1st International Conference on Higher Education Advances, HEAd'15
Universitat Politècnica de València, València, 2015
DOI: http://dx.doi.org/10.4995/HEAd15.2015.448

Learning Analytics for E-Learning Content Recommendations

A. S. Perera*, S. Tharsan*

*Dept. of Computer Science and Engineering,
University of Moratuwa, Moratuwa 10400, Sri Lanka.

Abstract: E-Learning systems have caused a rapid increase to the amount of learning content available on the web. It has become a time consuming and a daunting task for e-learners to find the relevant content that they should study. Existing e-learning technology lacks the automated capability to provide guidance for students to prioritize and engage in the most vital course content. The students who are unable to find out the most suitable resources, for their studies and the assignments, may waste most of their time on browsing and searching. Some of the "good-students" can indirectly act as good guides to other students. Average learners could follow the content adopted by good students in the process of learning. It is possible to capture the behaviour of "good-students" and expose it as a form of automated guiding. For this to work it is important to be able to predict students who are going to be successful at the end of the course based on their performance during the early part of the course. This work demonstrates the use of data mining techniques on e-Learning data to enable "Good-students" to indirectly guide "Average-Students" to find the most relevant content on an e-Learning environment.

Keywords: e-Learning; Learning Analytics; recommender systems; data mining;

Introduction

The popularity of e-learning has created a huge amount of educational resources and as a result locating suitable learning references is a formidable challenge (Lauria & Baron, 2011). In a typical University e-learning environment, students are expected to follow all the course material that is available in the e-learning systems like Moodle and Sakai (Lauria & Baron, 2011). But, when the time is limited for each course and with the availability of a large number of learning material they need to prioritize the items to maximize on the learning process. Excess amount of material lead the students to spend more time on browsing and filtering to identify information that suits their needs better, rather than on actually spending time on the content. If they are guided to select the most suitable learning material among all of the existing ones, they could spend more time in learning, than in browsing. Ability to identify "good learning material" is not easy for all the students. But it would not be a hard task for "good learners/good students" to choose "good learning material". Hence, suggesting the "highly accessed materials" by the "good students" to the average students would help average students to get involved in studies rather than wasting time on searching. This work therefore aims to propose an intelligent recommender system that is based on the artificial intelligence gained from the access patterns of "good learners" to the students who are not capable of identifying relevant learning material that are suitable for their studies. The idea of learning from best students or good learners is also strongly supported by the Social Learning Theory, that states that people can learn by observing the behaviour of others and the outcome of their behaviour (Ghauth & Abdullah, 2011).

This work used the weblog data generated by the Moodle learning management system which is installed at the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka, to analyse and build the prediction and recommender models. The Moodle log contains a rich set of data that can help to design and construct models for the recommendation system.

Identifying the good students and recommending learning resources are two key problems dealt within the methodological framework. The quality of the learning material recommendations has an important effect on a student's future learning behaviour (Lu, 2004).

Learning Analytics and Recommender Systems

E-learning has been used for more than a decade. But recently learning analytics has received considerable attention in higher education (Lauria & Baron, 2011). The continuous stream of weblog data being collected and stored in course management systems can be used as input to build the predictive models that can be used to implement data driven decision making systems. Despite the identification of the vast potential, learning analytics remains as an immature field that is yet to be explored broadly across a range of higher educational institutions, student populations and learning technologies (Lauria & Baron, 2011 ; Monk, 2005; Castro, 2007). All access to a learning management system can be tracked. Course access logs contain all the student interactions in a chronological order, indicating when a given URL was requested by a particular client (Zaïane, 2001). Log files generally has the login name of the user who generated the request, the date and time of the request, the method of the request, the name of the requested file, the result of the request (success, failure, error, etc.), the size of the data sent, the URL of the page, data about the application and some information about the data exchanged between the client and the server. These log entries are not in a format that is usable for data mining. Data mining applications require reformatting and clean-up of data in order to identify the necessary information to build up the predictive models (Spiliopoulou, Faulstich, & Winkler, 1999).

