowledge containing objects, relations among them, explanations, examples and exercises, (2) teachers' knowledge as a strategy for the process of learning and teaching and (3) students' knowledge as a model which is dynamically generated as a result of overlaying it with teachers' knowledge.

In recent years, many intelligent adaptive learning and testing systems [3][6][7][9][10][12][14][19][20][23] have been proposed to offer learners customized courses in accordance with their aptitudes and learning results. For achieving the adaptive learning, in many learning environments [3][6][9][10][14][23], a predefined concept map of a course, as domain knowledge in ITS, which provides teachers for further analyzing and refining the teaching strategies, is often used to generate adaptive learning guidance for learners.

Though concept map as a navigation tool in adaptive learning system is widespread, 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 thesis, we propose a Two-Phase Concept Map Construction (TP-CMC) algorithm to automatically construct a concept map of a course by 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 form input data, and Phase 2 is used to transform the mined association rules into prerequisite relationships between learning concepts for creating concept map. Therefore, in the first phase, we apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic form, 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 $Q_1.L \rightarrow Q_2.L$ 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 [12][23]. 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 thesis are to:

(1) Apply Fuzzy Set Theory to transform the numeric testing records of learners into symbolic form, 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.

Chapter 2. Related Work

2.1 Concept Map

Concept map, developed by Novak [16] in 1984, is a technique for organizing or representing knowledge as networks. Networks consist of nodes (points/vertices) and links (arcs/edges). Nodes represent concepts and links represent the relations among concepts. Links can be non-, uni- or bi-directional. According to Jonassen et al. [7], concept maps are “representations of concepts and their interrelationships that are intended to represent the knowledge structures that humans store in their minds”.

Concept map has been widely applied in the evaluation of students’ learning in the school system, policy studies, and the philosophy of science to provide a visual representation of knowledge structures. In many disciplines various forms of concept map are already used as formal knowledge representation systems, for example: semantic networks in artificial intelligence, bond graphs in mechanical and electrical engineering, CPM and PERT charts in operations research, Petri nets in communications, and category graphs in mathematics.

An example of a concept map is shown in Figure 1. In this example, each node is a learning concept, and each uni-directional link denotes relationship “prerequisite of “.

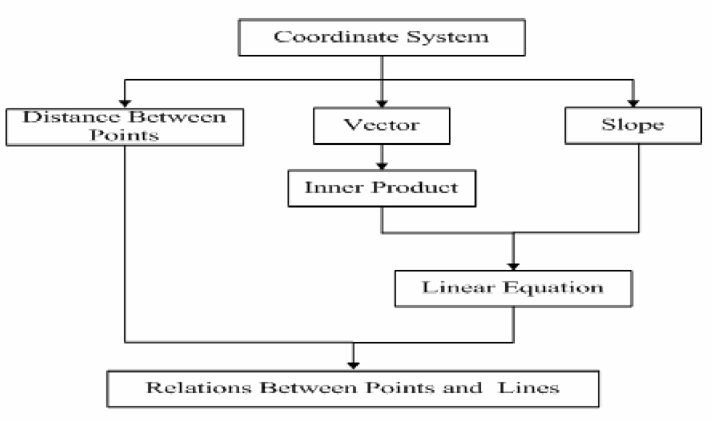

Figure 1. An example of concept map

Concept maps are very useful in that they can be considered from different points of views and a wide variety of different forms of concept map have been defined and applied in various domains and can be done for several purposes [4][10][20]. The functions of concept map are shown as follows.

(1) Knowledge representation (to design a complex structure, etc.)

(2) Knowledge acquisition/communication (to communicate complex ideas, etc.)

(3) Meaningful learning for education (to help brain storming in mapping process, etc.)

(4) Navigation tools in adaptive learning system (to find learning paths by assessment and diagnosis for understanding of learning concept)

2.2 Uses of Concept Maps as Navigational Tools in Education

As described in Section 2.1, the concept map has many potential roles in education on hypermedia learning. They can be used to help designers in designing hypermedia or, as navigational tools, for helping learners to find an appropriate learning path. After the test results of students were analyzed based on concept maps, the students were given guidance on concepts needing improvement to enhance their learning performance. To model the learning effect relationships among concepts, a conceptual map-based notation, Concept Effect Relationships (CER), is proposed by Hsu [14]. Consider two concepts, C_i and C_j, if C_i is prerequisite to the more complex and higher level concept C_j, then a concept effect relationship $C_i \rightarrow C_j$ exists. A single concept may have multiple prerequisite concepts and can also be a prerequisite concept of multiple concepts. In the computer-assisted instructional environment [3][4][6][9][10][14][23], for achieving the adaptability of learning, the predefined CER-like concept map of the course is often 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 to improve their learning performances.

2.3 Construction of Concept Map

Since the use of concept map as education tools in the hypermedia is widespread, the eliciting of concept maps from domain expert or experienced teachers becomes very important. However, the job for constructing the concept maps is still hard and very time consuming.

Therefore, due to the usefulness of the concept map, many approaches are proposed to construct the concept map. The construction of concept maps can be generally classified into manual [2], semi-automatic [21] and nearly automatic [13][17] three categories. In semi-automatic or nearly automatic construction of concept maps, the technique of extracting predicates from a text file or dialog using syntactic and discourse knowledge is often used. Moreover, the browsing behavior and testing records of learners can even be analyzed to construct the concept map.

The following approaches proposed to construct prerequisite relationships among learning concepts of the concept maps are built by analyzing the testing records of students.

Appleby, et al. [3] 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

Table 1. Relative Skills Frequency

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

Later, based upon statistical prediction and approach of Hsu, et al. [14], a CER Builder was proposed by Hwang [12]. Firstly, CER Builder finds the test item that most students failed to answer correctly and then collects the other test items which were 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 relationships among concepts of test items, which may decrease the quality of concept map.

Tsai, et al. [23] proposed a Two-Phase Fuzzy Mining and Learning Algorithm. In the first phase, **L**ook 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 is applied to find the misconcept map indicating the missing concepts during students learning. The obtained misconcept map as recommendation can be fed 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.

Thus, in this thesis, we propose a Two Phase Concept Map Construction (TP-CMC) to construct the complete concept map with influence weights among learning concepts of a course. For improving [23], we apply anomaly diagnosis process to reduce the noise rules and then we take multiple rule types into account to further analyze the mined rules for refining the quality of concept map. Therefore, according to the analysis results, we propose an algorithm to automatically construct the concept map of a course.

Chapter 3. Two Phase Concept Map Construction (TP-CMC)

As mentioned above, the concept map of a course is quite useful. However, the construction is time consuming. Therefore, in this thesis, we propose an approach to automatically construct the concept map as a directional graph with influence weights among learning concepts of a course.

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.

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

As shown in Figure 2, 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.

