Expert Systems with Applications 37 (2010) 4496–4503

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Using data mining techniques to automatically construct concept maps for adaptive learning systems

Shyi-Ming Chen *, Shih-Ming Bai

Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC

ARTICLE INFO

Keywords:
Adaptive learning systems
Apriori algorithm
Concept maps
Concepts-relationship analysis
Data mining

ABSTRACT

It is obvious that to construct concept maps for providing the learning guidance to learners is an important research topic of adaptive learning systems. Because the existing method to construct concept maps only considers the association rules that questions are not correctly answered, it will miss some information about questions that are correctly answered by the learners. Moreover, the existing method also has the drawback that it will build unnecessary relationships or lose some relationships between concepts in the constructed concept maps. In this paper, we present a new method to automatically construct concept maps based on data mining techniques for adaptive learning systems. The proposed method can overcome the drawbacks of the existing method. It provides us a useful way to automatically construct concept maps in adaptive learning systems.

1. Introduction

In recent years, some researchers focused on the research topics of adaptive learning systems (Bai & Chen, 2008a, 2008b, 2008c, 2008d; Carchiolo, Longheu, & Malgeri, 2002; Chen & Bai, 2009; Hwang, 2003; Lee, Lee, & Leu, 2009; Novak, 1998). Bai and Chen (2008a) presented a method for evaluating student learning achievement using fuzzy membership functions and fuzzy rules. Bai and Chen (2008b) presented a method for automatically constructing concept maps based on fuzzy rules for adaptive learning systems. Bai and Chen (2008c) presented a method for automatically constructing grade membership functions for students' evaluation for fuzzy grading systems. Chen and Bai (2009) presented a method for learning barrier diagnosis of learners based on fuzzy rules. Carchiolo et al. (2002) presented a method for dealing with adaptive formative paths in a web-based learning environment. Hwang (2003) presented a method for applying the conceptual map model for developing intelligent tutoring systems. Novak (1998) presented a method for using concept maps as facilitative tools in schools and corporations.

To construct concept maps for providing the learning guidance to learners is an important research topic of adaptive learning systems. Bai and Chen (2008b) presented a method for automatically constructing concept maps based on fuzzy rules. Lee et al. (2009) presented a method to automatically construct concept maps for conceptual diagnosis of e-learning. Lee et al. used the Apriori algorithm (Agrawal & Srikant, 1994) for automatically constructing concept maps and proposed the remedial-instruction path (RIP) algorithm to build an intelligent concept diagnostic system (ICDS). However, Lee et al.'s method (2009) for automatically constructing concept maps has the following drawbacks:

(1) Because it only considers the association rules that questions are not correctly answered, it will miss some information about questions that are correctly answered by the learners.
(2) It will build unnecessary relationships or lose some relationships between concepts in the constructed concept maps.

In this paper, we present a new method to automatically construct concept maps for adaptive learning systems based on data mining techniques to overcome the drawbacks of Lee et al.'s method (2009). The proposed method constructs concept maps based on the analysis of concept maps and the data mining techniques. It provides a useful way to construct concept maps for adaptive learning systems.

The rest of this paper is organized as follows: In Section 2, we briefly review Lee et al.'s method (2009) for automatically constructing concept maps. In Section 3, we present a new method to automatically construct concept maps based on the analysis of concept maps and the Apriori algorithm (Agrawal & Srikant, 1994) for adaptive learning systems. In Section 4, we use an example to illustrate the process of automatically constructing concept maps based on the proposed method. The conclusions are discussed in Section 5.

* Corresponding author. Addrress: Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, 43, Section 4, Keelung Road, Taipei 106, Taiwan, ROC. Tel.: +886 2 27376417; fax: +886 2 27301081.
E-mail address: smchen@mail.ntust.edu.tw (S.-M. Chen).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.12.060

2. A review of Lee et al's method for automatically constructing concept maps

Lee et al. (2009) presented a method to automatically construct concept maps for conceptual diagnosis of e-learning. They used the Apriori algorithm (Agrawal & Srikant, 1994) for automatically constructing concept maps. In the following, we briefly review Lee et al.'s method for automatically constructing concept maps, shown as follows:

Step 1: Set the conceptual weights of concepts in test questions by the teacher.

Step 2: Recording the test portfolio of each learner.

Step 3: Perform data mining based on the Apriori algorithm (Agrawal & Srikant, 1994) to find out the large itemsets.

Step 4: Based on the large itemsets obtained in Step 3, generate the association rule $Q_i \rightarrow Q_j$ indicating that if the learner gives a wrong answer to question Q_i, then he/she gives a wrong answer to question Q_j, too.

