Chapter 7.10
A Data Mining Approach to Diagnosing Student Learning Problems in Science Courses

Gwo-Jen Hwang
National University of Tainan, Taiwan, R.O.C.

ABSTRACT

In recent years, researchers have attempted to develop more effective distance education systems. Nevertheless, students in network-based learning environments may need additional guidance and assistance when they encounter problems in learning certain concepts. Therefore, it is important to provide learning guidance in a distance learning environment. In this paper, we propose a data mining approach that is capable of assisting teachers to provide information needed for guiding students during the learning process. Several experiments on science courses have shown the effectiveness of applying the novel approach.

INTRODUCTION

In recent years, many researchers have applied computer techniques to the development of computer-assisted tutoring systems (Hopper, 1992; Hwang, 1998, 2002; Sun & Chou, 1996; Wong et al., 1998; Yoshikawa et al., 2000). Owing to the growing popularity of computer-assisted instructions, computer-based testing has received an increasing amount of attention (Wainer, 1990). For example, Jacob and Chase (1992) indicated that computers can employ more varied methods of presentation than paper-and-pencil tests, including 3-D diagrams, motion effects, rotating geometric forms, and so forth. Besides the traditional multiple-choice, fill-in-the-blank, and short-essay-type questions, Rasmussen et al. (1997) suggested that Web-based instruction also could allow students' progress to be evaluated through participation in group discussions and portfolio development. Furthermore, Khan (1997) suggested that designers of Web-based instruction systems could create facilities to allow students to submit comments on courseware design and delivery. Later, Chou (2000) presented the CATES system, a collective

and collaborative project intended to integrate an interactive testing system with theoretical and practical research on complex technology-dependent learning environments.

Although many researchers have identified testing and evaluation as important issues in computer-based instruction and have suggested design strategies and techniques, few systems have attempted to diagnose student learning problems. Most conventional testing systems assign a score or status indicator to each student after conducting a test, thus determining the learning status of that student but without considering how to improve the student's learning status. To cope with this problem, Hwang (2003) proposed a concept-effect relationship model to demonstrate how the learning status of certain concepts possibly can be influenced by the learning status of other concepts. After the test results of students were analyzed, based on concept-effect relationships, the students were given guidance in areas that needed improvement and on how to improve their learning status.

Although the concept-effect relationship model appears desirable, based on experimental results, its application is time-consuming for teachers unfamiliar with computer programming. Previous experience in applying the concept-effect relationship model to tutoring has revealed that most teachers require assistance to define concept-effect relationships (Hwang, 2003). Although Hwang et al. (2003) have proposed a straightforward algorithm to estimate the possible implication relationships among concepts, it is still time-consuming for the teachers to construct a complete set of concept-effect relationships.

To cope with this problem, this investigation proposes a data mining approach to assisting teachers in defining and analyzing concept-effect relationships. A testing and diagnostic system has been implemented, based upon this approach. For each student, the system can give learning suggestions by analyzing the answer sheets and the relationships between the subject concepts and the test items. The following sections present this novel approach along with practical experiences in applying it to several science courses. Experimental results have demonstrated the effectiveness of this novel approach in diagnosing student problems and improving their learning status.

RELEVANT RESEARCH

In many pedagogic and psychologic literatures, *conception* is defined as the common attributes of the same category and the objects or events that are given the same name (Ausubel, 1963, 1968; Ausubel et al., 1978). During tutoring, students learn new concepts and new relationships among previously learned concepts; this knowledge can be represented as a concept map (McAleese, 1998). Concept maps have been used in education, policy studies, and the philosophy of science to provide a visual representation of knowledge structures and argument forms. They provide a complementary alternative to natural language as a means of communicating knowledge (Shaw & Gaines, 1992).

Salisbury (1998) indicated that learning information (e.g., facts, names, labels, paired associations) is often a prerequisite to achieving efficient performance in a more complex, higher-level skill, especially in science courses. For example, the names and abbreviations of chemical elements and their atomic weights must be learned well in order to comprehend scientific writing or chemical formulas. That is, effectively learning a scientific concept normally requires learning some basic concepts first. Based on this notation, a special concept map model named concept-effect *relationship*, has been proposed for diagnosing student learning problems (Hwang, 2003). The following subsections describe how this model can be used to find the poorly learned and well-learned concepts of a student and how it generates learning guidance.