ABSTRACT

Student behavior analysis is an active research topic in distance education in recent years. In this article, we propose a new method called Boosting to investigate students’ behaviors. The Boosting Algorithm can be treated as a data mining method, trying to infer from a large amount of training data the essential factors and their relations that influence students’ academic successes. Based on the trained model, it is possible to predict students’ academic successes and to assist them to adjust their learning behaviors. Among the essential factors selected by the Boosting algorithm, the analysis and comparison are conducted between on-campus students and off-campus students. More importantly, these findings are of great importance to academic administrators, faculty members, and instructional developers in order to improve the teaching modes and online courseware design.

Keywords: behavior analysis; boosting; data mining; distance education

INTRODUCTION

Distance education is emerging as an increasingly important component of higher education (Lewis, 1997) because it provides opportunities for those who are unable to attend universities and for those who are unable to completely put their jobs aside to be on-campus students.

There are many significant differences between distance education and traditional face-to-face education. Traditional education is a centralized, local, and classroom-teacher-focused approach, whereas distance education is a decentralized, global, network-based, and student-focused approach. Distance education is more flexible in that students actively can select their favorite learning modes instead of passively accepting predetermined modes; thus, they can improve their learning efficiency. Our experiences also have shown that students in distance education behave quite differently from those in traditional education. Despite potential benefits of distance education, the effectiveness varies significantly, making it important to identify those factors that influence students’ academic successes. Thus, student
behavior analysis in distance education has become an active research topic in recent years.

Being aware of the importance of behavior analysis, many have done work on this topic. Wilson (2000) studied the relationship between individual characteristics and the use of computer-mediated communication systems (CMCS) in a team project situation. His study includes three general categories of potential determinants: demographic, experiential, and personality. The results show that CMCS will be adopted and used successfully by the same types of students who perform well in courses conducted via traditional face-to-face communication, while personality may influence academic success in unanticipated ways. Different with Wilson’s study, Tolmie et al. (2000) studied external factors that determine successful implementations, such as size of group, knowledge of other participants, student experience, clarity about task, ownership of task, need for system, type of system, and prior experience. Xie et al. (2001) compared the behavior differences between on-campus and off-campus students in order to facilitate improving the effect of instruction. The comparison was made from four areas: learning time, course browsing path, intercommunication, and adaptability toward online learning. Liu et al. (2003) investigated in the Web-based learning environment the behaviors of students in different semesters, from different areas (on-campus or off-campus) and under different instruction methods. Akahori et al. (2001) developed a 10-day training course in three different locations and investigated the factors influencing the effectiveness of training. The study shows that a Web-based training support system and CD-ROM materials are very effective in improving a learner’s knowledge and skills. In general, these analyses and comparisons are very useful for educators and educational designers who plan and conduct online learning courses.

It would be very useful to know the degree that each factor impacts distance education and the insight relations among different factors. However, most investigation approaches, including the aforementioned analyses and comparisons, are conducted manually by emphasizing different aspects (i.e., all the factors are selected, categorized, and analyzed manually by investigators). To address such issues, we introduce a machine learning algorithm called Boosting to address the problem of student behavior analysis in distance education. The main idea of boosting is to combine many simple and moderately inaccurate classifiers into a single, highly accurate classifier (Schapire, 1999). If each weak classifier depends only on a single feature, the boosting process, which selects a new weak classifier in each stage, can be viewed as a feature selection process (Viola, 2001). Thus, it can deal with both discrete-valued attributes and continuous-valued attributes and does not require each dimension of the training data to be normalized. This advantage is critical, because factors of student behaviors vary significantly in types and scales. Applied to distance education, the boosting algorithm can be treated as a data mining method, trying to infer from a large amount of training data the essential factors and their relations that influence students’ academic successes. Furthermore, based on the learned model, it may be possible to anticipate the students’ academic successes and to assist them to adjust their learning behaviors. More importantly, these findings are generally valuable in order for academic administrators and instructional developers to improve their teaching modes.

This article is organized as follows. In the second section, we will provide a brief introduction of the boosting algorithm and then present how to adapt it to the problem of student behavior analysis. In the third section, we will describe an online course that served as the experimental environment for the study. In the fourth section, we will present and discuss the results of behavior analysis based on the boosting algorithm. Finally, we will give concluding remarks in the fifth section.

**BOOSTING-BASED ANALYSIS APPROACH**

To better understand the proposed method, in this section we provide a very brief introduction to the boosting algorithm.