



Predicting students' final performance from participation in on-line discussion forums



Cristóbal Romero*, Manuel-Ignacio López, Jose-María Luna, Sebastián Ventura

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

ARTICLE INFO

Article history:

Received 17 January 2013

Received in revised form

16 May 2013

Accepted 17 June 2013

Keywords:

Distance education and telelearning

Asynchronous discussion forums

Social network analysis

Data mining

Predicting performance

ABSTRACT

On-line discussion forums constitute communities of people learning from each other, which not only inform the students about their peers' doubts and problems but can also inform instructors about their students' knowledge of the course contents. In fact, nowadays there is increasing interest in the use of discussion forums as an indicator of student performance. In this respect, this paper proposes the use of different data mining approaches for improving prediction of students' final performance starting from participation indicators in both quantitative, qualitative and social network forums. Our objective is to determine how the selection of instances and attributes, the use of different classification algorithms and the date when data is gathered affect the accuracy and comprehensibility of the prediction. A new Moodle's module for gathering forum indicators was developed and different executions were carried out using real data from 114 university students during a first-year course in computer science. A representative set of traditional classification algorithms have been used and compared versus classification via clustering algorithms for predicting whether students will pass or fail the course on the basis of data about their forum usage. The results obtained indicate the suitability of performing both a final prediction at the end of the course and an early prediction before the end of the course; of applying clustering plus class association rules mining instead of traditional classification for obtaining highly interpretable student performance models; and of using a subset of attributes instead of all available attributes, and not all forum messages but only students' messages with content related to the subject of the course for improving classification accuracy.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Internet forums, web forums, discussion boards, message boards, discussion groups, or bulletin boards are on-line discussion sites where people can hold conversations in the form of posted messages. They are one of the most popular tools for supporting students' communication and collaboration in web-based teaching-learning environments and one of the best ways of sharing ideas, posting problems, commenting on posts by other students, and obtaining feedback (Raghavan, Catherine, Ikbal, Kambhatla, & Majumdar, 2010). An on-line discussion forum, being an asynchronous tool, allows students to maintain discussions related with their learning process at any time and place, to have more time to structure and organize their thoughts, and to communicate simultaneously or even participate in multiple discussions at the same time (Cobo et al., 2011). Forums also enable student interactions and social interactions to occur outside a traditional classroom, and students, especially those of an introvert character, feel less threatened about expressing their views or asking questions (Cheng, Paré, Collimore, & Joordens, 2011). Forums play an important role in students' collaborative learning, in which two main actions are carried by students: writing and reading (Rau, Gao, & Wu, 2008). The students can adopt different attitudes in terms of these two main actions, thus defining different behaviour profiles such as active learners, lurkers, etc. and different type of discussions such as announcement, questioning, clarification, interpretation, conflict and assertion (Rabbany et al., 2011). On the other hand, a synchronous forum allows instructors to analyze student participation in forums, to develop monitoring and even assessment tasks (Lopez, Luna, Romero, & Ventura, 2012). However, the instructor's view of the output of a threaded forum can normally be limited to reviewing a

* Corresponding author. Fax: +34 957 218630.

E-mail address: cromero@uco.es (C. Romero).

transcript or printed version of the written dialogue produced by students. With hundreds of contributions to review in an entire on-line forum, the instructor lacks a comprehensive view of the information embedded in the transcript. In fact, even in on-line courses with a small number of students, there could be thousands of messages generated in a few months using these types of forums. The instructor is faced with the difficulty of interpreting and evaluating the learning and quality of the participation reflected in the students' contributions. Evaluating the participation of students in such a case is very difficult, considering that current e-learning environments do not provide many indicators or information regarding the structure of interactions between students (Dringus & Ellis, 2005). A solution to this problem is the use of quantitative and qualitative participation indicators in forums, such as the number of contributions and the quality of the contributions, together with Data Mining (DM) for discovering and building alternative representation and models for the data underlying discussion forums. DM is a knowledge discovery process to extract information from a dataset and to transform it into an understandable structure for further use (Klogsen & Zytchow, 2002) and is known as EDM (Educational Data Mining) when it uses the unique kinds of data that come from an educational setting (Romero & Ventura, 2013). EDM is unobtrusive or archival research that uses educational data collected in the past to reach a finding and thus the researcher does not interact directly with subjects of the study (Denning, 2008). In this paper, we use an authentic educational setting – a first-year course in Computer Science with 114 university undergraduates – in which, in parallel to the traditional classroom, all students participate in an on-line discussion forum. Therefore, there is no pre-post assessment before-after intervention or treatment, because there is no such intervention or treatment (McMillan & Schumacher, 2006). We propose to use different data mining approaches to predict whether students will pass the course or not, starting from participation in an on-line discussion forum. We have also developed a new Moodle's module for gathering all the forum indicators. The objective of our study is to find a reply to the following research questions:

