Analyzing Students' Interaction Based on their Responses to Determine Learning Outcomes

Fazel Keshtkar
Southeast Missouri State University
One University Plaza, Cape Girardeau, MO, USA
fkeshtkar@semo.edu

Jordan Cowart
Southeast Missouri State University
One University Plaza
Cape Girardeau, MO, USA
jrcowart1s@semo.edu

Ben Kingen
Southeast Missouri State University
One University Plaza
Cape Girardeau, MO, USA
bwkingen@semo.edu

Andrew Crutcher
Southeast Missouri State University
One University Plaza,Cape Girardeau, MO, USA
alcrutcher1s@semo.edu

ABSTRACT

Online learning platforms such as Moodle and MOOC (Coursera, edX, etc) have become popular in higher education. These platforms provide information that are potentially useful in developing new student learning models. One source of information provided by these platforms is in the form of student interaction with one another, instructors, and the platform itself. These interactions contain various activities such as: participation in forum discussion, how frequently a student is logged into their account, and frequency of reading posted activities, etc. Using Data Mining techniques, namely clustering algorithms to find students with similar behavior patterns, our goal is to develop a student model that can be conducted by learning these interaction patterns. In doing so, we aim to develop a method by which to provide students with different guidelines and instructions that will help to improve their performance. This research is in progress and our data include Moodle online courses in computer science in different semesters.

Keywords

Online Learning, Student Behaviors, Student Outcomes, Moodle, Data Mining, Clustering, Educational Data Mining

1. INTRODUCTION

Detecting students' performance is one of the most crucial tasks in online learning and educational data mining (EDM), a task which falls under the scope of classification/clustering or other algorithms. Various learning methods have been applied to detect course results and academic performance with each learning algorithm performing differently with different datasets [4]. The No Free Lunch Theorem states that it is difficult to choose a specific model or classification algorithm for this difficult task [2]. Therefore, discovering and applying appropriate methods for a specific dataset should yield a significant improvement in the effectiveness of a given learning algorithm. Our approach will apply learning algorithms based on metadata, as they have proven to be sufficient to address this problem [2]. These meta-learning algorithms have been studied by exploring metadata to adopt suitable algorithm based on data mining and machine learning techniques [5]. In this research, we propose to apply various classifications/clustering models, evaluation measurements, and statistical analysis test to predict the performance of students' learning outcomes based on new dataset. This paper focuses on a portion of our statistical analysis, namely the examination of student response times to professor activity.

2. DATASET

Our dataset contains student and professor metadata from eleven courses over two semesters at Southeast Missouri State University. The metadata is in the form of log data from the online learning platform that the school uses, Moodle. In order to determine which of the features the metadata provides, we have performed rudimentary statistical analysis using SPSS. A basic overview of our dataset is provided in Table 1.

Table 1. Course Overview

Course	Number of Students	Number of interactions	Average Interactions
CS1	12	4281	356.75
CS2	**53**	**14006**	**264.26**
CS3	23	3891	169.17
IS1	33	26682	808.55
IS2	31	20049	646.74
IS3	10	7906	790.60
IS4	13	13311	1023.92
IS5	19	10986	578.21
IS6	30	31433	1047.77
IS7	7	13150	1878.57
UI1	27	7127	263.96

2.1 Data Processing

For this portion of analysis, we analyze CS2 (bold in Table 1.) for students' interaction response times; this was due to the large sample size it provided with respect to the other courses in our dataset. There were five students that failed to complete this course, so they were dropped from the dataset for this particular portion of analysis to prevent data skewing in the later weeks of the class.

3. METHODOLOGY

We propose that applying data mining techniques and statistical analysis of metadata from an online learning platform will allow us to derive insights into student interaction patterns. Using these insights, we theorize that a student learning model can be developed by learning these interaction patterns. In doing so, we aim to develop a method by which to provide students with different guidelines and instructions that will help to improve their performance.

3.1 Feature Selection

Our dataset explicitly provides the following features: the course in which an activity occurred, the time of occurrence, the IP address from which an activity originated, the user which performed the action, the action occured (course, user, assignment, and grade view), and information about the action completed.

There are also metadata that are not explicitly provided in the dataset but can be extracted. For example, our dataset does provide with specifics of the activity that the student is performing (e.g. posting to a forum, content of their posts, etc.) . However, we aware that a student is automatically logged out from their Moodle account if they have not performed an activity within 15 minutes. Using this knowledge, we can then determine when a student is logged out, approximately number of times they login, and the time interval between logins. We are aware that that there may be more metadata hidden within our dataset that maybe found upon closer examination that we plan to consider for future research.

