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
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.

© 2005 Elsevier Ltd. All rights reserved.

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, $f_{AB}$ represents the amount of learners with wrong answers of skill $A$ and right answers of skill $B$. If $f_{AB} > f_{BA}$, 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 Fuzzy Mining Association Rule Algorithm (LFMAAlg) 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
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 $CS_{Q_i}$, 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 Mining Algorithm (LFMAAlg) 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 LFMAAlg 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.
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.
Example 2. Table 3 shows the fuzzified testing grades of learners on Q4 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 \( \text{Max}(L, M, H) \) function.

Then, by applying FIA-NR algorithm, we can get the Difficulty and Discrimination of every quiz. For example, the \( P_{4H} \) and \( P_{4L} \) of Q4 are \( P_{4H} = \frac{R_{4H}}{N_{4H}} = \frac{H+L+L}{3} = \frac{1+0+0}{3} = \frac{1}{3} \) and \( P_{4L} = 0 = 0 \), respectively. Therefore, its Difficulty \( D_4 \) and Discrimination \( D_4 \) are \( D_4 = 1 - \frac{P_{4H}+P_{4L}}{2} = 1 - \frac{\frac{1}{3}+0}{2} = \frac{5}{6} = 0.83 \) and 0.33 respectively. Thus, learners’ grade on Q4 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>
<thead>
<tr>
<th>Group</th>
<th>High</th>
<th>Middle</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner ID</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Total (100)</td>
<td>77</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td>Grade level = Max(L, M, H)</td>
<td>H</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

Table 4
Difficulty and discrimination degree of each quiz

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty (0 to 1)</td>
<td>0.25</td>
<td>0.42</td>
<td>0.42</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>Discrimination (−1 to 1)</td>
<td>0.5</td>
<td>0.83</td>
<td>0.83</td>
<td>0.33</td>
<td>0.5</td>
</tr>
</tbody>
</table>
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 LFMAAlg Algorithm, the support value of every itemset $x$ in candidate $C_\ell$ 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 $(A \cup B) = \sum_n^{\ell} \text{Min}(A, B)$, 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: LFMAAlg Algorithm

Symbol Definition:
- $\alpha_\ell$: The minimum support threshold in the $\ell$-large itemset.
- $C_\ell$: The $\ell$-Candidate itemset.
- $L_\ell$: The $\ell$-large itemset.
- $\lambda$: The minimum confidence threshold.

Input: The test records of learners after Fuzzification and Anomaly Diagnosis.
The minimum support threshold $\alpha_\ell$ and $\lambda$.

Output: The fuzzy association rules of test records of learners.

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

Step2: Generate the association rules according to the given $\lambda$ in $L_\ell$.

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

Fig. 6. Mining process of Large 2 Itemset.
Table 5
The mining results (Conf_i > 0.8)

<table>
<thead>
<tr>
<th>Rule types</th>
<th>Mined rules</th>
<th>Conf_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>L–L</td>
<td>Q2 · L → Q3 · L</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Q3 · L → Q2 · L</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Q2 · L → Q5 · L</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Q3 · L → Q5 · L</td>
<td>0.90</td>
</tr>
<tr>
<td>L–H</td>
<td>Q1 · L → Q5 · H</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Q5 · L → Q1 · H</td>
<td>0.82</td>
</tr>
<tr>
<td>H–H</td>
<td>Q2 · H → Q3 · H</td>
<td>0.91</td>
</tr>
<tr>
<td>H–L</td>
<td>Q5 · H → Q1 · L</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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 ith mined association rule. For example, the Confidence (Conf_i) of rule Q_2 · L → Q_3 · 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.
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 \( CS_{Q1} = \{ \cup_i a_i \} \) and \( CS_{Q2} = \{ \cup_i b_j \} \), the set of concept-pair is \( CS_{Q1} \) JOIN \( CS_{Q2} = \{ \cup_i^k (a_i, b_j) \} \),

### Table 7
The explanations of rule types

<table>
<thead>
<tr>
<th>Rule type</th>
<th>Description of learning scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>L → L</td>
<td>If the association rule ( Q_i \cdot L \rightarrow Q_j \cdot L ) is mined, it means that the ( CS_{Q_i} ) is the prerequisite of ( CS_{Q_j} ), represented as ( CS_{Q_i} \rightarrow^{pre} CS_{Q_j} ). That is why getting low grade on ( Q_i ) might imply getting low grade on ( Q_j ).</td>
</tr>
<tr>
<td>H → H</td>
<td>If the association rule ( Q_i \cdot H \rightarrow Q_j \cdot H ) is mined, it means that the ( CS_{Q_i} ) is the prerequisite of ( CS_{Q_j} ).</td>
</tr>
<tr>
<td>L → H</td>
<td>If the association rule ( Q_i \cdot L \rightarrow Q_j \cdot H ) is mined, it means that the ( CS_{Q_i} ) is the prerequisite of ( CS_{Q_j} ) because ( CS_{Q_i} ) may be not learned well resulting from ( CS_{Q_j} ).</td>
</tr>
<tr>
<td>H → L</td>
<td>If the association rule ( Q_i \cdot H \rightarrow Q_j \cdot L ) is mined, it means that the ( CS_{Q_i} ) is the prerequisite of ( CS_{Q_j} ).</td>
</tr>
</tbody>
</table>