1. What DM techniques are best for predicting student performance starting from participation in on-line forums? Traditionally, supervised classification algorithms have been used to do this task. However, other unsupervised DM algorithms, such as clustering algorithms, can also be used for the task and provide similar accuracy and comprehensible models.
2. What attributes are the best predictors? Are all attributes or available variables about forum usage data relevant or can similar accuracy be obtained by using just one subgroup of selected attributes?
3. What messages are the best predictors? Are all instances or available messages relevant or perhaps some messages may be irrelevant to the contents of the course and a better prediction can be obtained by using only those messages related to the course.
4. It is possible to make an early prediction? Is it necessary to wait until the end of the course in order to obtain good prediction accuracy or it is possible to do it before the end of the course?

The paper is arranged in the following way. Firstly, the related background is described. Next, the proposed approach is presented together with the material and methods used. Then, the experimental results and discussion are described. Finally, conclusions and future work are dealt with.

2. Background

Predicting students' performance is one of the oldest and most useful applications of EDM and its goal is to estimate the unknown value of students' performance, knowledge, score or mark from other information, aspects or behaviour of those students (Romero & Ventura, 2013). This is a difficult problem to solve due to the large number of factors or characteristics that can bear influence on students' performance, such as demographic, cultural, social, or family factors, socio-economic status, psychological profile, previous schooling, prior academic performance, interactions between student and the faculty, etc. (Araque, Roldan, & Salguero, 2009). Different techniques have been applied for predicting students' performance depending on the variable to be predicted (Hämäläinen & Vinni, 2011): classification (when the predicted variable is a categorical value), regression (when the predicted variable is a continuous value) or density estimation (when the predicted value is a probability density function). It is also important to notice that most of the current research on the application of EDM for predicting student performance has been applied primarily to the specific case of higher education or university students (Kotsiantis, Patriarchas, & Xenos, 2010) and more specifically to web-based education or e-learning (Romero, Espejo, Romero, & Ventura, 2013). This is mainly due to the fact that the use of Learning Management System (LMS) such as Moodle, Ilias, Claroline, Atutor, Blackboard, WebCT, TopClass, etc. is increasing exponentially due to its ability to create powerful, flexible and engaging on-line courses and experiences. These systems accumulate a vast amount of information in relation to visits and times, resources viewed, assessments, and activities in chat rooms and on-line forums, which is very valuable in analyzing students' behaviour, predicting performance and assisting courseware authors in detecting possible errors, shortcomings and improvements (Romero, Ventura, Espejo, & Hervás, 2008; Romero, Ventura, & García, 2008). One of the most popular LMS activities are on-line discussion forums, because they are a promising strategy for collaboration and higher-order thinking. Some researchers have also detected the predictive relationship between discussion participation and learning (Koschmann, Kelson, Feltovich, & Barrows, 1996). While agreement exists that participation in on-line discussions can enhance student learning, it has also been identified that there is an increasing need to investigate the impact of participation on student course performance (Palmer, Holt, & Bray, 2008). Some examples are: Alstete and Beutell (2004) found that, in on-line courses, the strongest indicator of student performance was the use of discussions forums; Patel and Aghayere (2006) found that forum participation was positively correlated with the student's course grade in two undergraduate civil engineering courses; Shaw (2012) proved that a web-based programming language learning course supported with on-line forums and active student participation increases learning performance as measured by student learning scores; and Cheng et al. (2011) found that students who participated in a forum tended to have better performance in an introductory psychology course. Most of these approaches only use quantitative information such as message frequencies (the number of initiations and replies, the number of messages read, thread lengths, and response time from previous messages) to correlate them with course grades by stepwise multivariate linear regression analysis (Palmer et al., 2008), or only use the number of posts and the number of page views to find relations between forum participation and course performance by regression analysis (Cheng et al., 2011); or use the frequency of access and the duration of sessions to establish several categories of learners by cluster analysis, which depict the differences among the cohort in terms of participation (Khan, Clear, & Sajadi, 2012); or use agglomerative hierarchical clustering for

Download English Version:

<https://daneshyari.com/en/article/348499>

Download Persian Version:

<https://daneshyari.com/article/348499>

[Daneshyari.com](https://daneshyari.com)