Finally, we consider statistical features that have been extracted. For this portion of the analysis we considered how quickly the students responded to activities made by the professor; these activities include: updates to materials, posting of assignments, and updating student grades. We have computed the average student response time per activity, a sample standard deviation for the response time per activity, the total average response time and a sample standard deviation for the course during the first two weeks and the entirety of the course. We have also computed the top ten activities that resulted in the quickest average response times and the top ten activities that resulted in the slowest average response times.

4. RESULTS AND DISCUSSION

One of our goals was to explore trends in how students interact with their course over the duration of a semester and, more specifically, how quickly they react to activities performed by their professor. We noticed that when a professor interacts with Moodle, they typically perform a lot more than one action. For our statistics, we counted the time it took for each student in the course to respond to the last update to the course page that a professor made in a continuous block of interactions. For each of these interactions, we then calculated the average response time per student and the overall average response time for that particular activity. The average response times per activity are shown in Figure 1. We can see that student activity has fluctuates throughout the semester, but further analysis is needed to determine possible causes for these fluctuations. The only readily explainable peak is activity thirty, which occurred during a five day break.

5. RELATED WORKS

Wang [6] has indicated a need for the examination of log analysis within online learning platforms, namely the examination of

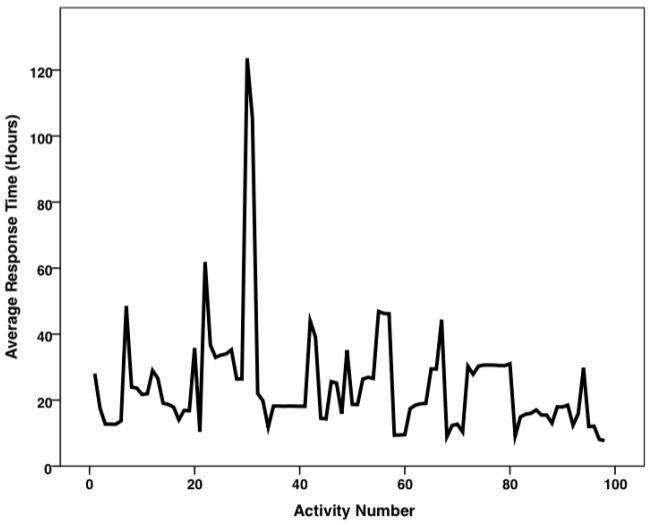

Figure 1. Average response times per activity.

indicators of participation such as use of discussion forums, quiz completion rate, and video usage. The research of Yudelson et al [7]. indicates that finding and analyzing certain sub-populations within a student body can produce a better predictive model than that of examining the entire population; importantly, these sub-populations tend produce a more substantial data footprint [7]. The research of Coffrin et al. indicates that student interactivity and success during the first two weeks of a course strongly related to their outcomes at the end of the course. They also suggested that identifying students based on their patterns of engagement presents the opportunity of tailored feedback to these sub-populations [1].

6. ACKNOWLEDGMENTS

This research is funded by GRFC grant, Southeast Missouri State University.

7. REFERENCES

1. Coffrin, C., Corrin, L., Barba, P., Kennedy, G. Visualizing patterns of student engagement and performance in MOOCs. Proceedings of the Fourth International Conference on Learning Analytics and Knowledge. 2014.
2. Hamälainen, W., Vinni M. Classifiers for educational data mining; Handbook of Educational Data Mining. Chapman & Hall/CRC. 2011.
3. Ho T.K., Basu M. Complexity measures of supervised classification problems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3):289-300, 2002.
4. Romero, C. and Ventura, S. Data Mining in Education. Wire Interdisciplinary Reviews: Da-ta Mining and Knowledge Discovery, 3:12-27. 2013.
5. Song, Q, Wang, G, Wang, C. Automatic recommendation of classification algorithms based on dataset characteristics. Pattern recognition. 45, 2672–2689, 2012.
6. Wang, Y. MOOC Learner Motivation and Learning Pattern Discovery - A Research Prospectus Paper. Proceedings of the 7th International Conference on Educational Data Mining. 2014.
7. Yudelson, M., Fancsali, S., Ritter, S., Berman, S., Nixon, T., Joshi, A. Better Data Beat Big Data. Proceedings of the 7th International Conference on Educational Data Mining. 2014.