Student Behavior Patterns in a Virtual Learning Environment

Priscila Valdiviezo
Utniversidad Técnica Particular de Loja, Departamento de Ciencias de la Computación y Electrónica, Loja, Ecuador, pmvaldiviezo@utpl.edu.ec

Ruth Reátegui
Utniversidad Técnica Particular de Loja, Departamento de Ciencias de la Computación y Electrónica, Loja, Ecuador, rmreategui@utpl.edu.ec

Marcia Sarango
Utniversidad Técnica Particular de Loja, Titualción de Sistemas Informáticos y Computación, Loja, Ecuador, mysarangox@utpl.edu.ec

ABSTRACT

This work focuses on the identification of student behavior patterns obtained from their interactions on a virtual learning Environment (VLE). Clustering techniques were used to classify certain indicators and to obtain groups of students with similar characteristics. The activities performed are directly related to four Computer Science degree courses in the Distance Education modality. Generally, our results show that students interacted more with online forum, followed by the quiz, tasks, instant messaging, resources, and twitter. The knowledge acquired via the data mining techniques helped to discover certain characteristics of their online interaction, which should be taken into account when enhancing the teaching-learning process.

Keywords: Clustering, Virtual Learning Environments, Data mining, Distance Education, E-learning.

1. INTRODUCTION

Data mining techniques play an important role in E–learning since they contribute to the discovery of information about user interaction with web-based systems. This information can be used by the tutor to improve learning and to develop students' interaction via web platforms. In addition, data mining helps to identify key problems related to the usage of the system and the usage of online tools.

The main applications of data mining techniques in the context of distance or e-learning education are personalized systems, recommending systems, and the detection of irregularities, among others.

In terms of the benefits provided by the above techniques, the focus of the present study is to identify patterns of behavior related to student interactions via the Virtual Learning Environment (VLE) at the Universidad Técnica Particular de Loja (UTPL). In order to do this, a preliminary analysis of the corresponding database was carried out. Essentially, this consisted of selecting components with information about the various kinds of online activities performed by students. After extracting the data about the students' interaction with the VLE, a clustering data mining technique was used to identify students with similar characteristics. This was done to assess their course participation, and their usage of the available tools on the VLE. We shall now look at how data mining is use in E-learning.

This article begins with a review of data mining in e-learning that was support with some previous studies related with this research. In addition we showed the methodology used in this work, and the application of data mining algorithms to determine the behaviors of students. Finally, we showed the conclusions of this research.

2. DATA MINING IN E-LEARNING

According to Hand, et al. (2001) data mining is defined "as a process that consists of a set of tools of different sciences such as (Statistics, Computer Science, Mathematics, Engineering, etc.)", which aims to extract hidden knowledge or non- trivial information from large volumes of data, with the goal of providing solutions to specific problems. Educational Data Mining (EDM) uses data extraction from education systems to analyze students' patterns of behavior (Baker, 2010). For this reason, the analysis of 'student learning' is recommended for improving current educational practices.

In the educational field, data mining helps teachers and educational institutions to use patterns and trends from student interaction with the aim of improving both the learning process and the quality of education (Romero et al., 2008). These data include the students' use of interactive learning environments, computer-supported collaborative learning, or data from schools and universities. Moreover, the data often have multiple levels of meaningful hierarchy, which need to be determined by properties in the data itself, rather than using previous data. Issues of time, sequence, and context also play important roles in the study of educational data (International Educational Data Mining Society, s.f).

There are certain tasks in the educational field which can be done with the help of data mining techniques: data analysis and visualization, information support for instructors, student counseling/guidance, predictions of student performance, student behavior modeling, detection of undesirable behavior, student grouping, social network analysis, and so on (Romero and Ventura, 2010).

In general, web-based educational systems have a lot of information recorded in log files, for example, interactions between students and the online learning systems, details of student successes and failures, student grades, and knowledge levels. Data mining applications differ from other applications because of their effect on the field of teaching. In other words, these applications aim to improve the learning process– something which is traditionally subjective and difficult to measure (Romero et al., 2008).

To sum up, we could say that data mining is used to look for new patterns and to develop new algorithms and models. In this way, it primarily focuses on technical aspects.

3. RELATED WORKS

In order to find behavioral patterns of students, namely those which could be used as student models, various data mining techniques have been applied to the data that was collected for the analysis of educational systems. These techniques are explained in more detail below.

Some activities and results that should be considered when analyzing the student's interaction with the system are: tasks performed, the order and time in which the activities were carried out, and the percentage of exercises performed correctly (Peredes and Rodríguez, 2004). In Brusilovsky and Millan (2008) focused on five features when viewing the user as an individual: the user's knowledge, interest, goals, background and individual traits.