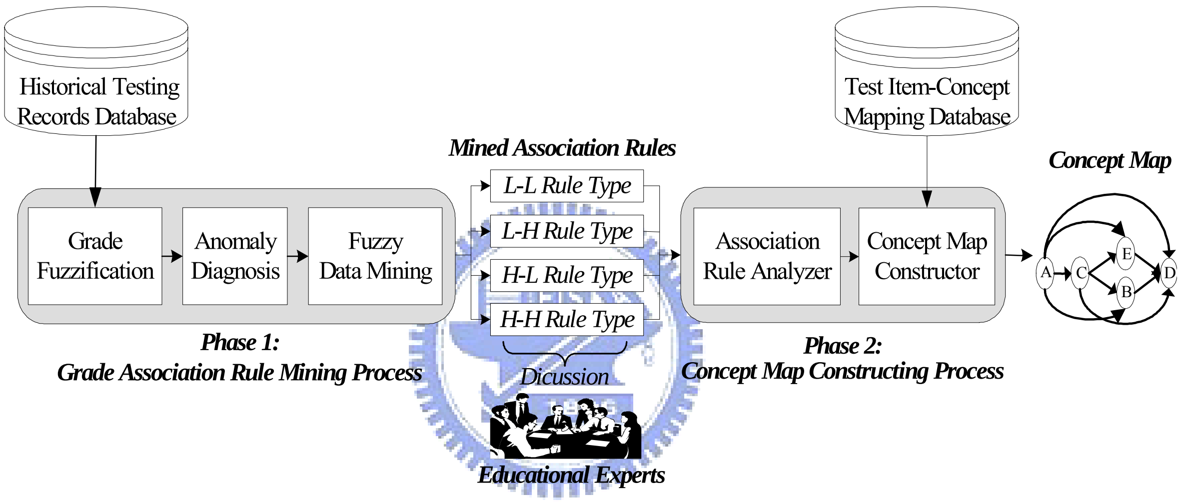

Figure 2. The Flowchart of Two Phase Concept Map Construction (TP-CMC)

3.1 Grade Fuzzy Association Rule Mining Process

In [23], the Look Ahead Fuzzy Association Rule Miming Algorithm (LFMAlg) can be used to find the associated relationship information embedded in the testing records of learners. In the first phase, we propose an anomaly diagnosis process, a preprocessing, to improve LFMAlg and reduce the input data before the mining process.

In the following, three steps of Phase 1 shown in Figure 2 will be briefly described.

(1) Grade Fuzzification

Firstly, we apply Fuzzy Set Theory to transform these numeric testing data into symbolic form. 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. Then, the association mining approach can be used to find the association rule among these testing items.

(2) Anomaly Diagnosis

Based upon Item Analysis for Norm-Referencing of Educational Theory [1][17], 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 using difficulty and discrimination of test item, called Anomaly Diagnosis, to refine the test data.

(3) Fuzzy Data Mining

Then, we apply LFMAlg [23] to find the grade fuzzy association rules of quizzes from the historical testing data. In this thesis, we analyze the prerequisite relationships among learning concepts of quizzes according to 4 association rule types, ***L-L, L-H, H-L, H-H***. We use Q_i.L notation to denote that the *i*th question (Q) was tagged with low (L) degree, e.g., $Q_2.L \rightarrow Q_3.L$ 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

In the second phase, based upon the heuristic of our observation in real learning situation, we further analyze the mined rules. According to the analysis result, we propose a heuristic algorithm to construct the concept map of a course.

In the following, the process of Phase 2 shown in Figure 2 will be briefly described.

(1) Association Rule Analyzer

Firstly, we analyze the four association rule types, L-L, L-H, H-H, and H-L, to generate related prerequisite relationships among concept sets of test item based on our observation in real learning situation. The result of analysis is used to define the edge between nodes of concept set and provide teachers with the possible learning scenario of students for further refining the test sheet.

(2) Concept Map Constructor

Then, based on the prerequisite relationships of concept sets described above 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 mapping.

Chapter 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 Figure 3. In the fuzzification result, "Low", "Mid" and "High" denotes "Low Grade", "Middle Grade" and "High Grade" respectively. $Q_i.L$ denotes the value of LOW fuzzy function, $Q_i.M$ denotes the value of MIDDLE fuzzy function i, and $Q_i.H$ denotes the value of HIGH fuzzy function for the quiz i. By given membership functions, the fuzzification of testing records is described in Example 1.

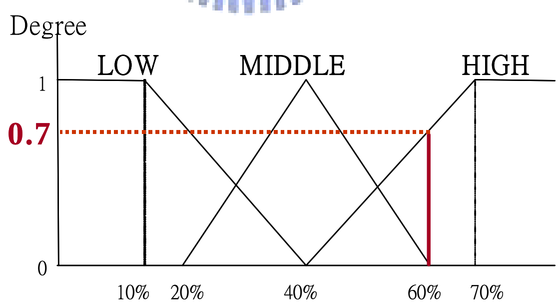

Figure 3. The Given Membership Functions of Each Quiz's Grade

Example 1:

In Figure 4, assume there are 10 testing records with 5 quizzes of learners and the highest grade on each quiz is 20.

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

Figure 4. The Fuzzification of Learners' Testing Records

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 Figure 5. By the anomaly diagnosis process, a test item will be deleted if it has low discrimination.

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* of each student in the high(H)/Low(L) group.

***N_{iH} /N_{iL}*:**The number of learners in high/low group.

PiH* & *PiL*:**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 $P_{iL} = \frac{R_{iL}}{N_{iL}}$.

Step3: 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)**.

Figure 5. 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 Figure 4, 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.

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

Then, by applying FIA-NR algorithm, we can get the *Difficulty* and *Discrimination* of every quiz. For example, the P_{4H} and P_{4L} of Q_4 are $P_{4H} = \frac{R_{4H}}{N_{4H}} = \frac{H+L+L}{3} = \frac{1+0+0}{3} = \frac{1}{3}$ and $P_{4L} = \frac{0}{3} = 0$. Therefore, its *Difficulty* P_4 and *Discrimination* D_4 are $P_4 = 1 - \frac{P_{4H} + P_{4L}}{2} = 1 - \frac{1/3+0}{2} = \frac{5}{6} = 0.83$ 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 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 [23] as shown in Figure 6 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*($A \cup B$) = $\sum_1^n Min(A, B)$, where n is the number of learners. For example, in Figure 4, *support*(Q_1.L, Q_3.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 $C_\ell = NULL$.

1.1: Generate and insert the ℓ-itemset into C_ℓ

1.2: $\alpha_\ell = \max(\frac{\alpha_{\ell-1}}{2}, \alpha_{\ell-1} - \frac{\alpha_{\ell-1}}{(\ell-1)\times c})$, where $\ell > 1$ 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 .

Figure 6. Look ahead Fuzzy Association Rule Mining Algorithm (LFMAlg)

Example 3:

For the data shown in Examples 1 and 2, Figure 7 shows the process of finding the association rules by LFMAlg algorithm.

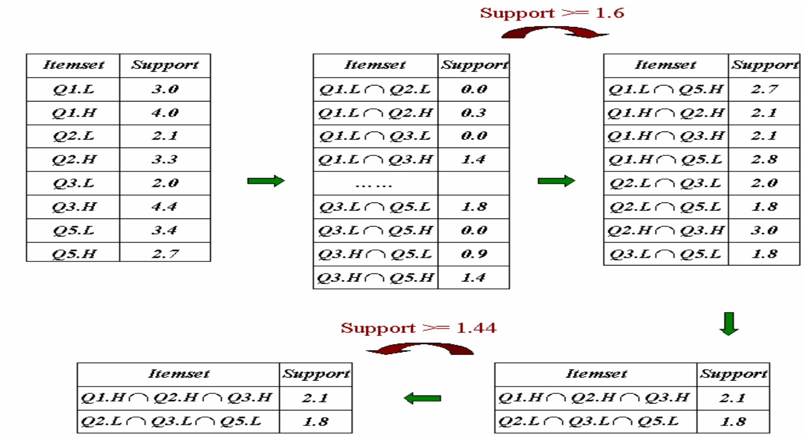

Figure 7. Mining Process of LFMAlg algorithm

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 $Q_2.L \rightarrow Q_3.L$ can be obtained as follows.

$$Q_2.L \rightarrow Q_3.L : Confidence = \frac{suuport(Q_2.L \cap Q_3.L)}{suuport(Q_2.L)} = 0.95$$

Table 5. The Mining Results (***Confi*** > 0.8)

Large 2 Itemset		
Rule Types	***Mined Rules***	***Confi***
L-L	$Q_2.L \rightarrow Q_3.L$	0.95
	$Q_3.L \rightarrow Q_2.L$	1.00
	$Q_2.L \rightarrow Q_5.L$	0.86
	$Q_3.L \rightarrow Q_5.L$	0.90
L-H	$Q_1.L \rightarrow Q_5.H$	0.90
	$Q_5.L \rightarrow Q_1.H$	0.82
H-H	$Q_2.H \rightarrow Q_3.H$	0.91
H-L	$Q_5.H \rightarrow Q_1.L$	1.00

Chapter 5. Concept Map Constructing Process

In Phase 2, the rules mined in Phase 1 will be analyzed based upon the heuristics of our observation in real learning situation. Accordingly, we propose a heuristic algorithm to automatically construct the concept map of a course. The Concept Map Constructing Process shown in Figure 2 is described as follows.