### Table 8
Result by analyzing the prerequisite relationships in Table 6

<table>
<thead>
<tr>
<th>Rule type</th>
<th>Association rules of quiz</th>
<th>Prerequisite relationship of concept set</th>
<th>Conf,</th>
<th>( W_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>L–L</td>
<td>Q2 \cdot L → Q3 \cdot L</td>
<td>( CS_{Q2} \rightarrow^{pre} CS_{Q3} )</td>
<td>0.95</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Q3 \cdot L → Q2 \cdot L</td>
<td>( CS_{Q3} \rightarrow^{pre} CS_{Q2} )</td>
<td>1.00</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Q2 \cdot L → Q5 \cdot L</td>
<td>( CS_{Q2} \rightarrow^{pre} CS_{Q5} )</td>
<td>0.86</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Q3 \cdot L → Q5 \cdot L</td>
<td>( CS_{Q3} \rightarrow^{pre} CS_{Q5} )</td>
<td>0.90</td>
<td>1.0</td>
</tr>
<tr>
<td>L–H</td>
<td>Q1 \cdot L → Q5 \cdot H</td>
<td>( CS_{Q5} \rightarrow^{pre} CS_{Q1} )</td>
<td>0.90</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Q5 \cdot L → Q1 \cdot H</td>
<td>( CS_{Q1} \rightarrow^{pre} CS_{Q5} )</td>
<td>0.82</td>
<td>0.8</td>
</tr>
<tr>
<td>H–H</td>
<td>Q2 \cdot H → Q3 \cdot H</td>
<td>( CS_{Q2} \rightarrow^{pre} CS_{Q3} )</td>
<td>0.91</td>
<td>1.0</td>
</tr>
<tr>
<td>H–L</td>
<td>Q5 \cdot H → Q1 \cdot L</td>
<td>( CS_{Q5} \rightarrow^{pre} CS_{Q1} )</td>
<td>1.00</td>
<td>0.8</td>
</tr>
</tbody>
</table>
where $a_i \neq b_j$ and $k \leq n \times m$. 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 $i\}$
- $E = \{C_iC_j | i \neq j\}$

The node, $C_i$, denotes the learning concept and the edge, $C_iC_j$, which connects $C_i$ and $C_j$, denotes that $C_i$ is the prerequisite of $C_j$. The $C_iC_j$ 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$_Q$ for $Q_iL$ or $Q_iH$ and CS$_Q$ for $Q_jL$ or $Q_jH$.
- 2.2: Create a new prerequisite edge with confidence ($Conf_i$) and weight ($W_i$) of rule to connect the CS$_Q$ and CS$_Q$ according to Table 6.

**Step 3**: In CM, repeatedly execute this step until no cycle is found.

- 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$_Q$ in CM

- \{ Insert node’s corresponding learning concepts according to Test item–Concept Table \}

**Step 6**: For each edge CS$_Q$ CS$_Q$ 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 $k$th concept-pair ($C_i, C_j$) according to weight ($W_k$) and confidence ($Conf_k$) of CS$_Q$ CS$_Q$.

**Step 4**: Output the $CM$.

---

Fig. 7. The transforming of association rules.

Fig. 8. Concept Map Constructing (CMC) Algorithm.
of relationship between learning concepts. The formulation of $IW_k$ is $( (k - 1) \times IW_{k-1} + W_k \times \text{Conf}_{k}/k, 1 \leq k \leq n$, where $n$ is the amount of $C_iC_j$.

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 $f_{12} > f_{12}$ 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 $f_{12} > f_{12}$, where $f_{12} = \text{Count}(Q_1 \cdot H \cap Q_2 \cdot L)$ and $f_{12} = \text{Count}(Q_1 \cdot L \cap Q_2 \cdot H)$. 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 $C_iC_j$ 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

<table>
<thead>
<tr>
<th>Relative quizzes frequency</th>
<th>(Q_1) Higher</th>
<th>(Q_1) Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q_2) Higher</td>
<td>$f_{12}$</td>
<td>$f_{12}$</td>
</tr>
<tr>
<td>(Q_2) Lower</td>
<td>$f_{12}$</td>
<td>$f_{12}$</td>
</tr>
</tbody>
</table>

Fig. 9. The process of Concept Map Constructing Algorithm.
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 Circularität is our concern. As shown in Fig. 10, we create three concept maps by using different approaches and evaluate their

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

<table>
<thead>
<tr>
<th>Rule</th>
<th>Prerequisite relationship</th>
<th>Conf(_i)</th>
<th>(W_i)</th>
<th>(IW_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q_1 \rightarrow Q_5 \rightarrow H)</td>
<td>CS(<em>{Q5}) \rightarrow CS(</em>{Q1})</td>
<td>0.90</td>
<td>0.8</td>
<td>(W_1 \times \text{Conf}_1 = 0.9 \times 0.8 \approx 0.72)</td>
</tr>
<tr>
<td>(Q_5 \rightarrow H \rightarrow Q_1 \rightarrow L)</td>
<td>CS(<em>{Q5}) \rightarrow CS(</em>{Q1})</td>
<td>1.00</td>
<td>0.8</td>
<td>(\frac{2(1-0.72)(0.8)}{0.8} = 0.76)</td>
</tr>
</tbody>
</table>

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 Circularty. 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 Q4 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 \( \overline{ED} \) and \( \overline{BD} \) 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 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

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

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).
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


