Learning Objects and Ontologies to Perform Educational Data Mining

F. Castro, M.A. Alonso
Research Center on Systems and Information Technologies, Autonomous University of Hidalgo's State, Pachuca, Hidalgo, Mexico
fcastro@lsi.upc.edu, marial@uaeh.edu.mx

Abstract - *E-learning systems have established as a strong alternative to traditional distance education. One of the most valuable, but unfortunately, less used in online educational courses is the learning objects (LO) technology. Internet allows the gathering of plenty of information on students' online behavior; however, this information is in raw format, making difficult the knowledge extraction. Moreover, information about the usability of components in the online course, LO, is rarely obtained, and therefore, hardly used in the data mining (DM) process. The knowledge extracted from this information can be used to define personalization strategies tailored to the students' needs and requirements. In this brief study we introduce a platform to perform educational DM process based on gathered information from both, the students' navigational activities in the e-learning system and the information collected from the LO usability. Moreover, the gathered data are structured, organized and formalized by means of an educational ontology.*

Keywords: E-learning, Data Mining, Learning Objects, Ontology.

1 Introduction

Within a decade, the Internet has become a pervasive medium that has changed completely the way information and knowledge are transmitted and shared all over the world. The education community has not limited itself to the role of passive actor in this unfolding story, but it has been at the forefront of most of the changes. Indeed, the Internet and the advance of telecommunication technologies allow us to share and manipulate information in nearly real time. This reality is determining the next generation of distance education tools. Distance education arose from traditional education in order to cover the necessities of remote students and/or help the teaching-learning process, reinforcing or replacing traditional education.

This is e-learning, a new context for education where large amounts of information describing the continuum of the teaching-learning interactions are continuously generated and ubiquitously available. This could be seen as an approval: a lot of information readily available just a click away. But it could equally be seen as an exponentially growing nightmare, in which unstructured information strangles the educational system without providing any articulate knowledge to its actors.

Data Mining was born to tackle problems like this. As a field of research, it is almost contemporary to e-learning. Not just of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. At its most detailed, it can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, pre-processing and modelling to process evaluation and implementation [1]. It is nevertheless usual to pay attention to the Data Mining methods themselves.

Therefore, Data Mining can be used to extract knowledge from e-learning systems through the analysis of the information available in the form of data generated by their users. In this case, the main objective becomes finding the patterns of system usage by teachers and students and, perhaps most importantly, discovering the students' learning behavior patterns. This process is known as educational data mining.

Several research projects have dealt with the integration of data mining methods focusing on e-learning systems improvement. For a deeper inside into these projects the authors recommend [2, 3, 4], where a widespread and deep analysis on different learning platforms is performed, including LON-CAPA [5], AHA! [6], ALFANET [7], etc. Commonly, the existing platforms perform students' classification (using supervised neural networks, decision trees, fuzzy methods, association rules, etc.), and/or students' clustering (using Kohonen's self-organizing maps, EM, etc.).

Most of this research uses data mining techniques to analyze the available data. In turn, we can distinguish diverse Soft Computing-based approaches to e-learning process analysis, i.e. methods to classify students' based on their usage patterns on a web-based course [8, 9, 10], methods oriented towards system personalization [11]. For instance, a neural network model is proposed in [12] to recommend an adequate navigation strategy for the user. A methodology to improve the performance of developed courses through adaptation, using Evolutionary algorithms, is presented in [13]. And,

finally, methods that allow automatic detection of atypical students' behavior such as the Bayesian predictive distribution model to detect irregular learning proposed in [14], and the Generative Topographic Mapping model to detect atypical behavior on the grouping structure of the users of a real virtual campus, presented in [15].

However, just a short number of papers have devoted to include ontologies as a support to perform educational data mining processes, for instance, in [16], a framework for personalized e-learning based on aggregate usage profiles and a domain ontology were presented and a combination of Semantic Web and Web mining methods was used. The Apriori algorithm for Association Rules was applied to capture relationships among URL references based on the navigational patterns of students.

An ontology-based tool, within a Web Semantics framework, was implemented in [17] with the goal of helping e-learning users to find and organize distributed courseware resources. An element of this tool was the implementation of the Bisection K-Means algorithm, used for the grouping of similar learning materials. An approach to automate the classification process of Web learning resources was developed in [18]. The model organizes and labels learning resources according to a concept hierarchy extracted from the extended ontology of the ACM Computing Curricula 2001 for Computer Science.

Despite the existence of some research concerned with the mining of data generated by the use of e-learning systems, there is still a lack of standard methods and techniques to address some open problems in distance education. This is the case, for instance, of methods that could detect similarities in learning behaviors and group (cluster) students according to them. This strategy could be used, for instance, to provide teachers with an adaptive guidance tool to prevent student failure.