The Signals project at Purdue University, which applied the principles of business intelligence analytics to academia, facilitates to improve student success, retention, and graduation rates (Mazza & Milani, 2005). The project mines a large dataset continuously and applies statistical techniques to predict which students might be falling behind. The warnings provided by the Purdue Early Warning System (PAWS) tend to be general in nature and does not include the resources available for a specific course. These warnings alert the students, but do not provide any learning material recommendations. Different technologies can be used to build recommender systems. Technologies in the area of data mining such as collaborative filtering, content based filtering, rule based expert system, artificial network and fuzzy logic can be used for this purpose (Ghauth & Abdullah, 2011; Govaerts, et al., 2012; Lu, 2004). The recommender systems in (Ghauth & Abdullah, 2011) suggests learning materials based on the average ratings given by good learners. Though it outperforms with respect to the accuracy of recommendations when compared with other frameworks, the drawback of this framework is, it requires explicit rating of learning material. It would be very hard to get the majority of the students to rate the learning material when they are hard pressed to complete the assigned work. In order to address the issue of explicitly collecting rating information automatic recommendation could be more atractive based on browsing patterns of other successful learners (Zaiane, 2002). The successful learners' have the capability to discriminate the relevant materials from a large collection of resources. Hence, if the access pattern of the good students is identified, based on that, the resources that are used by the good students can be recommended to the average students.

The personalized recommendation approaches were first proposed and applied to e-commerce applications like many big web retailers, including Amazon.com and CDNow.com when the customers buy products. Providing personalized product recommendations would help the customers, to find products that they would like to purchase, by producing a list of recommended products for each given customer. Such recommendations are generated by the recommender systems, based on the analysis done, on the past transactions made by each customer (Lu, 2004). The content-based recommender systems has a common problem which requires a large set of key attributes. If the dataset is too small, obviously there is insufficient information to learn about the customer profile. Hence, if a customer visits the site, but has not made any purchase and the customer wants to buy a product which is not frequently purchased. In this case, what products need to be suggested, This is called a cold start problem(Ghauth & Abdullah, 2011). (Lu, 2004) proposes a framework called PLRS for recommending learning materials to students who may have different backgrounds, learning styles and learning needs. The PLRS tries to identify a student's need, and how to accurately find the learning materials which match the student's need. The framework comprises four components, each with different purposes. 'Getting student information', 'Identifying student requirements', 'Learning material matching analysis and 'Generating recommendations' are the main four components of the PLRS.

It has also been suggested that e-learning environments have a psychological and physical cost, as it "may affect student performance due to cognitive overload caused by too much information to process multiple resources in a very short time and because of vision problems caused by spending long time to read the materials off computer screens (Monk, 2005).

Based on the literature review, this work attempts to build a resource recommendation framework based on access behaviour of good learners. Though recommendation based on good learners' rating is the ideal solution, it cannot be implemented in most of the current LMSs at the moment as this requires explicitly getting the rating feedback from the good learners.

Methodology

The objective of this research is to build e-Learning resource recommender system based on the behaviour of the "good-students" in a particular course. This requires two prediction systems. One system to predict the "good-students" based on the initial activities carried out by the students during the first few weeks of the course. And the other system is to be able to recommend suitable e-Learning content to be able to succeed in the course. Both systems are modelled using available data mining techniques by extracting patterns from completed e-Learning courses. Predictive analytics model development from data mining requires the data extraction, transformation, loading, training, and validation. In addition to the above data privacy also needs to be assured when building the model using actual data from completed courses.

Dataset

The dataset was extracted from a Moodle course management system used at the University of Moratuwa, Sri Lanka. Ten courses were selected in the area of computer

science for the study and some characterisation information on the data is given in Table 1. Only a selected set of attributes from a total of 84 attributes are indicated in the table. In order to address the issue of privacy preservation the data was anonymized to avoid disclosure of sensitive data without removing the important patterns in the data based on (Zhang, et al, 2006). The input data was normalized using standard normalization techniques. And also the data was cleaned and integrated into a database based on a standard data warehousing schema that contains all the extracted data for the purpose of building the predictive models.