5.1 Association Rules Analyzer

(1) Analysis of association rules generated from Large 2 Itemset

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*** as follows.

Heuristic 1 :

Given two quizzes Q1 and Q2, if concepts of Q1 are prerequisite of concepts of Q2, we summarize the possible learning scenarios of students as follows.

- **Illustrations of rule Q1.L→Q2.L**

Scenario 1) Learners get low grade on Q1 implies that they must get low grade on Q2.

Scenario 2) Learners get low grade on Q2 implies that their grade on Q1 might be bad.

- **Illustrations of rule Q1.H→Q2.H**

Scenario 3) Learners get high grade on Q1 implies that they may also get high grade on Q2

Scenario 4) Learners get high grade on Q2 implies that they must get high grade on Q1.

- **llustrations of rule (Q1.H→Q2.L or Q2.L→Q1.H)**

Scenario 5) Learners get higher grade on Q1 (an easier quiz) but get lower grade on Q2 (a harder quiz).

As shown in Table 6, for convenience to explain the following process in this thesis, we adopt Scenario 1, 4, and 5 of ***Heuristic 1*** to get prerequisite relationships among concept sets of quizzes with parameterized possibility weight for each rule type, 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

Table 6. Prerequisite Relationship of Association Rule

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

In this thesis, 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.

Table 7. The Explanations of Rule Type

Rule	Description of Learning Scenario
$Q_i.L \rightarrow Q_j.L$	CS_{Qi} is the prerequisite of CS_{Qj}. That is, learners get low grade on Q_i implies that they must get low grade on Q_j.
$Q_i.H \rightarrow Q_j.H$	CS_{Qj} is the prerequisite of CS_{Qi}. That is, learners get high grade on Q_j implies that they must get high grade on Q_i
$Q_i.L \rightarrow Q_j.H$	CS_{Qj} is the prerequisite of CS_{Qi} That means learners get higher grade on Qj (an easier or simpler quiz) but get lower grade on Q_i (a harder or more complex quiz).
$Q_i.H \rightarrow Q_j.L$	CS_{Qi} is the prerequisite of CS_{Qj} That means learners get higher grade on Q_i (an easier or simpler quiz) but get lower grade on Q_j (a harder or more complex quiz).

Table 8 shows the result of transforming association rules in Table 5 by analyzing the prerequisite relationships in Table 6.

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	$Q_2.L \rightarrow Q_3.L$	$CS_{Q2} \xrightarrow{pre.} CS_{Q3}$	0.95	1.0
	$Q_3.L \rightarrow Q_2.L$	$CS_{Q3} \xrightarrow{pre.} CS_{Q2}$	1.00	1.0
	$Q_2.L \rightarrow Q_5.L$	$CS_{Q2} \xrightarrow{pre.} CS_{Q5}$	0.86	1.0
	$Q_3.L \rightarrow Q_5.L$	$CS_{Q3} \xrightarrow{pre.} CS_{Q5}$	0.90	1.0
L-H	$Q_1.L \rightarrow Q_5.H$	$CS_{Q5} \xrightarrow{pre.} CS_{Q1}$	0.90	0.8
	$Q_5.L \rightarrow Q_1.H$	$CS_{Q1} \xrightarrow{pre.} CS_{Q5}$	0.82	0.8
H-H	$Q_2.H \rightarrow Q_3.H$	$CS_{Q2} \xrightarrow{pre.} CS_{Q3}$	0.91	1.0
H-L	$Q_5.H \rightarrow Q_1.L$	$CS_{Q5} \xrightarrow{pre.} CS_{Q1}$	1.00	0.8

(2) Analysis of association rules generated from Large n (>= 3) Itemset

In addition to Large 2 itemset, Large 3 itemset may also help refining learning strategies. The possible scenario is as follows.

- We may find $CS_{Q1} \xrightarrow{pre.} CS_{Q2}$ and $CS_{Q2} \xrightarrow{pre.} CS_{Q3}$ generated from large 2 itemset, but we are not sure if concepts of Q_1 and Q_2 must be learned together to ensure concepts of Q_3 well learned. However, we can clarify the uncertainty if we find $CS_{Q1} \cap CS_{Q2} \xrightarrow{pre.} CS_{Q3}$ generated from large 3 itemset.

The scenarios of the larger n (>3) itemset are the same as that described above. Therefore, in this section, we extend ***Heuristic 1*** and only adopt the L-L, H-H rule types to help analyzing the prerequisite relationships between learning concept sets for further analyzing and refining the teaching strategies. For not losing focus on the analysis of Large n itemsets, now we only adopt rules with prerequisite relationship of N:1 ($CS_{Qi} \cap CS_{Qj} \cap \ldots \cap CS_{Qk} \xrightarrow{\text{pre.}} CS_{Qh}$) format after applying ***Heuristic 1***. As shown in Table 9, Scenario 1 and Scenario 4 of ***Heuristic 1*** are adopted here as an example.

Table 9**.** Prerequisite Relationships of Association Rule

Association Rules	**W*Ri***	**Prerequisite relationship**
$Qi.L \cap Qj.L \cap \ldots \cap Qk.L \rightarrow Qh.L$	1.0	$CS_{Qi} \cap CS_{Qj} \cap \ldots \cap CS_{Qk} \xrightarrow{\text{pre.}} CS_{Qh}$
$Qh.H \rightarrow Qi.H \cap Qj.H \cap \ldots \cap Qk.H$	1.0	$CS_{Qi} \cap CS_{Qj} \cap \ldots \cap CS_{Qk} \xrightarrow{\text{pre.}} CS_{Qh}$

The mining results of Large n (>= 3) itemsets are shown in Table 10. From the table, we can see concept sets of Q2, Q3, and Q5 seem to be prerequisite of each other. However, from Table 2, the Test Item–Concept Mapping Table, we know that learning concepts of Q2 and Q3 almost overlap. Thus, we may think rule (Q2.L∩Q3.L → Q5.L) is more meaningful than the others although the confidence of rule (Q2.L∩Q3.L → Q5.L) is smaller than others. With the prerequisite relationship $CS_{Q2} \cap CS_{Q3} \rightarrow CS_{Q5}$ of the rule, the learning strategies may be adapted to learn concepts of Q2 and Q3 together to ensure concepts of Q5 well learned.