It is very difficult and time consuming for teachers to thoroughly track and assess all the activities performed by all students. Moreover, it is hard to evaluate the structure of the course content and its effectiveness on the learning process as it develops. Therefore, it would be helpful to obtain objective feedback from learners in order to track the learning process and assess the online course structure effectiveness [19].

The possibility of tracking user behavior in such environments creates new possibilities for both web-based system architects and designers, but also for the pedagogical and instructional designers who create and organize the learning contents. One of the most interesting possibilities is the personalization of the e-learning process. Personalization arises from the knowledge extracted from the navigational behavior of the e-learning virtual environment users, mostly students in this particular scenario.

To deal with the majority of the problems addressed before, in this brief study we propose the integration of learning objects (LO), including an educational ontology, as a support, in the educational data mining process. The LO can, both, generate information about their usability using the metadata and, by means an ontology, provide a formalization of the entities from the e-learning domains. Learning objects are any digital resource that can be reused for learning or training, and constitute a valuable resource for e-learning course development. Moreover, the motivation for developing an ontology for educational data mining is many-fold. On the one hand, the educational data mining area is devoting many research efforts and rapidly increasing; however, one of the most challenging problems deals with developing a general framework for educational data mining. On the other hand, often, the data extracted from the educational system are raw and unstructured making difficult the knowledge extraction and the decision support stage. Moreover, the students could increase their course performance if the teachers properly provide them with a suitable and actionable feedback; this activity can be achieved by means an appropriate ontology. An educational ontology can be used for formalizing and describing educational scenarios.

The remaining of the paper is organized as follows: section 2 presents a description of our educational data mining proposal. Finally, section 3 wraps up the paper with some preliminary conclusions.

2 Proposed Educational Data Mining Approach

As has stated before, one of the first and more difficult steps in the data mining process is the pre-processing, cleaning and integration data, due to the fact that often the gathered data are in a raw and unstructured format, even, including mixing data.

In the proposed approach, the educational data mining process would benefit from the formalization of data and educational entities provided by an educational ontology. The proposed ontology can be used to discover relationships between the learning resources metadata and should follows the best practices in ontology engineering, for instance, it must be a deep/heavy-weight ontology avoiding multiple inheritance of classes, using a predefined set of relationships and using a top level ontology. Furthermore, the integration of learning objects (LO), and an educational ontology can, both, generate information about the usability of the LO and provide a formalization for the entities from the e-learning activities. Learning objects provide valuable support to deal with the standardization, interoperability, cooperation, transportation and reusability of educational content. The proposed educational ontology would helps and standardizes the data extraction and preparation in the educational data mining process and makes easier the knowledge extraction. Furthermore, the integration of LO and educational ontology will provide valuable support in the students' feedback activity, to help them to cover their academicals needs and requirements.

For the sake of space availability in Figure 1 we summarize the interaction between the actors and the virtual campus as well as the functionalities offered by the proposed platform. The main goal of the platform is to alleviate the virtual tutors' workload and to provide an effective and valuable feedback to students. To deal with these objectives the platform offers tools to discover relevant learning behavior patterns from students' interaction with the educational materials. The knowledge obtained can be used by teachers to design courses in a more effective way and to timely detect students with learning difficulties. The extracted knowledge can also be helpful for the students to know their own learning performance and therefore use more efficiently the educational resources.

The platform presented in this paper includes as part of its skills the identification of students' learning behavior models that allow both, students and teachers, to know the future performance of the student based on their current learning behavior and the course information available at that moment. This functionality allows teachers to give actionable feedback to those students that need it, such as the failing students. The platform also provides functionalities to forecast students' learning performance and to determine the most relevant features involved in the evaluation process, and their relative weight. Finally, it allows extracting logical rules that describe the students' learning behavior patterns, which are easily understandable by experts in educational domains. All the platform functionalities could be performed when the teacher or student considers relevant doing with the main characteristic that in each execution will be used the information available at the moment, i.e. using temporal models that, of course, give better performance prediction results when it has wealthy information that often happens at the last phases of the course. Using the information obtained by the toolbox, the teachers could prevent possible course failing to the students susceptible to obtain worse performance results in order to accomplish their learning requirements.

3 Preliminary Conclusions

The tracking of user navigational behavior in virtual campus e-learning environments makes the web mining of the resulting data bases possible. This opens new possibilities for the pedagogical and instructional designers who create and organize the learning contents; amongst the most interesting ones, the personalization of the e-learning process.