Other research (Petrushyna et al., 2011) analyzed data about interaction in forum that are used to support students who are learning English. The authors sought to recognize among other things patterns of student behavior and their role in the learning community. Also, in Blikstein (2011) describes an automatic technique to evaluate, analyze, and visualize students who learn computer programming. He works with snapshots of code in a programming subject, and uses different quantitative techniques to obtain information about student behavior, and finally classifies this information in terms of their experience of programming. In Anaya and Boticario (2009), data mining techniques are applied to statistical indicators of student interactions in the forums of a VLE, to obtain information about the group collaboration.

Talavera and Gaudioso (2004) discovered patterns by reflecting user behaviors in collaborative spaces. The main characteristic used to form groups was the parameter of the students' answers related to their interests and abilities. Likewise in Bouchet et al. (2012), analyzed trace data to identify patterns of behavior in a study of 51 college students learning about a complex science topic with an agent-based Intelligent Tutoring Systems (ITS). With the Expectation-Maximization clustering algorithm, three distinct groups of students were observed, who were distinguished by their test and quiz scores, their learning gains, the frequency of their note-taking, their note-checking, the proportion of sub-goals attempted, and the time spent reading. Moreover, they employed a differential sequence mining technique to identify differentially frequent activity patterns between the student groups and they interpreted these patterns in terms of relevant learning behaviors.

Furthermore, in Perera et al. (2008) identified groups of students using a clustering algorithm in a senior software development project, namely where students used the collaboration tool 'TRAC'. The authors extracted patterns by distinguishing between stronger and weaker groups, and also made observations about the success factors. The results highlighted the importance of leadership and group interaction. Patterns indicating good individual practices were also identified.

4. METHODOLOGY

Data that were considered and included in this study were obtained from students' interactions in the VLE based on Moodle. We worked with four different courses from the Computer Science major in the distance education modality at the Universidad Técnica Particular de Loja during the study period April - August 2011. Based on this information we were able to identify different student behaviors and different indicators that reflected the students' usage of the virtual platform.

The population studied in this research comprised 388 students from the following courses: Programming Logic (205), Discrete Mathematics (142), Artificial Intelligence (18), Seminar II (23). With the data we were able to later identify the actions of students and the necessary variables for the application of data mining. This was achieved by analyzing their online activities.

During this process, we evaluated student characteristics, which included the analytical and informational processes that were necessary for the application of data mining algorithms. Among the various aspects that we analyzed were data about students´ interaction via the virtual platform, which was based on the analysis of log registers.

5. DATA MINING TO ANALYZE STUDENT LEARNING BEHAVIOUR

Drawing on the VLE database, entities were selected that held the most useful information about students' actions performed on the virtual platform. In order to determine the students' online activity on the VLE during their studies, as well as their attributes, we performed an analysis of data extracted from the log table of the database. As a result, a series of concrete attributes were established, namely those which represented students' interaction with the system, which are the base for identifying the behaviors of students.
These attributes are described in Table 1, together with the key actions that students performed on the platform, and the modules on which these actions were performed.

Table 1. Attributes of students interactions in the VLE

Module	Action	Description of attributes
COURSE	VIEW	Number of hits to the selected course.
ASSIGNMENT	VIEW	Number of times the user accesses to the jobs submitted by the teacher.
FORUM	ADD_POST	Number of times the user sends a post to a forum.
FORUM	VIEW_DISCUSSION	Number of times the user checks the discussions within a forum.
FORUM	VIEW_FORUM	Number of times the user accesses or reviews the discussion forums.
FORUM	UPDATE_POST	Number of times the user updates the post that was sent to the forum.
FORUM	ADD_ DISCUSSION	Number of times the user adds a topic of discussion or debate.
RESOURCE	VIEW	Number of times the user accessed or downloaded resources uploaded by the teacher.
USER	UPDATE	Number of times the user updates their profile data.
USER	VIEW	Number of times the users check or access to their profile.
UPLOAD	UPLOAD	Number of times the user uploads a task.
MESSAGE	WRITE	Number of messages the user wrote.
MESSAGE	HISTORY	Number of visits to the messages recorded.
QUIZ	VIEW	Number of times the user checks or accesses to the questionnaires.
QUIZ	TIMESTART, TIMEFINISH	Time used to resolve a questionnaire.
QUIZ	ATTEMPT	Number of times the user tries to solve a questionnaire.