Table 10**.** Association Rules generated from Large n (>= 3) Itemset (Confidence > 0.8)

Rule(*Ri*)		**Prerequisite relationship of Concept Set of Quiz**	**Conf (*Ri*)**
L-L	Q2.L∩Q3.L → Q5.L	$CS_{Q2} \cap CS_{Q3} \rightarrow CS_{Q5}$	***0.9***
	Q3.L∩Q5.L → Q2.L	$CS_{Q3} \cap CS_{Q5} \rightarrow CS_{Q2}$	1
	Q2.L∩Q5.L → Q3.L	$CS_{Q2} \cap CS_{Q5} \rightarrow CS_{Q3}$	1

5.2 Concept Map Constructor

Based on the analysis of association rules generated from Large 2 itemset in Section 5.1, we propose a heuristic algorithm to automatically construct the concept map of a course. This algorithm has the function to detect the unreasonable prerequisite relationships of concept sets. According to our literature survey, there is no prior research similar to ours with respect to the capability of detection. For example, in Figure 8, the mined rules, $Q_1.L \rightarrow Q_2.H$ and $Q_1.H \rightarrow Q_2.L$, can be transformed into corresponding prerequisite relationship of concept set resulting in a confused relation as a cycle between concept set, called circularity. That is to say, concepts of Q_1 are prerequisite of concepts of Q_2 and concepts of Q_2 are prerequisite of concepts of Q_1, 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., $CS_{Q1} \rightarrow CS_{Q2} \rightarrow CS_{Q1}$.

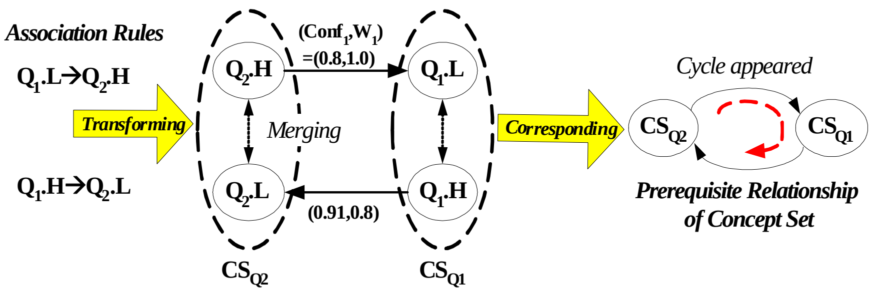

Figure 8. The Transforming of Association Rules

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* $CS_{Q1} = \{\bigcup_1^n a_i\}$

and $CS_{Q2} = \{\bigcup_1^m b_j\}$, *the set of concept-pair is* CS_{Q1} *JOIN* $CS_{Q2} = \{\bigcup_1^k (a_i, b_j)\}$, *where* $a_i \neq b_j$ and S.

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$\}$
- $E = \{\overrightarrow{C_iC_j} \mid i \neq j\} \cup \{\overrightarrow{S_iC_j} \mid$ for each concept set $S\}$

C_i denotes the learning concept as a node. $\overrightarrow{C_iC_j}$ denotes the edge connecting C_i and C_j, where C_i is the prerequisite of C_j. The $\overrightarrow{C_iC_j}$ has an *Influence Weight*, $IW_{(Ci \rightarrow Cj)}$, which denotes the degree of relationship between learning concepts C_i and C_j. The formulation of IW_k is = $\frac{(k-1)\times IW_{k\text{-}1} + W_k \times Conf_k}{k}$, $1 \leq k \leq n$, where n is the amount of $\overrightarrow{C_iC_j}$.

The proposed Concept Map Constructing (CMC) algorithm is shown in Figure 9.

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 and weight of rule to connect the CS_{Qi} and CS_{Qj} according to Table 6.
Step 3 : In C*M*, 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 $\overrightarrow{CS_{Qi}CS_{Qj}}$ 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 $\overrightarrow{CS_{Qi}CS_{Qj}}$.
Step 4: Output the ***CM***.

Figure 9. Concept Map Constructing (CMC) Algorithm

In CMC algorithm, the main purpose of *Cycle Detection Process* is to detect the unreasonable prerequisite relationship as a cycle among concept sets. Thus, the CMC algorithm can generate reasonable prerequisite relationships among concept sets of quizzes. Moreover, the Influence Weight, ***IW***, denotes the degree how the learning status of concept C_i or concept set *Si* 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_{Ri} denotes the possibility of the learning scenario of the association rule in our analysis. Thus, the educational experts can assign different value of W_{Ri} 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 Figure 10. In Figure 10.b, the edges, $\overrightarrow{CS_{Q1}CS_{Q5}}$ and $\overrightarrow{CS_{Q2}CS_{Q3}}$, which have lowest confidences in a cycle, will be deleted by *Cycle Detection Process*. Moreover, Table 11 shows the example of computing the Influence Weight of Concept-Pair (B, D) in Figure 10.f. Because the Concept-Pair (B, D) has two edges between CS_{Q5} and CS_{Q1}, we have to compute the Influence Weight twice.

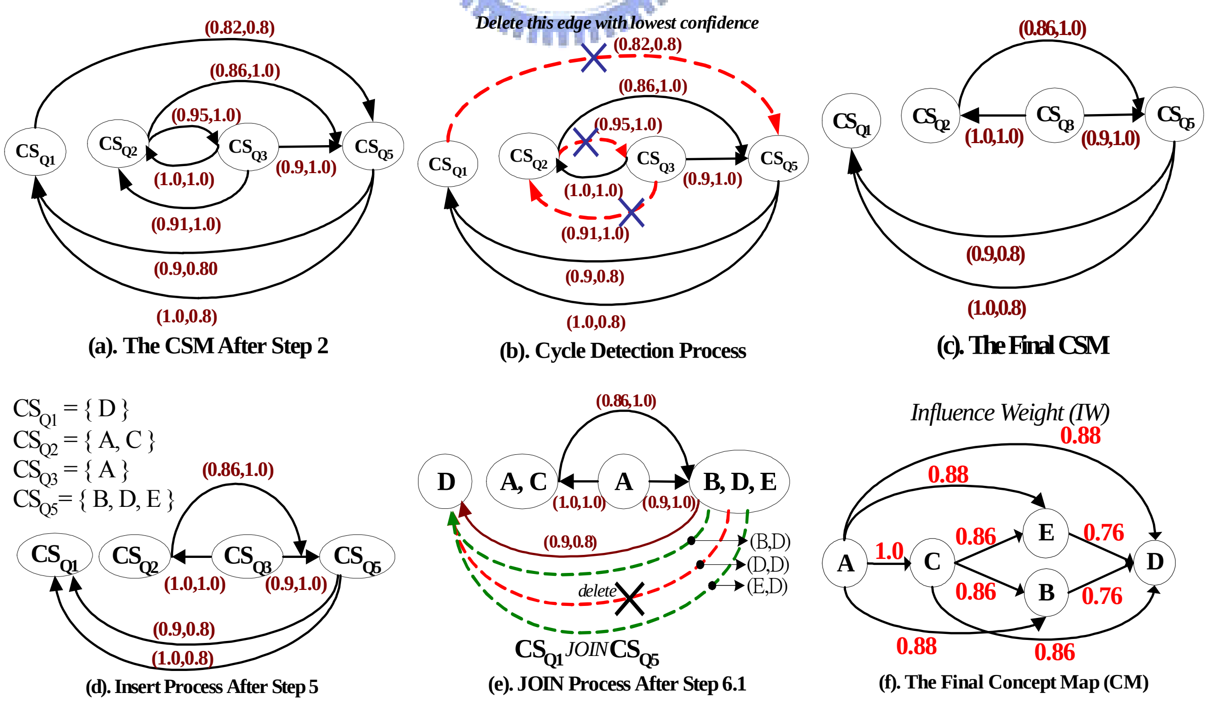

Figure 10. The Process of Concept Map Constructing Algorithm

Table 11. The Computing of the Influence Weight for Concept-Pair (B, D) in Figure 10.f

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

Chapter 6. Evaluating the Redundancy and Circularity of Concept Map

In this thesis, creating a concept map without *Redundancy* and *Circularity* is our concern. Therefore, we create four concept maps using different approaches and evaluate their 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 Figure 11, the prerequisite relationship between concept sets in Figure 11.a is created based upon analyzing L-L rule type only, and the Figure 11.c is created based upon analyzing L-L rule type and anomaly diagnosis we proposed. Then, the concept maps as Figure 11.b and Figure 11.d are transformed according to the **Test Item–Concept Mapping Table**. The Figure 11.e and Figure 11.f are created by our proposed approach.

