Computers in Human Behavior 36 (2014) 469–478

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Developing early warning systems to predict students' online learning performance

Ya-Han Hu [a,1], Chia-Lun Lo [a,1], Sheng-Pao Shih [*,b]

[a] Department of Information Management, National Chung Cheng University, 168, University Rd., Min-Hsiung Chia-Yi, Taiwan, ROC
[b] Department of Information Management, Tamkang University, No. 151, Yingzhuan Rd., Tamsui Dist., New Taipei City 25137, Taiwan, ROC

ARTICLE INFO

Article history:
Available online 7 May 2014

Keywords:
Learning management system
e-Learning
Early warning system
Data-mining
Learning performance prediction

ABSTRACT

An early warning system can help to identify at-risk students, or predict student learning performance by analyzing learning portfolios recorded in a learning management system (LMS). Although previous studies have shown the applicability of determining learner behaviors from an LMS, most investigated datasets are not assembled from online learning courses or from whole learning activities undertaken on courses that can be analyzed to evaluate students' academic achievement. Previous studies generally focus on the construction of predictors for learner performance evaluation after a course has ended, and neglect the practical value of an "early warning" system to predict at-risk students while a course is in progress. We collected the complete learning activities of an online undergraduate course and applied data-mining techniques to develop an early warning system. Our results showed that, time-dependent variables extracted from LMS are critical factors for online learning. After students have used an LMS for a period of time, our early warning system effectively characterizes their current learning performance. Data-mining techniques are useful in the construction of early warning systems; based on our experimental results, classification and regression tree (CART), supplemented by AdaBoost is the best classifier for the evaluation of learning performance investigated by this study.

1. Introduction

Recent technological innovations and the development of e-Learning platforms, such as web-based online-learning and multimedia technologies, not only overcome limitations of time and space, but also reduce learning costs. Educators can monitor students' learning processes by evaluating their learning portfolios using an learning management system (LMS) in an e-Learning environment (Macfadyen & Dawson, 2010). Through the use of an LMS, information about student learning behaviors and activities are retrieved from system logs or databases, and the data is then be analyzed by an early warning system. Educators can assess overall learning performances, and determine how well students are learning and what particular difficulties they might be having, so gaining insight into students that are at-risk of course failure (Campbell & Oblinger, 2007; Lust, Elen, & Clarebout, 2013). Past studies report that, logs of online activity and learner data stored in an LMS can be used to make forecasts of student future performance (Macfadyen & Dawson, 2010; Romero, Ventura, & García, 2008; Valsamidis, Kontogiannis, Kazanidis, Theodosiou, & Karakos, 2012). LMS prompts academically at-risk students to study more effectively, while educators can track the progress of their students and generate timely feedback (Kotsiantis, Pierrakeas, & Pintelas, 2004). However, the development of a precise LMS that can assess student learning performances using web-based learning portfolios is a challenging task. Data mining attempts to obtain valuable knowledge from data stored in large repositories; the strategy has been considered an appropriate method of knowledge discovery to excavate implicit information (Duan & Da Xu, 2012; Dunham, 2002). In the field of education, data mining is concerned with developing methods to explore the unique types of data that describe learners, and applies these methods to provide a better understanding of the learners, thus, data mining may reveal information that will benefit the learners. Recently, researchers used data mining techniques to analyze learning portfolios, make predictions, and to construct models of student learning performance (Macfadyen & Dawson, 2010). Dringus and Ellis (2005) argued that, from a systems view of LMS, given the textual nature of most asynchronous data held in systems, assessment and the acquisition of valuable information is hindered by limitations of the query and reporting toolset

* Corresponding author. Tel.: +886 2 2621 5656x2874; fax: +886 2 2620 9737.
E-mail addresses: yahan.hu@mis.ccu.edu.tw (Y.-H. Hu), allenlo.tw@gmail.com (C.-L. Lo), sbao@mail.tku.edu.tw (S.-P. Shih).
[1] Tel.: +886 5 272 0411; fax: +886 5 272 1501.

http://dx.doi.org/10.1016/j.chb.2014.04.002
0747-5632/

provided by the systems. LMS with data mining technique can successfully integrate online learning systems to further improve student academic performances (Hanna, 2004; Valsamidis et al., 2012; Wu, 2013).

Previous studies investigated the learning portfolios recorded by LMS systems to understand learner behavior (Gaudioso & Talavera, 2006), determine learning system effectiveness (Mostow et al., 2005; Yu, Jannasch-Pennell, Digangi, & Wasson, 1998; Yu & Wu, 2013), predict academically at-risk students (Essa & Ayad, 2012), or develop an early warning system to provide decision support system for instructors (Kotsiantis, 2012; Macfadyen & Dawson, 2010). However, most of the research dataset was obtained from traditional classroom settings, not from online learning courses; there is not much integrated LMS data that describes all activities undertaken by students during courses, which is available for analysis of academic performances (Hwang, Hsiao, & Tseng, 2003). To our knowledge, no study has predicted student learning performance using learning portfolio datasets when a fully online course is in progress. Therefore, time-dependent variables and other activities undertaken by students during the semester were included in our study, to predict student learning performance using data mining techniques. To identify *time-dependent variables* regarding the students' learning behaviors is a critical task in the development of an LMS. In this study, *time-dependent variables* denote those variables that varied during the learning activity processes (i.e., change with time). In this manner, the study addressed the following research questions:

- *How can data mining techniques accurately predict student learning performance based on activities in a fully online course?*
- *With the inclusion of time-dependent variables, how early in the semester can the early warning system accurately predict student learning performance?*
- *Which data mining technique offers superior predictive power regarding learning performance, when a fully online class is in progress?*

We analyzed a fully online undergraduate information literacy course offered at a national university in Taiwan during 2009. The analysis included 330 student learning portfolios. A "fully online" course refers to one in which all content delivery, communication and assessment is carried out via the LMS (Macfadyen & Dawson, 2010). We evaluated learning portfolio data, obtained while the class was in progress, as our dataset, and built a prediction model to predict at-risk students. To understand the effects of time-dependent variables on academic performance, we collected online course time-dependent variables to build reliable prediction models for an early warning system using C4.5 (Quinlan, 1993), classification and regression tree (CART) (Breiman, Friedman, Stone, & Olshen, 1984), logistic regression (LGR) (Sumner, Frank, & Hall, 2005), and adaptive boosting (AdaBoost) (Freund & Schapire, 1996) as part of our data mining strategy.

The remainder of this study is organized as follows. In Section 2, we review previous studies on the development of LMS and EDM research. Section 3 explains the proposed classification techniques used to build the early warning model and data collection procedure. Section 4 describes the experimental results of classification systems. Section 5 presents the system development and evaluation. Section 6 concludes the study.

2. Literature review

2.1. Learning portfolio and LMS

Learning portfolios are a collection of the events and learning activities undertaken by learners. Building a portfolio is a flexible, evidence-based process that combines reflection with documentation, and encourages student engagement in ongoing and collaborative analysis of learning to provide purposeful and selective outcomes for both improving learning outcomes and assessing the learning process (Zubizarreta & Millis, 2009). Traditionally, a learning portfolio relies on manual data collection and a writing-centered learning process. Difficulties in data storage, searching, and management are obstacles to the development and implementation of learning portfolio evaluation. By contrast, a web-based learning portfolio can be automatically collected, stored, and managed by LMS when learners interact with an e-Learning management system. Consequently, there has been a significant research effort into learning performance assessment using a web-based learning portfolio approach (Rasmussen, Northrup, & Lee, 1997). Previous studies revealed how a learning portfolio can assist instructors in correlating learner behaviors with their learning performance (Agrawal & Srikant, 1994; Hanna, 2004; Macfadyen & Dawson, 2010; Sadler-Smith, 2001). Drawing on the work, and building on a study conducted by Wang and Newlin (2002), researchers identified a strong relationship between LMS usage and learning performance (Campbell & Oblinger, 2007; Goldstein & Katz, 2005). Campbell and Oblinger (2007) further proposed that educators can directly benefit from the analysis of LMS data and development of an early warning system, by identifying at-risk students and implementing early intervention strategies. Wang, Newlin, and Tucker (2001) investigated student behaviors using LMS, and argued that online learning activities can provide an early warning index of student academic achievement. However, among these studies, there are few studies into the effectiveness of LMS-based early warning systems.

2.2. Educational data mining research

The analysis of student usage activity recorded in LMS is becoming increasingly important to EDM (Baker, 2010; Baker & Yacef, 2009). The objectives of these EDM studies include: understanding learner behavior (Chang, Kao, Chu, & Chiu, 2009; Gaudioso & Talavera, 2006), determining the effectiveness of learning systems (Mostow et al., 2005), identifying academically at-risk students (Essa & Ayad, 2012), and developing an early warning system and decision support system for instructors (Gaudioso, Montero, & Hernandez-Del-Olmo, 2012; Kotsiantis, 2012; Macfadyen & Dawson, 2010).

Castro, Vellido, Nebot, and Mugica (2007) reviewed EDM research conducted from 1999 to 2006 and concluded that most EDM studies dealt with Classification and Clustering problems associated with online platforms. The techniques used by the reviewed studies varied with research hypothesis and data characteristics. We reviewed EDM reports related to our study, which were published during the past decade (as shown in Table 1).

The research issues broadly divide into two categories. The first category includes reports that make predictions about online test performances, for example, Anozie and Junker (2006) provided an online examination platform that could timely and effectively predict an examination result. Kotsiantis et al. (2004) predicted student performances under a distance learning system by employing ML techniques. Guruler, Istanbullu, and Karahasan (2010) used classification techniques to identify individual student characteristics and their association with future success. The second category includes studies that make predictions about student learning performance through various feature sets. For example, Muehlenbrock (2005) used decision tree (DT) techniques to help students using e-Learning systems to further improve their e-Learning performance. Etchells, Nebot, Vellido, Lisboa, and Mugica (2006) used Fuzzy Inductive Reasoning (FIR) and Orthogonal Search-Based Rule Extraction (OSRE) techniques to construct a