Standard data mining pre-processing steps were carried out to identify the impact of each of the attributes towards building the prediction model and a reduced set of attributes were selected out of the original 84 attributes for the respective prediction models. These attributes contained direct attributes and also derived attributes. For the student performance prediction model the selected most predictively influential attributes included the grade obtained by students for assignments, number of forum discussions, number of wiki entries made by the student, number of completed assignments, number of visits to the course page, number of resources accessed by the student, and the number of resources accessed by the students close to deadlines. For the resource recommendation model attributes such as the file size of the resource, location of the resource with respect to the other resources within the week, week on which the resources was made available for students, and resources accessed by the good students were some of the attributes that were predictively influential. The input dataset was also appropriately split into a training and validation set.

Table 1 Descriptive Characterization of Input data

Course Code and Year	No of Students	Average Final Grade	No of Open Discussions	No of Discussion Replies	No of Wiki Entries	No of Assignments
CS5105ISec-2008	42	78.32	99	56	164	5
CS5105ISec-2009	40	79.14	7	189	214	4
CS5404CNS-2010	32	75.07	110	71	252	5
CS5105ISec-2010	31	76.41	19	97	127	4
CS5401SND-2011	29	46.95	44	34	81	2
CS5404CNS-2011	29	71.85	37	45	196	6
CS5105ISec-2011	23	76.19	12	56	98	3
CS5404CNS-2012	31	72.65	73	80	204	6
CS5105ISec-2012	24	78.59	22	72	90	5
CS5404CNS-2009	33	76.90	133	47	451	4

Data Mining for Predictive Models

In the process of data mining for predictive models it is important to evaluate multiple techniques to identify the most appropriate technique (Han, et al, 2012). This work evaluated Decision Tree, Neural Networks, SVM, k-NN, Naive Basis, Rule Induction, Perceptron, Linear Regression, Polynomial Regression, Vector Linear Regression, and Gaussian Process in standard forms as available in literature (Han, et al, 2012).

The respective accuracy values for the predictive models are shown in Figure 1. As indicated by the figure most of the classification techniques are providing a significantly high accuracy. All the classification techniques were executed multiple times with different settings to obtain the maximum possible accuracy. Also this shows that the data pre-processing steps carried out on the input data have been successful in extracting the most relevant data for the purpose of building the classification model.