Step 5: Based on the comparison chart of the conceptual weight relationships in test questions, convert each association rule " $Q_i \rightarrow Q_j$" obtained in Step 4 into the relationship between concepts, shown as follows:

$$\begin{aligned} \mathrm{Q}_i \rightarrow \mathrm{Q}_j \rightarrow W_{\mathrm{C}_x\mathrm{C}_y} &= \mathrm{C}_x \rightarrow \mathrm{C}_y \\ &= \mathrm{confidence}(\mathrm{Q}_i \rightarrow \mathrm{Q}_j) \times R_{\mathrm{Q}_i\mathrm{C}_x} \times R_{\mathrm{Q}_j\mathrm{C}_y}, \end{aligned}$$

where Q_i and Q_j are questions, "$Q_i \rightarrow Q_j$" denotes the association rule from Q_i to Q_j, $W_{C_xC_y}$ denotes the relevance degree between C_x and C_y, "$C_x \rightarrow C_y$" denotes the relationship from C_x to C_y with respect to the association rule $Q_i \rightarrow Q_j$, "confidence $(Q_i \rightarrow Q_j)$" denotes the confidence of the association rule $Q_i \rightarrow Q_j$, $R_{Q_iC_x}$ denotes the conceptual weight of concept C_x in question Q_i, $R_{Q_jC_y}$ denotes the conceptual weight of concept C_y in question $Q_j, 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant m, 1 \leqslant x \leqslant p$ and $1 \leqslant y \leqslant p$.

Step 6: For each relationship between concepts, put the concepts and their relationship edges into the "preliminary concept map" with the relevance degrees of the relationship obtained in Step 5.

Step 7: If there are more than one edge between concept C_i and concept C_j in the preliminary concept map, then remove the relationship edge between concepts C_i and C_j with smaller relevance degrees and keep the relationship between concepts C_i and C_j having the largest relevance degree.

Step 8: Analyze the priority relationship between concepts. Calculate the number NP of parent concepts and the number NC of child concepts of each concept in the concept map, then sort the priority sequence of concepts in the concept map to construct the complete concept map.

However, Lee et al.'s method has the following drawbacks:

(1) Because it only considers the association rules that questions are not correctly answered, it will miss some information about questions that are correctly answered by the learners.

(2) It will build unnecessary relationships or lose some relationships between concepts in the constructed concept maps.

In the next section, we will present a new method to automatically construct concept maps based on data mining techniques to overcome the drawbacks of Lee et al.'s method (2009).

3. A new method to automatically construct concept maps for adaptive learning systems

Assume that there are n learners $S_1, S_2, \ldots, S_n$ and assume that there are m questions $Q_1, Q_2, \ldots, Q_m$ in a test paper, then we can get a grade matrix G, shown as follows:

$$G = \begin{array}{c} \\ Q_1 \\ Q_2 \\ \vdots \\ Q_m \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & \cdots & S_n \end{array} \\ \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{bmatrix} \end{array}$$

where $g_{ij} \in [0,1]$, $g_{ij} = 0$ denotes that the learner S_j does not correctly answer the question Q_i, $g_{ij} = 1$ denotes that the learner S_j correctly answers the question Q_i, $1 \leqslant i \leqslant m$, and $1 \leqslant j \leqslant n$. Moreover, assume that there is a questions-concepts mapping matrix QC, shown as follows:

$$QC = \begin{array}{c} \\ Q_1 \\ Q_2 \\ \vdots \\ Q_m \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_p \end{array} \\ \begin{bmatrix} qc_{11} & qc_{12} & \cdots & qc_{1p} \\ qc_{21} & qc_{22} & \cdots & qc_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ qc_{m1} & qc_{m2} & \cdots & qc_{mp} \end{bmatrix} \end{array}$$

where qc_{ij} denotes the conceptual weight of the concept C_j in the question $Q_i, 0 \leqslant qc_{ij} \leqslant 1, 1 \leqslant i \leqslant m$, and $1 \leqslant j \leqslant p$. The proposed method for constructing the concept map based on the concepts appearing in the questions of a test paper is now presented as follows:

Step 1: Based on the testing records of the learners and the Apriori algorithm (Agrawal & Srikant, 1994), mine the following two kinds of association rules between questions:

(1) If the question Q_i is correctly learned by the learner, then the question Q_j is also correctly learned by the learner.

(2) If the question Q_i is incorrectly learned by the learner, then the question Q_j is also incorrectly learned by the learner.

Construct the association rules from each question in the large 1-itemset to all other questions and calculate the confidence of each mined association rule by the Apriori algorithm, where the confidence "$\mathrm{conf}(Q_i \rightarrow Q_j)$" of an association rule "$Q_i \rightarrow Q_j$" is calculated as follows:

$$\mathrm{conf}(Q_i \rightarrow Q_j) = \frac{\sup(Q_i,\ Q_j)}{\sup(Q_i)}, \tag{1}$$

where Q_i is a question in the large 1-itemset, Q_j is a question in the test paper, "$Q_i \rightarrow Q_j$" denotes the association rule from Q_i to Q_j, "$\mathrm{conf}(Q_i \rightarrow Q_j)$" denotes the confidence of the association rule "$Q_i \rightarrow Q_j$", "$\sup(Q_i,\ Q_j)$" denotes the support of the 2-itemset $(Q_i,\ Q_j)$, "$\sup(Q_i)$" denotes the support of the large 1-itemset $Q_i, i \neq j, 1 \leqslant i \leqslant m$, and $1 \leqslant j \leqslant m$.

Step 2: Construct two kinds of question-relationship maps based on the associated rules derived in Step 1. For the associated rules that the learner failed the question Q_i and then failed the question Q_j, build a relationship from question Q_i to question Q_j in the constructed "failure-to-failure