5.1 EXPERIMENTS WITH ALGORITHMS

Within the framework of the student activities done via the virtual platform, two key indicators were established:

- Course participation: representing a student's interaction in a particular course by measuring the contribution they have made to the platform. Data can subsequently be used to recommend alternative content to students, especially those who demonstrate low motivation. Furthermore, it provides an opportunity for recommending more complex content. In this case, the more complex material would be recommended for those students who demonstrated a high level of participation.

- Usage of online tools: To address the students' action and usage of online forum, instant messaging, online resources, distance tasks, usage of twitter, tests, etc. Table 2 provided details of the actions performed using each VLE tool.

Table 2: Actions performed within the tools linked to the courses

Tool	Action
Forum	View forum
	Add post
	See discussion in the forum
	Update post
	Add a topic of discussion or debate
Resources	view_resourse
Tasks	Display tasks proposed by professor
	Send, or upload tasks
Twitter	Send Twitter message
Messages	Write message
	View message history
Quiz	Answer quiz
	View questions

The two indicators were measured using three criteria:

- Permanent (P), referring to a high level of participation and interaction with the tools.
- Moderate (M), referring to a medium level which includes both interaction and usage of online tools.
- Low (E), referring to minimum values (low access and minor usage of tools) during the course.

In order to assess the students' online activity based on their interactions with the VLE at UTPL, to establish online work groups, and to determine who made full use of the tools provided by the platform, the Cluster K-Means Algorithm was used to determine the students' level of participation in the course. When utilizing this technique, we focused mainly on a specific group of students and their interaction patterns. In this way, we could recommend actions and resources for those users with similar characteristics.

Although there are other clustering algorithms such as EM, which assigns a distribution probability for each cluster, we decided that we should perform the experiment with the K-Means Algorithm -- as it is one of the most widely used methods for this type of research. Also, it has been widely proven that is gives more accurate results for similar experiments (Valdiviezo et al., 2010).

5.2 RESULTS AFTER APPLYING THE ALGORITHMS

For data mining processing, the WEKA tool was used. Three experiments were performed with the K-Means Algorithm in order to obtain the best possible result. With this algorithm, we were able to evaluate each attribute of the indicators.

For all three experiments, three groups of students were classified according to the activities that were performed, which was based on the following levels: Low (E) for minimum values; Moderate (M) for medium values; and Permanent (P), for high range values.

In the third experiment, we noticed more consistent results. In addition, there were similar groups with minimum, intermediate and high values of participation, as well as a high usage of online tools. There was also greater similarity between the distribution of activities and lower values in the addition of square errors. In this way, the third experiment was the most adequate means to gauge the group results.

According to the results from the third experiment in cluster 0 (group 1), students who presented a high level (permanent) of participation in the course were grouped together. This could be seen by the high-level of access to the courses and the continual usage of the online tools.

Characteristics obtained of groups

In this section shows the interpretation and results obtained for each group, considering the indicators listed above and the ideas of Santos and Boticiario (2004) that manifest that the students are characterized as: Participatory , which measures activity in the different tools of the VLE, and not collaborative, which refers to students that contribute little in the forums and on twitter.

- **Group 1 -** Cluster 0: This group comprises students who have higher participation access and replies on VLE forums, which implies that the more a student accesses a course, the greater the probability that they will respond, update, view, add a discussion and/or review an online topic. With regard to the downloading of resources, the students who access the courses more frequently also downloaded resources more frequently. In the case of distance tasks or homework, the more frequently students accessed the course, the greater the number of tasks they uploaded onto the system. With regard to messages, the students who frequently accessed the system also had higher values concerning the writing of messages and the reviewing of the history of messages. In relation to twitter, it could be observed that students with more frequent access to the courses send more tweets. Finally, concerning access to VLE quizzes, the more these were reviewed, the greater the number of attempts were made to solve them. In light of the above, we could conclude that this is a participative and collaborative group with a high-level (permanent usage) of online tools (forums, resources, tasks, messaging, twitter, and quizzes).

- **Group 2 -** Cluster 1: Comprising students with a medium level of access to courses. In terms of online tools, we determined that this group made an intermediate number of responses to the forum, and that the frequency that they downloaded resources was rated as average. Moreover, the downloading of materials depended mainly on the regularity of access to the VLE during the course, i.e. in terms of messaging, twitter and uploading homework tasks. In addition, the results showed their online tool usage was average. This group also showed average values in the number of attempts they made at answering quizzes; however, they also showed high values for the amount of time they took to take the quiz. So in this cluster, the level of tool usage (forum, resources, tasks, messaging, twitter, and questionnaire) is medium (moderate).

- **Group 3 -** Cluster 2: This group comprises students with a low frequency of access to their courses. Their interactions in terms of replying to forums, downloading resources, uploading files, sending homework-tasks, messages, tweets, and accessing tests, was lower than the other groups. Therefore, they comprised a group of students who rarely used online tools, which means that they were a group with a low level of online participation and collaboration.