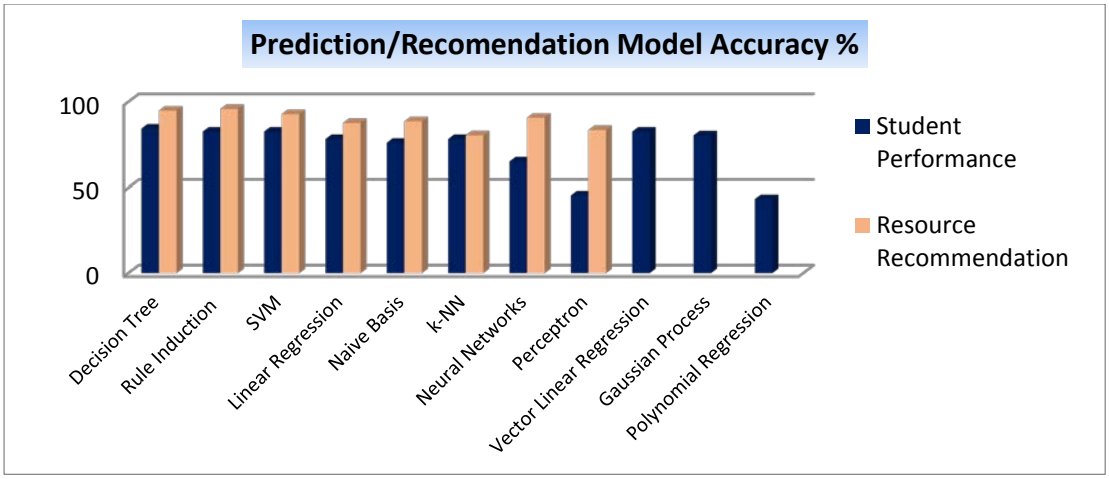

Figure 1 Prediction Model Accuracy

Results and Discussion

High accuracy levels indicated on figure 1 clearly indicates the ability to use data mining techniques for the purpose of building a recommender system for students engaged in eLearning courses. Out of the Data Mining techniques carried out Decision Trees shows the highest overall accuracy for both predictive models and can be suggested as the technique to be used for a future recommendation application to be built based on the work presented on this paper. Decision Tree algorithm extracts rules based on the patterns in the data in the process of building the predictive model. One major advantage of Decision Tree approach is that human domain experts can look at the rules and formulate an interpretation of the rules to further validate the predictive models. Also these rules can be very easily implemented on the e-Learning platform to execute on the existing database to perform the recommendation tasks. It is very important to have light weight predicative technique to reduce the workload on the e-Learning computing environment.

The predictive models described in this work could be implemented as a plugin for the Moodle that would recommend the resources to the students who require assistance to identify the most relevant materials. It is also possible to use the student performance classification model as a standalone application with some slight modifications as an early warning system to identify the students who are at a risk of failing, so that they can be monitored with close surveillance. Same ideas can also be extended to Massive Online Open Courses (MOOCS) as an additional feature to recommend resources based on other users. One minor drawback of the application of the system would be the loss of opportunities for students to gain experience in searching for content. Applicability of the proposed learning analytics to other subject areas other than computer science needs to be also evaluated as future work.

Conclusions

An additional benefit that can be provided for students in an eLearning environment is ability to recommend the most appropriate resource for the learning process. All academics expect all students to go through all the learning material given on course webpages. But in practice most of the time students do not have sufficient time to go through all the resources and would be at an advantage if the student has some from

of guidance to identify the most appropriate resource from the list of resource available on the course page. The proposed solution has the capability to provide individualised automated feedback and guidance for students to prioritize and engage in the required course content. This will allow students who do not have knowledge to find out the most suitable resources, links and references for their studies and the assignments, without wasting time. The proposed solution captures the behaviour of "good-students" to indirectly act as "good" guides to their fellow students. The process and the data mining models described in this work can be used to identify "good-students" in a class before the final exam and use their access pattern as a guide and a recommender with high accuracy. The student performance prediction has an accuracy of 85% and the recommendation system has an accuracy of 95% based on the validation conducted on the input data. It is important to note that the high accuracies are obtained through a process of a rigorous data pre-processing based on the dataset available for this study. The pre-processing process and the predicative models can be implemented on any e-Learning environment that produces similar data.

References

Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems., *Evolution of Teaching and Learning Paradigms in Intelligent Environment*, 183–221.

Ghauth, K. I., & Abdullah, N. A. (2011). The Effect of Incorporating Good Learners' Ratings in e-Learning Content-based Recommender System, *Educational Technology & Society, 14(2)*, 248–257.

Govaerts, Verbert, Duval, & Pardo S., "The student activity meter for awareness and self-reflection," in CHI '12 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, USA, 869–884.

Han J., Kamber. & Pei, j. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.

Lauria, E. J. M., & Baron, J. (2011). Mining Sakai to Measure Student Performance, *Opportunities and Challenges in Academic Analytics*.

Lu, J. (2004). Personalized e-learning material recommender system. In *In: Proc. of the Int. Conf. on Information Technology for Application* 374–379.

Mazza, R., & Milani, C. (2005). Exploring Usage Analysis in Learning Systems: Gaining Insights from Visualisations, *AIED Workshops (AIED'05), 65-72.*

Monk, D. (2005). Using data mining for e-learning decision making. *Electronic Journal of E-Learning*, *3*, 41–54.

Spiliopoulou, M., Faulstich, L. C., & Winkler, K. (1999). A Data Miner analyzing the Navigational Behaviour of Web Users. *Proc. of the Workshop on Machine Learning in User Modelling of the ACAI99.*

Zaiane, O. R. (2002). Building a recommender agent for e-learning systems. *International Conference on Computers in Education, 2002. Proceedings* vol.1, 55-59.