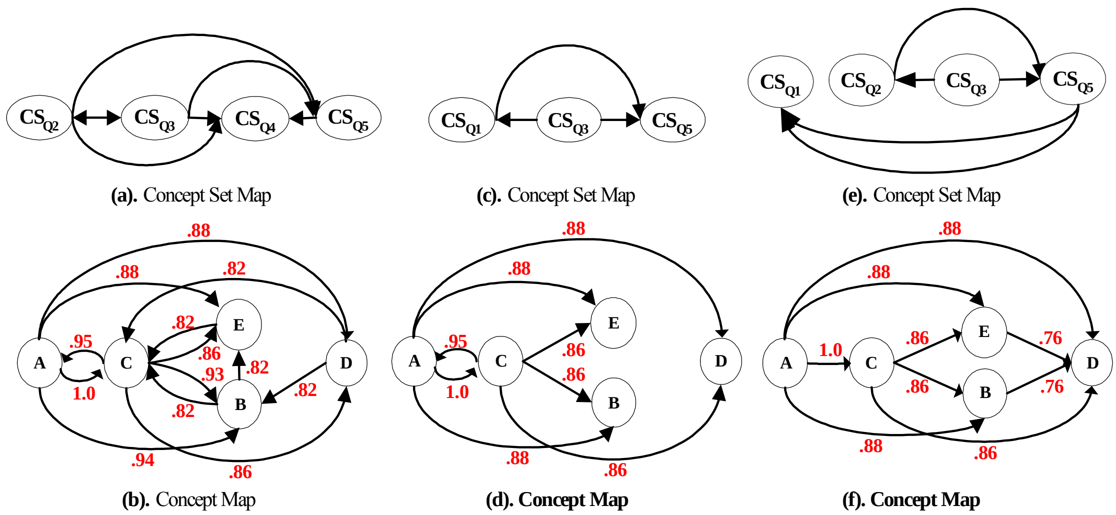

Figure 11. The (a) and (b) created based up analyzing L-L rule type only. The (c) and (d) are created based upon Anomaly Diagnosis and analyzing of L-L rule type. The (e) and (f) created by our approach (only large 2 itemset).

Based upon these results using different approaches, we can conclude that the concept map we proposed has the following characteristics.

- ***Non-Redundancy*** : The *anomaly diagnosis* can filter many useless test records with low discrimination for refining the input data. For example, in Figure 11.a, the Q_4 with low discrimination results in generating many co-prerequisite links as a cycle in Figure 11.b.

- ***Non-Circularity*** : The *cycle detection process* can delete these cycles, e.g., the cycle between A and C in 10.d, to make the concept map un-ambiguous.

Moreover, analyzing association rule with L-L, L-H, H-L, and H-H types generated from Large 2 itemset can refine the concept map, e.g., the edges $\overrightarrow{ED}$ and $\overrightarrow{BD}$ connect the node D only in Figure 11.f.

Chapter 7. The Implementation of TP-CMC

In this chapter, we describe our implementation of the Two Phase Concept Map Construction. The TP-CMC is realized in Java (jdk1.4.2), JGraph (graphic tool for Java), PHP web language and MySQL DBMS.

7.1 Construction of Related Database

The Two Phase Concept Map Constructor provides a friendly graphical user interface to help educational experts or domain experts to construct the Historical Testing Records Database and Test Item-Concept Mapping Database. As shown in Figure 12, you can construct Test Item-Concept Mapping Database of a new course in Figure 12.a, Figure 12.b, and Figure 12.c and construct Historical Testing Records Database of students in Figure 12.d.

For constructing Test Item-Concept Mapping Database of a new course, the course name, number of quizzes and learning concepts of the testing paper must be input first in Figure 12.a and then in Figure 12.b, input the learning concepts contained in the testing paper and the highest grade on each quiz of the testing paper. After the two steps described above, domain experts or teachers can choose the learning concepts of each quiz on the web directly in the Test Item-Concept Mapping choosing table shown in

Figure 12.c.

For constructing Historical Testing Records Database of students, the file recording students' grades on each quiz which domain experts and teachers upload must observe the grade format in Figure 12.d.

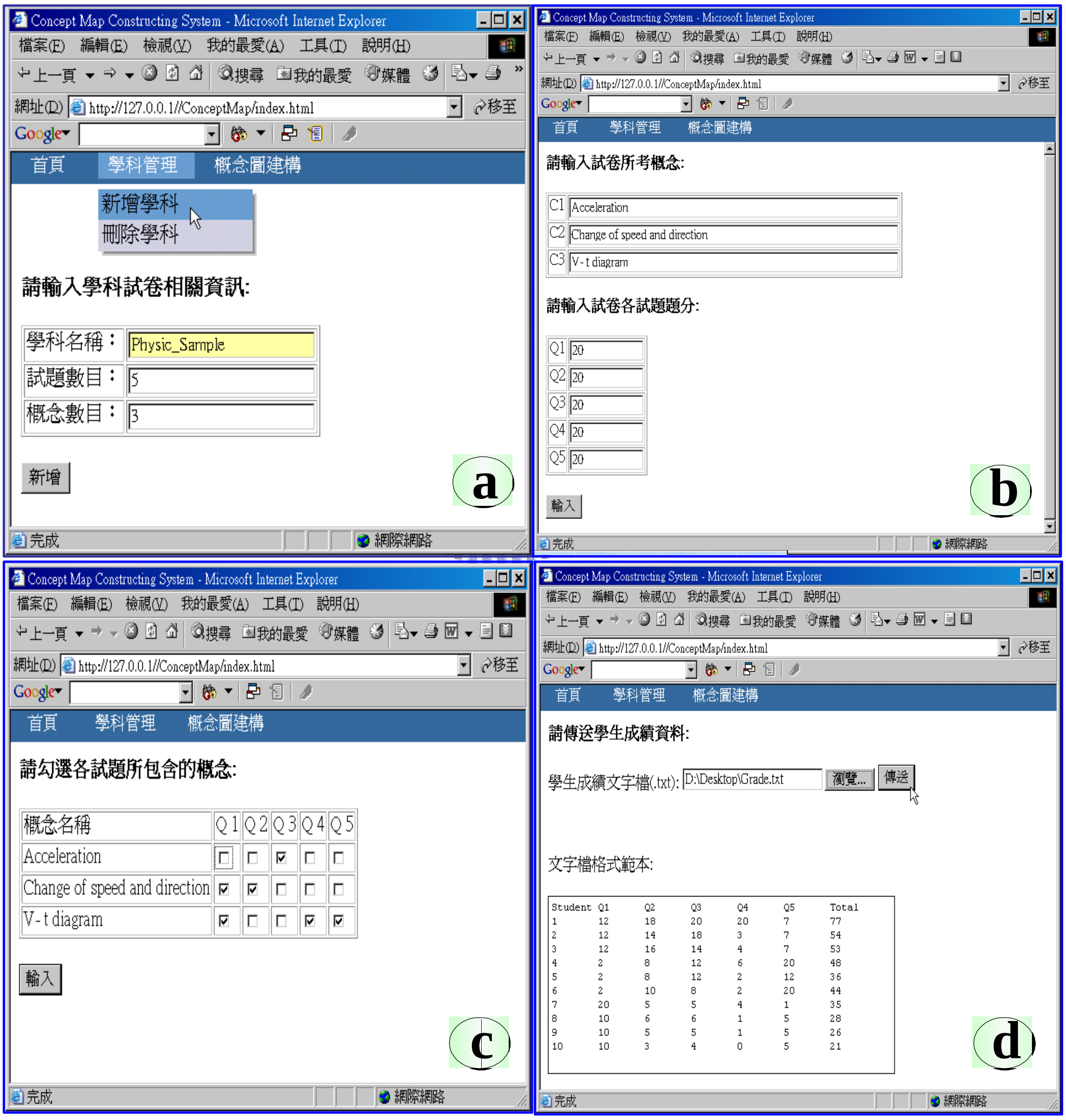

Figure 12. The process of constructing Test Item-Concept Mapping Database and Historical Testing Records Database

7.2 Construction of Concept Maps

As shown in Figure 13.a, the Two Phase Concept Map Constructor provides friendly user interface to help educational experts or domain experts adjusting the parameters to construct the course concept maps. Moreover, in the presentation of the constructed concept map, as Figure 13.b and Figure 13.c shows, each node is draggable in our constructor for a better view without obscurity.