One of the most difficult and time-consuming activities for teachers in distance education courses is the evaluation process, due to the fact that, in this type of courses, the review process is better accomplished through collaborative resources such as e-mail, discussion forums, chats, etc. As a result, this evaluation usually has to be carried out according to a large number of parameters, whose influence in the final mark is not always well defined and/or understood. Therefore,

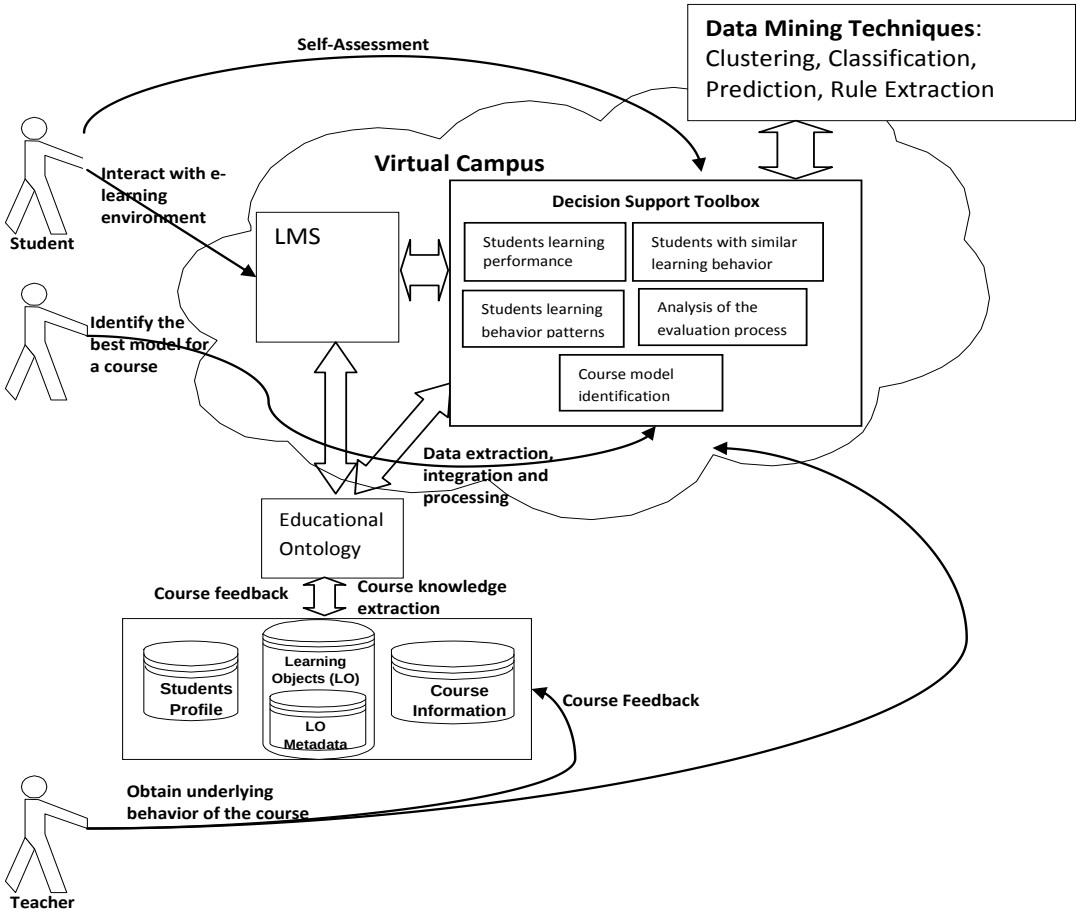

Figure 1. Proposed platform functionality.

it would be helpful to discover features that are highly relevant for the evaluation of groups (clusters) of students. In this way, it would be possible for teachers to provide feedback to students, on their learning activity, online and in real time in a more effective way than if we tried to do it individually. This option, using clustering methods, could reduce teachers' workload. In this sense the grouping of students with similar learning behavior could significantly help in the e-learning personalization, by the fact that the teacher can send grouped students' feedback instead to perform that in an individualized way, alleviating their workload, and even, the learning groups' behavior can be used to create greatly cohesive, organize and interactive students group based on their performance and/or learning behavior.

In order to obtain concluding remarks, nowadays, we are performing a set of deeply and detailed experiments with the aim to identify qualitatively and quantitatively the course improvements, accomplished using the resulting knowledge provided for the proposed platform. Partial results, obtained by the authors, have indicated that the teachers workload have been reduced considerably, optimizing their course activities, and providing a better quality of the students' feedback; accordingly the students learning and course performance have significantly increased. As a future trend of the presented platform, we would include a Computer Adaptive Tests (CAT) methodology in order to improve the e-learning evaluation process. We think that this could significantly help teachers in the evaluation procedure due to the fact that this methodology could provide valuable teacher support in reducing the e-learning teacher workload. Preliminary results of the authors are very encouraging in this educational topic.

