Available online at www.sciencedirect.com
ScienceDirect
Computers & Education 51 (2008) 368–384
COMPUTERS & EDUCATION
www.elsevier.com/locate/compedu

Data mining in course management systems: Moodle case study and tutorial

Cristóbal Romero *, Sebastián Ventura, Enrique García

Department of Computer Sciences and Numerical Analisys, University of Córdoba, 14071 Córdoba, Spain

Received 5 March 2007; received in revised form 19 May 2007; accepted 25 May 2007

Abstract

Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational context. This work is a survey of the specific application of data mining in learning management systems and a case study tutorial with the Moodle system. Our objective is to introduce it both theoretically and practically to all users interested in this new research area, and in particular to online instructors and e-learning administrators. We describe the full process for mining e-learning data step by step as well as how to apply the main data mining techniques used, such as statistics, visualization, classification, clustering and association rule mining of Moodle data. We have used free data mining tools so that any user can immediately begin to apply data mining without having to purchase a commercial tool or program a specific personalized tool.

Keywords: Distance education and telelearning; E-learning; Evaluation of CAL systems; Data mining; Web mining

1. Introduction

Course management systems (CMSs) can offer a great variety of channels and workspaces to facilitate information sharing and communication among participants in a course. They let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums, chats, file storage areas, news services, etc. Some examples of commercial systems are Blackboard (BlackBoard, 2007), WebCT (WebCT, 2007) and TopClass (TopClass, 2007) while some examples of free systems are Moodle (Moodle, 2007), Ilias (Ilias, 2007) and Claroline (Claroline, 2007). Nowadays, one of the most commonly used is Moodle (modular object oriented developmental learning environment), a free learning management system enabling the creation of powerful, flexible and engaging online courses and experiences (Rice, 2006).

These e-learning systems accumulate a vast amount of information which is very valuable for analyzing students' behaviour and could create a gold mine of educational data (Mostow & Beck, 2006). They can record

* Corresponding author. Fax: +34 957 218630.
E-mail address: cromero@uco.es (C. Romero).

0360-1315/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2007.05.016

any student activities involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers (Mostow et al., 2005). They normally also provide a database that stores all the system's information: personal information about the users (profile), academic results and users' interaction data. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually. Instructors and course authors demand tools to assist them in this task, preferably on a continual basis. Although some platforms offer some reporting tools, it becomes hard for a tutor to extract useful information when there are a great number of students, (Dringus & Ellis, 2005). They do not provide specific tools allowing educators to thoroughly track and assess all learners' activities while evaluating the structure and contents of the course and its effectiveness for the learning process (Zorrilla, Menasalvas, Marin, Mora, & Segovia, 2005). A very promising area for attaining this objective is the use of data mining (Zaïane & Luo, 2001).

In the last few years, researchers have begun to investigate various data mining methods to help instructors and administrators to improve e-learning systems (Romero & Ventura, 2006). Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections (Klosgen & Zytkow, 2002). Data mining is a multidisciplinary area in which several computing paradigms converge: decision tree construction, rule induction, artificial neural networks, instance-based learning, bayesian learning, logic programming, statistical algorithms, etc. And some of the most useful data mining tasks and methods are statistics, visualization, clustering, classification and association rule mining. These methods uncover new, interesting and useful knowledge based on students' usage data. Some of the mains e-learning problems or subjects to which data mining techniques have been applied (Castro, Vellido, Nebot, & Mugica, in press) are dealing with the assessment of student's learning performance, provide course adaptation and learning recommendations based on the students' learning behaviour, dealing with the evaluation of learning material and educational web-based courses, provide feedback to both teachers and students of e-learning courses, and detection of atypical student's learning behaviour.

Nowadays, data mining tools are normally designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex for educators to use and their features go well beyond the scope of what an educator might require. As a result, the CMS administrator is more likely to apply data mining techniques in order to produce reports for instructors who then use these reports to make decisions about how to improve the student's learning and the online courses.

This knowledge, however, can be useful not only to the providers (educators) but also to the users themselves (students), as it can be oriented towards different ends for different partakers in the process (Zorrilla et al., 2005). It could be oriented towards students in order to recommend learners' activities, resources, suggest path pruning and shortening or simply links that would favor and improve their learning or to educators in order to get more objective feedback for instruction. Instructors can evaluate the structure of course content and its effectiveness in the learning process and also classify learners into groups based on their needs for guidance and monitoring. Learners' regular and irregular patterns could be determined allowing the most frequently made mistakes to be identified and more effective activities to be elaborated. There could be more orientation towards obtaining parameters and measures to improve site efficiency and adapt it to the behaviour of the users (optimal server size, network traffic distribution, etc.) and to organize better institutional resources (human and material) and educational offer.

Data mining has been applied to data coming from different types of educational systems. On one hand, there are traditional face-to-face classroom environments such as special education (Tsantis & Castellani, 2001) and higher education (Luan, 2002). On the other, there is computer-based education as well as web-based education such as well-known learning management systems (Pahl & Donnellan, 2003), web-based adaptive hypermedia systems (Koutri, Avouris, & Daskalaki, 2005) and intelligent tutoring systems (Mostow & Beck, 2006). The main difference between one and the other is the data available in each system. Traditional classrooms only have information about student attendance, course information, curriculum goals and individualized plan data. However, computer and web-based education has much more information available because these systems can record all the information about students' actions and interactions onto log files and databases.

This paper is oriented to the specific application of data mining in computer-based and web-based educational systems (in particular, course management systems). It is arranged in the following way: Section 2 describes the general process of applying data mining to e-learning data, especially to Moodle usage