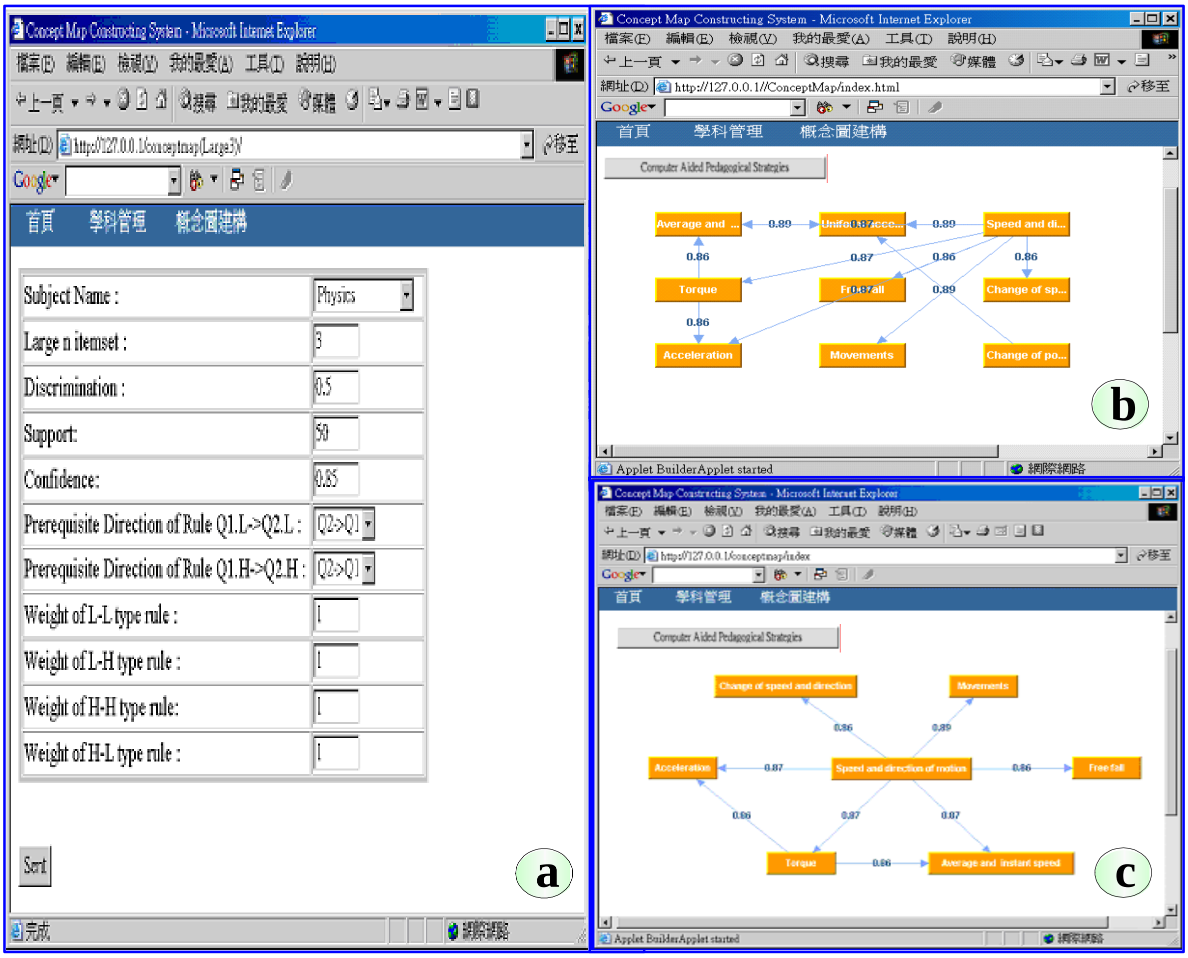

Figure 13. The process of constructing the concept map

The *Large n itemset* value represents the layer frequent itemsets found. The *discrimination* value represents the lowest acceptable degree of the discrimination of the test item. The *Support, Confidence* values represent the mining thresholds in LFMAlg. Cause of the design of LFMAlg, the *Support* value is not between 0 ~ 1. From the experiment, we find that better concept maps can be constructed when the *Support* value is set half of the number of the students participating in the testing. The *Confidence* value is between 0 ~ 1, representing the lowest acceptable connection level for two quizzes the grade level students get based upon the conditional probability.

The *Prerequisite Direction of Rule Q1.L→Q2.L or Q1.H→Q2.H* values decide which scenario of Heuristic 1 chosen. Take *Prerequisite Direction of Rule Q1.L→Q2.L* as an example, Figure 14.a represents the adoption of prerequisite direction Q1→Q2 and Figure 14.b represents the adoption of prerequisite direction Q2→Q1. As shown in Figure 14, learning concept ***(Speed and direction of motion)*** in Figure 14.a is more complex than other concepts, however, in Figure 14.b learning concept ***(Speed and direction of motion)*** is prerequisite of most learning concepts. Though Scenario 1 is more intuitive than Scenario 2, the concept map constructed in Figure 14.b is more reliable than concept map in Figure 14.a. Therefore, in our approach, the *Prerequisite Direction of Rule* is configurable according to different domains and learner's backgrounds. Finally, the *Weight of L-L, L-H, H-H and H-L* value is between 0 ~ 1, representing how reliable the rule type domain experts or experienced teachers think is.

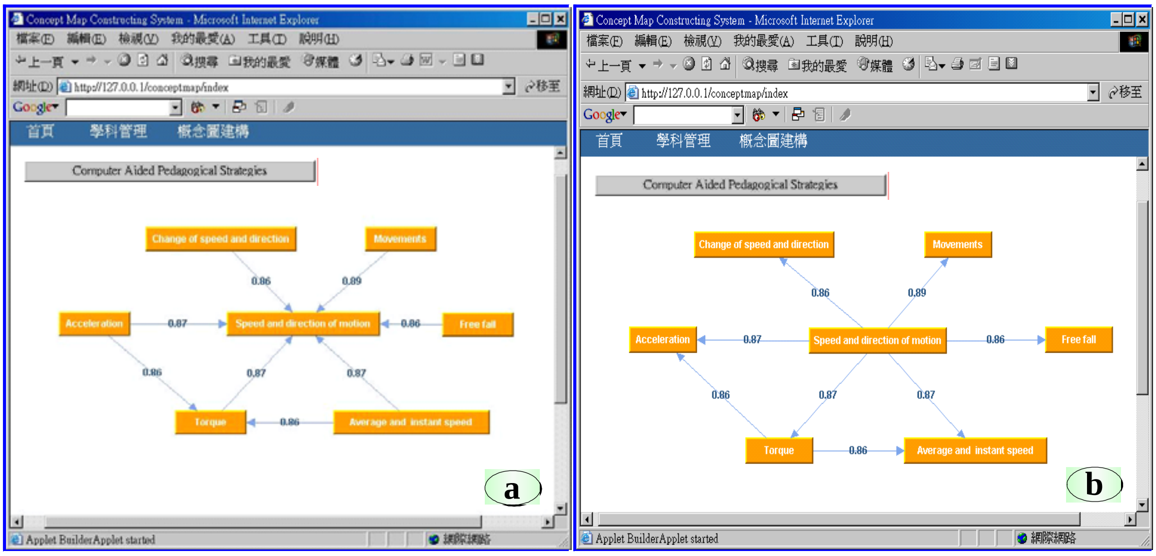

Figure 14. The concept maps (a) and (b) are created with Scenario 1 and 2 of Heuristic 1 by TP-CMC approach respectively. (Discrimination 0.5, Support=50, Confidence=0.85)