As one can see, there were a group of students with a low level of participation and interaction with the virtual platform and low usage of the online tools. This clearly shows that the resources available on the virtual platform were not fully explored by the students in these courses.

1. CONCLUSIONS

Applying data mining techniques to the data obtained from student interactions via the VLE allowed us to determine how effectively these tools contributed to education. In addition, they enabled us to discover how these tools were used by the students in their courses.

In this paper, the results obtained by applying the clustering algorithm to the three groups indicated that the greatest level of student interaction was in forum, followed by quizzes, online tasks, instant messaging, the usage of online resources, and twitter. Twitter was the tool with the least amount of interaction because it addresses social and informal topics rather than academic issues (unlike other tools on the platform).

We also observed a group of students who were at risk of not completing the course or degree subject. Similarly, some students might fall behind in their academic studies. By studying these kinds of problems, professors or tutors could make more informed decisions about their teaching practices on VLE. What's more, it would give them an opportunity to look for educational alternatives so that students could interact more effectively with the online tools, as well as increase their level of participation, for example, by providing more feedback, giving opportune answers to students, uploading more resources that would foster the learning process, and finally, proposing tasks that would promote increased collaboration among student peers, etc.

REFERENCES

Anaya, A. R., Boticario, J. G. (2009). "Clustering Learners according to their Collaboration". *Proceedings Computer Supported Cooperative Work in Design (CSCWD 2009)*, pp. 540–545, April 2009.

Baker, R.S.J.d. et al. (2010). "Contextual Slip and Prediction of Student Performance after Use of an Intelligent Tutor". *Proceedings UMAP*, pp. 52-63.

Blikstein, P. (2011). "Using learning analytics to assess students' behavior in open-ended programming tasks". *Proceedings I Learning Analytics Knowledge Conference (LAK 2011).* pp. 110—116, Feb. 2011.

Bouchet, F., Azevedo, R., Kinnebrew, J.S., Biswas, G. (2012). "Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning". *Proceedings 5th International Conference on Educational Data Mining (EDM 2012)*, pp. 56-72. June 2012.

Brusilovsky, P., Millan, E. (2008). "User Models for Adaptive Hypermedia and Adaptive Educational Systems". *In The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science*, P. Brusilovsky, A. Kobsa and W. Neidl, Eds., New York: Springer-Verlag, , Vol. 4321, pp. 3-53.

Hand, D.J., Mannila, H., and Smyth, P. (2001). "Principles of Data Mining". *MIT Press*. Cambridge, MA. USA.

International Educational Data Mining Society, http://www.educationaldatamining.org/ [05.01.2013]

Peredes, P., Rodríguez, P. A. (2004). "A mixed approach to modelling learning styles in adaptive educational hypermedia". *Proceedings IASTED Conference on Web-Based Education (WBE 2004)*, pp. 16-18. February 2004.

Perera, D., Kay, J., Yacef, K., Koprinska, I., and Zaiane, O. (2008). "Clustering and Sequential Pattern Mining of Online Collaborative Learning Data". *IEEE Trans. Knowledge and Data Eng.*

Petrushyna, Z., Kravcik, M., Klamma, R. (2011). "Learning analytics for communities of lifelong learners: a forum case". *Proceedings In: 11th IEEE International Conference on Advanced Learning Technologies*, pp. 609-610.

Romero, C., Ventura, S., and Garcia, E. (2008). "Data mining in course management systems: Moodle case study and tutorial". *Computers & Education*, Vol. 51, pp. 368–384.

Romero, C.,Ventura, S. (2010). "Educational data mining: a review of the state of the art". *IEEE Transactions on Systems*, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40, pp. 601–618. Nov. 2010.

Santos, O., Boticario, J. (2004). “Supporting a collaborative task in a web-based learning environment with Artificial Intelligence and User Modelling techniques”. *In: 6º SIIE’04.* Caceres, Spain.

Talavera, L., Gaudioso, E. (2004). “Mining student data to characterize similar behavior groups in unstructured collaboration spaces”. *In Workshop on artificial intelligence in CSCL*, pp. 17-23.

Valdiviezo, P. M., Santos, O. L., Boticario, J. G. (2010). “Aplicación de métodos de diseño centrado en el usuario y minería de datos para definir recomendaciones que promuevan el uso del foro en una experiencia virtual de aprendizaje”. *In. RIED*, J.G. Boticario, O.C. Santos, A. Rodríguez-Ascaso, Eds, Vol. 13, pp. 237-264.

Authorization and Disclaimer