Acknowledgements

The authors gratefully acknowledge funding from Mexican Secretary of Public Education through the project FOMIX-HGO-2009-C01-131024.

4 References

[1] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRIPS-DM 1.0 Step by Step Data Mining Guide. CRISP-DM Consortium (2000).

[2] Castro, F., Vellido, A., Nebot, A. & Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems. In L.C. Jain, R.A. Tedman & D.K. Tedman, Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62 (pp. 183-221). Germany: Springer-Verlag.

[3] Croock, M., Mofers, F., Van Veen, M., Van Rosmalen, P., Brouns, F., Boticario, J., Barrera, C., Santos, O., Ayala, A., Gaudioso, E., Hernández, F., Arana, C., Trueba, I. (2002). State-of-the-Art. ALFanet/IST-2001-33288 Deliverable D12. Open Universiteit Nederland. URL: http://learningnetworks.org/

[4] Van Rosmalen, P., Brouns, F., Tattersall, C., Vogten, H., van Bruggen, J., Sloep, P., Koper, R. (2005). Towards an Open Framework for Adaptive Agent-Supported e-Learning. International Journal Continuing Engineering Education and Lifelong Learning 15(3–6) 261-275.

[5] Minaei-Bidgoli, B., Tan, P.N., Punch, W.F. (2004). Mining Interesting Contrast Rules for a Web-based Educational System. In Proceedings of The 2004 International Conference on Machine Learning and Applications, ICMLA'04. Louisville, KY.

[6] Romero, C., Ventura, S., De Bra, P., De Castro, C.(2003). Discovering Prediction Rules in AHA! Courses. In Proceedings of the User Modelling Conference. Johnstown, Pennsylvania (pp. 35-44).

[7] Van der Klink, M., Boon, J., Rusman, E., Rodrigo, M., Fuentes, C., Arana, C., Barrera, C., Hoke, I., Franco, M. (2002). Initial Market Study, ALFanet/IST-2001-33288 Deliverable D72. Open Universiteit Nederland. URL: http://learningnetworks.org/downloads/alfanet-d72-initialmarket -studies.pdf.

[8] Minaei-Bidgoli, B. and Punch, W.F. (2003). Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System: Genetic and Evolutionary Computation Conference.

[9] Mullier, D. (2003). A Tutorial Supervisor for Automatic Assessment in Educational Systems: International Journal on E-Learning.

[10] Stathacopoulou, G.D. and Grigoriadou, M. (1999). Neural Network-Based Fuzzy Modeling of the Student in Intelligent Tutoring Systems: International Joint Conference on Neural Networks, Washington.

[11] Brusilovsky, P. (2001). Adaptive hypermedia: User Modeling and User Adapted Interaction, Ten Year Anniversary Issue, 11 (1/2), 87-110.

[12] Mizue, K. and Toshio, O. (2001). N3: Neural Network Navigation Support–Knowledge-Navigation in Hyperspace: The Sub-symbolic Approach: Journal of Educational Multimedia and Hypermedia 10(1), 85-103.

[13] Romero, C., Ventura, S. and De Bra, P. (2004). Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware: User Modeling and User-Adapted Interaction, 14(5), 425-464.

[14] Ueno, M. (2003). LMS with irregular learning processes detection system: World Conference on E-Learning in Corp., Govt., Health, & Higher Ed., 2003(1), 2486-2493.

[15] Castro, F., Vellido, A., Nebot, A. and Minguillón, J. (2005). Detecting Atypical Student Behaviour on an e-learning System: VI Congreso Nacional de Informática Educativa, Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación (SINTICE'2005). Granada, España, September 14-16.

[16] Markellou, P., Mousourouli, I., Spiros, S., Tsakalidis, A.: Using Semantic Web Mining Technologies for Personalized e-Learning Experiences. In: Uskov, V. (ed.): The Fourth IASTED Conference on Web-based Education, WBE-2005. Grindelwald, Switzerland (2005)

[17] Tane, J., Schmitz, C., Stumme, G.: Semantic Resource Management for the Web: An e- learning Application. In: Fieldman, S., Uretsky, M. (eds.): The 13th World Wide Web Conference 2004, WWW2004. ACM, New York (2004) 1–10

[18] Singh, S.P.: Hierarchical Classification of Learning Resources through Supervised Learning. In: World Conference on e-Learning in Corp., Govt., Health., & Higher Education (2004) 178–183

[19] Zaïane, O.R. and Luo, J. (2001). "Towards evaluating learners' behavior in a Web-based distance learning environment", in Proc. of IEEE International Conference on Advanced Learning Technologies (ICALT'01), Madison, WI, August 6-8, pp 357-360.