Moreover, after click on button "Computer Aided Pedagogical Strategies", as shown in Figure 15, rules generated from Large 3 Itemset can provide the information for refining pedagogical strategies. For example, prerequisite relationship $Q_8 \cap Q_{38} \rightarrow Q_5$found may provide information for teachers that sound learning of concepts of Q_8 and Q_{38} may make learning concepts of Q_5 better.

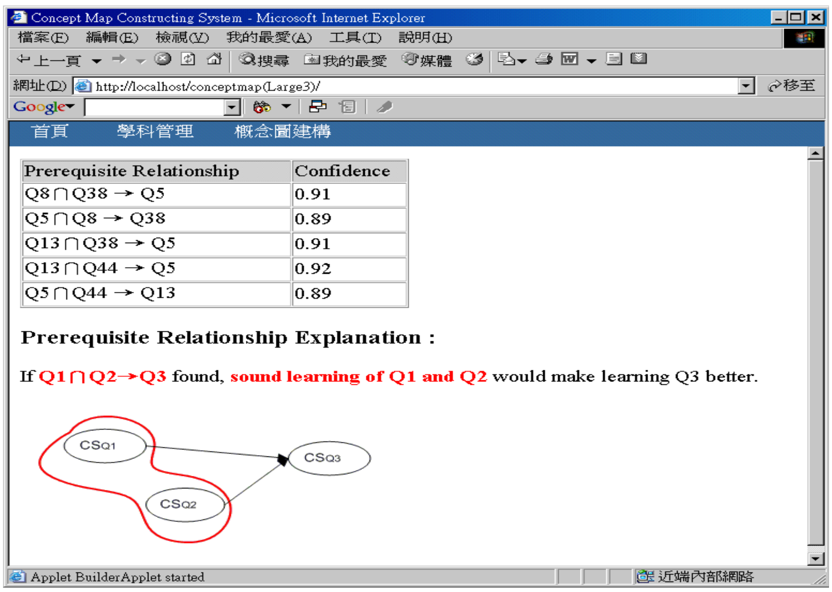

Figure 15. Association Rules Generated from Large 3 itemset (Discrimination 0.5, Support=50, Confidence=0.85)

Chapter 8. The Experiment of TP-CMC

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

8.1 Experimental Results in Physics Course

The participants of the experiment are the 104 students of junior high school in Taiwan and the domain of the examination is the Physics course. The related statistics of testing results and related concepts of testing paper are shown in Table 12 and Table 13.

Table 12. 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 13. 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

As shown in Figure 16.a, Figure 16.b, and Figure 16.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 Figure 16.a and Figure 16.b show that the prerequisite relationships between learning concepts are very disorderly 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 Figure 16.c.

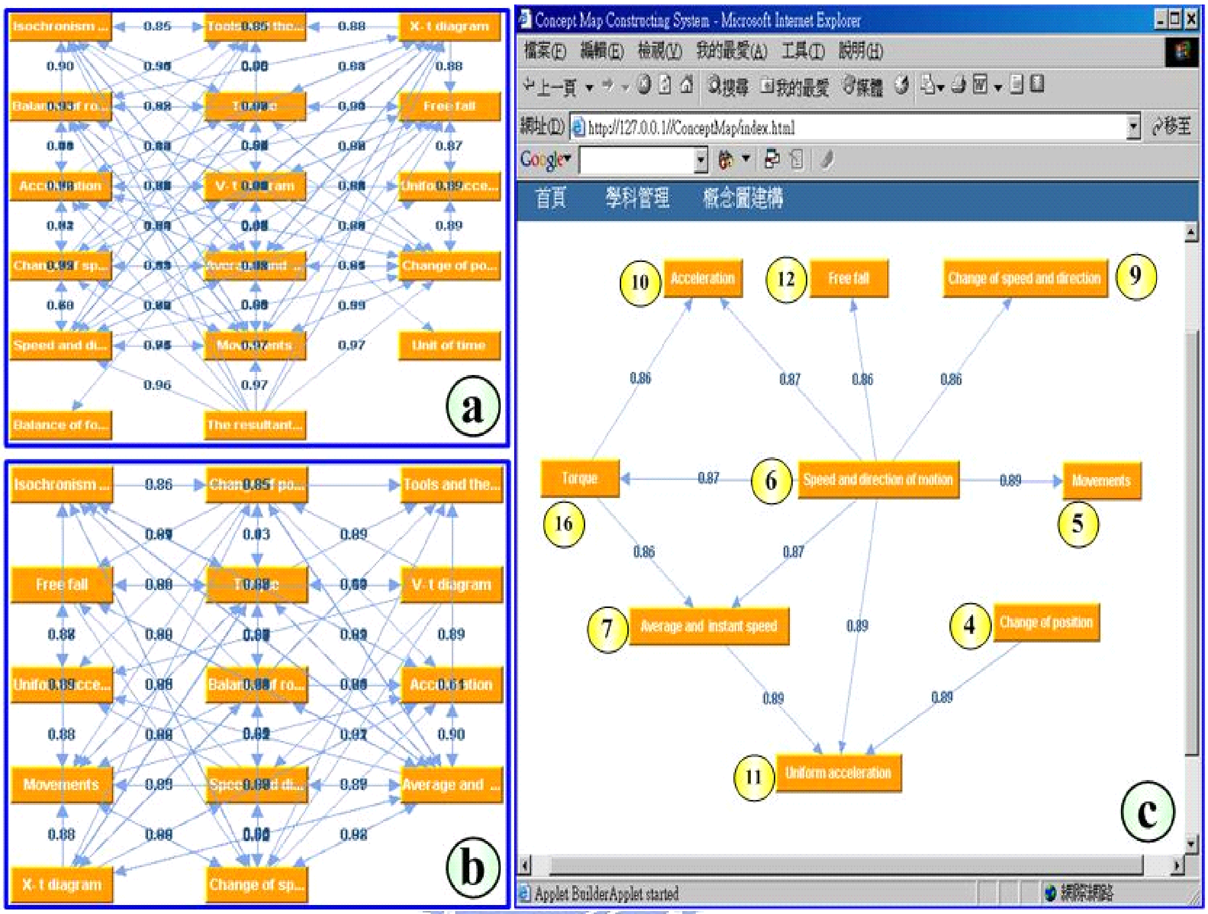

Figure 16. The concept maps (a), (b), and (c) with Discrimination 0.0, 0.3, and 0.5 are created by TP-CMC approach respectively. (Support=50, Confidence=0.85)

The comparison of percentage of rules found and concepts involved are shown in Figure 17, the anomaly diagnosis function in the first phase indeed reduces many ambiguous and useless relationships among learning concepts. From the figure, we know the lower the discrimination is, the larger the variation of the number of rules found is. However, the percentage of concepts involved doesn't change so obviously.

