

E-learning Web Miner: A data mining application to help instructors involved in virtual courses

D. GARCÍA-SAIZ

University of Cantabria, Spain

AND

M.E. ZORRILLA PANTALEÓN

University of Cantabria, Spain

In this demo we present a data mining application, called E-learning Web Miner (EIWM), which aims to help instructors involved in distance education to discover their students' behavior profiles and models about how they navigate and work in their virtual courses which are offered in Learning Content Management Systems. The main characteristic is that the users do not require data mining knowledge to use EIWM, they only have to send a data file according to one of the templates provided by the application and request the results. The application carries out the Knowledge Discovery in Database (KDD) process itself. Furthermore, the application provides an interface based on web services which allow its integration and use by any external software.

Key Words and Phrases: Educational Data Mining, Learning Content Management Systems, Towards Parameter-free Data Mining, Data Mining Web Services

1. INTRODUCTION

Our goal is to describe a data mining application, called E-learning Web Miner (EIWM), implemented in the University of Cantabria which assists the instructors involved in virtual education in their teaching activity in the sense that this application helps instructors to discover on the one hand, the distance students' behavior based on their navigation and demographic data and on the other hand, how they surf and work in a distance course offered in an e-Learning platform such as Moodle or Blackboard. The patterns and models which it generates help instructors to better understand the learning process and to analyze the course organization effectiveness (design, tasks, resources used, and so on).

EIWM has been developed following a Service Oriented Architecture (SOA) and implemented by means of web services (a more extended description is found in [Zorrilla and García-Saiz 2011]) with the aim of it being easily extended and choreographed with other web services offered by third parties. The two main characteristics of this Data Mining Application are that: it offers a set of templates which resolve some of the common questions of instructors involved in virtual courses and it is configured to be used by non-data mining experts. As far as we know there is not a similar tool available and a clear necessity for a tool which addresses this issue exists according to the extensive research activity which is being carried out in this field [Romero and Ventura 2010].

2. WORKING WITH E-LEARNING WEB MINER

EIWM currently offers instructors, through an easy web-based interface, the possibility of

answering three different questions: What kinds of resources are frequently used together (forum, mail, etc.) in each learning session; What is the student session profile; and What is the profile of the students who enroll in the course.

The end-user chooses the question he wants the application to answer and indicates where the data file to process is stored or selects the course if this is hosted in Blackboard, the official Learning Content Management System of our University. Next, EIWM carries out the KDD process and returns the result to the instructor in textual and graphic format.

The KDD process to answer each question is defined by means of templates. The templates contain the input attributes which the data file must have as well as the pre-processing tasks, the mining algorithms and the parameter-setting which is adequate for obtaining the patterns. These templates have been defined and validated by means of a previous experimentation carried out in several virtual courses taught at the University of Cantabria.

As an example, we describe briefly the student profile template. The objective of this template is to group students according to their activity in the e-learning platform and their demographic data. The template establishes as input parameters: gender, age, number of sessions in the course, time spent in the course, average sessions per week, average time spent per week; and as data mining algorithms for obtaining the patterns: EM and SimpleKMeans from Weka. EM algorithm is used to determine the number of clusters with which the SimpleKMeans algorithm will be executed. We generate the patterns with SimpleKMeans because it is one of the most used in practical problems, its execution is quick and furthermore, the results which it offers are easy to understand statistically and graphically. Before the mining process starts, a pre-processing task is carried out in order to evaluate the quality of the data for the process, for example, to eliminate correlated or highly unbalanced data or outliers.

3. CURRENT AND FUTURE WORK

Currently, we are extending EIWM with a new template. The goal is to predict the final mark which the student will obtain in the course according to the global activity the student has carried out during the course. Likewise, our group is improving *yacaree*, a parameter-free rule miner [Balcázar 2011], to respond adequately to educational data needs [Zorrilla et al. 2011]. Once this task is completed, the algorithm will be used by the application. Next, we will continue adding new templates and, in the near future, our objective is to deploy our application as a service in Cloud Computing, offering this as a Software-as-a-Service (SaaS). According to [Software & Information Industry Association 2006], SaaS is a model for software delivery that allows lower total cost of ownership and better return on investment for subscribers.

4. REFERENCES

- BALCÁZAR, J. L. 2011. Parameter-free association rule mining with *yacaree*. In EGC, A. Khenchaf and P. Poncelet (Eds) *Revue des Nouvelles Technologies de l'Information Series*, vol. RNTI-E-20. Hermann-Editions, 251–254.
- ROMERO, C. AND VENTURA, S. 2010. Educational Data Mining: A Review of the State-of-the-Art. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. Vol 40. Issue 6. Pages 601 – 618.
- SOFTWARE & INFORMATION INDUSTRY ASSOCIATION. 2006. *Software-as-a-Service: A Comprehensive Look at the Total Cost of Ownership of Software Applications*. Retrieved May, 2011 from <http://www.winnow.com/saas.pdf>
- ZORRILLA, M.E., AND GARCÍA-SAIZ, D. 2011. A Service Oriented Architecture to provide data mining services for non-expert data miners. Accepted. To appear in *Decision Support Systems*.
- ZORRILLA, M.E., GARCÍA-SAIZ, D., AND BALCÁZAR, J.L. 2011 *Towards Parameter-Free Data Mining: Mining Educational Data with *yacaree**. Submitted to *EDM 2011*