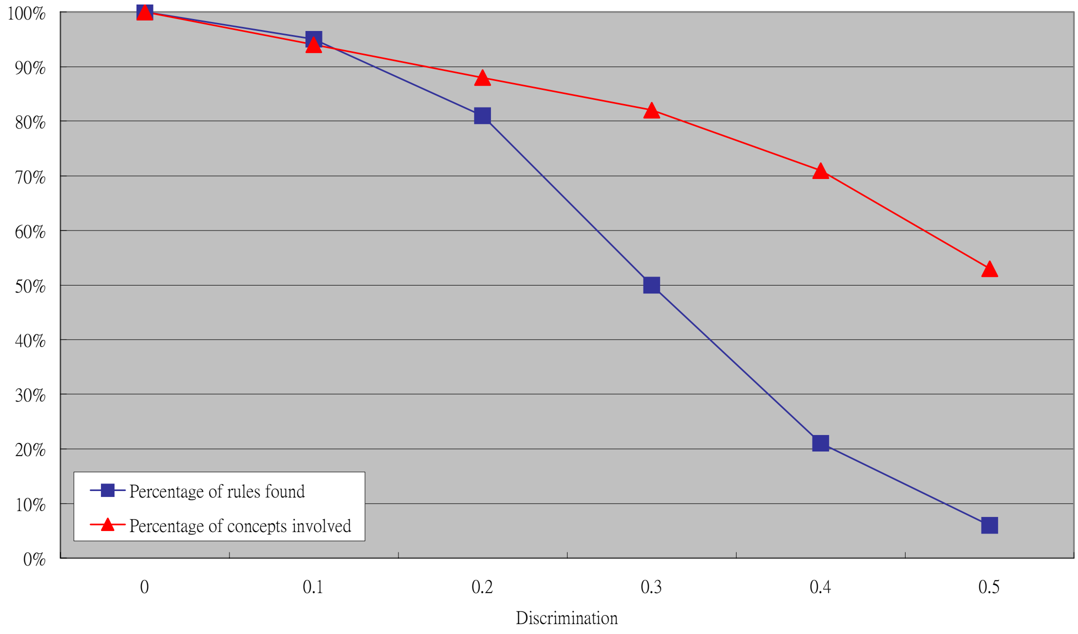

Figure 17. The comparison of percentage of rules found and concepts involved

Besides, the rationality of the concept map constructed has been also discussed with educational experts. The prerequisite relationships among the learning concepts are compatible with teachers' teaching strategies. 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 Figure 16.c represents that if students don't 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***.

Chapter 9. Conclusion and Future Work

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 thesis, 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 feasible.

In the near future, we may further analyze the rules with Large 2 itemset from combinational view and 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 more experiments based upon real learning testing records, too.

Reference

[1] Test Diagnostics, http://www.edtech.vt.edu/edtech/id/assess/diagnostics.html.

[2] IHMC, Cmap Tools, http://cmap.ihmc.us/

[3] J. Appleby, P. Samuels, and T.T. Jones , "Diagnosis–A Knowledge-based Diagnostic Test of Basic Mathematical Skills" , *Computers & Education* , 28(2), February, 1997, pp. 113-131.

[4] Eric Bruillard & Georges-Louis Baron, "Computer-based concept mapping: a review of a cognitive tools for students", *International Conference on Educational Uses of Communication and Information Technologies*, Beijing, China, August 21-25, 2000, pp. 331-338

[5] K.E. Chang, S.H. Liu, and S.W. Chen, "A Testing System for Diagnosing Misconceptions in DC Electric Circuits", *Computers & Education*, 31(2), September, 1998, pp. 195-210.

[6] V. Carchiolo, A. Longheu, and M. Malgeri, "Adaptive Formative Paths in a Web-based Learning Environment", *Educational Technology & Society,* 5(4), October 2002, pp. 64-75.

[7] Jonassen, D., Beissner, K. & Yacci, M., "*Structural knowledge: techniques for representing , conveying and acquiring structural knowledge*", 1993, Hillsdale (N. J.): Erlbaum.

[8] G. Frosini, B Lazzerini, and F. Marcelloni, "Perform Automatic Exams", *Computers & Education*, 31(3), November, 1998, pp. 281-300.

[9] H. Gamboa, "Designing Intelligent Tutoring Systems : A Bayesian Approach", *Proc. of Ana*

Fred 3rd International Conference on Enterprise Information Systems (ICEIS'2001), 2001, pp. 452-458.

[10] J.L. Gordon, “Creating knowledge maps by exploiting dependent relationships”, *Knowledge-Based Systems*, 13(2000), pp.71-79

[11] G.J. Hwang, “A Conceptual Map Model For Developing Intelligent Tutoring System”, *Computers & Education*, 40(3), April, 2003, pp. 217-235.

[12] G.J. Hwang, C.L. Hsiao, and C. R. Tseng, “A Computer-Assisted Approach to Diagnosing Student Learning Problem in Science Course”, *Journal of Information Science & Engineering*, 19(2), 2003, pp. 229-248.

[13] Chen H., and Lynch KJ, "Automatic Construction of Networks of Concepts Characterizing Document Database”, *IEEE Transactions on Systems, Man and Cybernetics*, (22:5), 1992, pp. 885-902.

[14] C. S. Hsu, S. F. Tu, and G. J. Hwang, “A Concept Inheritance Method for Learning Diagnosis of a Network-based Testing and Evaluation System”, *proc. of The 7th International Conference on Computer-Assisted Instructions*, 1998, pp. 602-609.

[15] Alenka Kavčič, "Adaptation Techniques in Adaptive Hypermedia Systems", *Proceedings of the 22nd International Convention MIPRO'99, Conference on Multimedia and Hypermedia Systems*, Volume 1, Opatija, Croatia, May 1999

[16] J.D. Novak, *Learning, Creating, and Using Knowledge: Concept Maps As Facilitative Tools*

in Schools and Corporations, Lawrence Erlbaum Assoc 1998.

[17] Ana Oliveira, Francisco Camara Pereira, and Amilcar Cardoso, "Automatic Reading and Learning from Text", *Proc. of the International Symposium on Artificial Intelligence*, 2001, pp. 302-310

[18] W.J. Popham, *Classroom Assessment: What Teachers Need to Know*, Pearson Allyn & Bacon, 1999, pp.222-227.

[19] K.A. Papanikolaoua, M. Grigoriadoua, G.D. Magoulasb, and H. Kornilakisa, “Towards New Forms of Knowledge Communication: The Adaptive Dimension of a Web-based Learning Environment”, *Computers & Education*, 39(4), December, 2002, pp. 333-360.

[20] Brain R. Gaines and Mildred L.G. Shaw, “Concept maps as hypermedia components”, *International Journal of HumanComputer Studies*, 43(3), 1995, pp. 323--361.

[21] Hajime Saito, Takashi Ohno, Fuminori Ozaki, Kenji Saito, Takashi Maeda, and Azuma Ohuchi, “A Semi-Automatic Construction method of Concept Map Based on Dialog Contents”, *Proceedings of International Conference on Computers in Education (ICCE'2001)*, Seoul (Korea), 2001, pp. 628-635

[22] E. Triantafllou, A. Pomportsis, and S. Demetriadis, “The Design And The Formative Evaluation of An Adaptive Educational System Based on Cognitive Styles”, *Computers & Education*, 41(1), August, 2003, pp. 87-103.

[23] C.J Tsai, S. S. Tseng, and C.Y Lin, "A Two-Phase Fuzzy Mining and Learning Algorithm for

Adaptive Learning Environment", *Proc. of Int'l Conf. on Computational Science (ICCS'01)*, *Lecture Notes in Computer Science (LNCS 2074)*, Vol. 2, pp. 429-438, CA, USA, May, 2001.