ARE YOU PAYING ATTENTION? ASSESSING STUDENTS ATTENTION AND PARTICIPATION ON LEARNING MANAGEMENT SYSTEM

Nuket Savaskan Nowlan, M. Omair Shafiq, Ali Arya
School of Information Technology, Carleton University (CANADA)

Abstract

In this study, we investigate students' online activity patterns as recorded by a Learning Management System (LMS), along Attention and Participation dimensions, to see if there is any correlation between students Attention activity pattern, Participation activity pattern, and grade-based performance. We hypothesize that LMS-based tracking of attention and participation activities can be used as an early measure of performance and the required support. Our findings show potential for such an approach.

1 INTRODUCTION

Educational technologies are more effective when they can be adapted to the needs of individual learners [3]. Adapting students' needs can only be done through on-going observation and assessment which with a proper design and implemented method might be possible. Further, support and customization should be offered accordingly. Traditional assessment methods of learner's evaluation are based on assignments. Assignments submitted by students, quizzes and exams taken, discussions, and forum contributions are evaluated by instructors in assessing the learner's course performance. However, this traditional approach often results in missing valuable information such as learner's early engagement with the course that may affect future success.

Ability to record students every interaction in CBLEs is providing a very valuable opportunity to educational organizations to anticipate students need and provide timely support; ideally, before student fails the course, the exam. The challenge, of course, is to interpret user data collected by the system in a way to provide suggestions of actions to improve student's performance.

Data Mining (DM) is a discipline that utilizes machine learning and databases. Educational Data Mining (EDM) uses methods defined by DM, improves and adopts them to educational setting to gain insight into students learning and setting inside learning occurs [19]. EDM is a relatively new discipline. The International Conference on Educational Data Mining is the primary EDM conference, which debuted in 2008. The overall objective of this study is to provide a contribution in the field of EDM.

In this study, we will be investigating the role of attention and participation in student’s performance (grades). In the early stage of a course, viewing course components, learning about the administrative information of the course, deadline and groups reading the discussion, overall paying attention are the activity students are mostly involved. Therefore, this explorative study asks:

RQ: Can we identify emerging patterns for student’s early activities in a course indicating their future performance?

We used the Learning Management System (LMS) data captured from a fully online course on Moodle (Modular Object-Oriented Development Learning Platform, http://www.moodle.org). Data was anonymized and offered with creative common license by Moodle Pty Ltd [11][12]. To the best of our knowledge, this data has never been investigated through indexing with time to gain better insight into students learning process. Hussain et al. [13]used the same publicly available data to identify students at risk of failing and features providing a strong indication. Hussain at all preferred using the data as static data ad used K-mean clustering method. Greene et al. [14]however, found evidence suggesting the timing of these activities and treating Moodle data as timing series provides additional information that provides a better model with higher prediction rate. The course used in this study was not a fully online course.

In this study, after splitting data and creating time series on attention and participation features, we apply time series-based clustering analysis focusing early period of the course. To the best of our knowledge, this specific investigation has never been done on Moodle data before, on a fully online course.
2 RELATED WORK

One of the advantages of EDM is enabling researchers to investigate attributes contributing learning and identify relationship in between and to performance. Relationship between students’ online discussion forum activity and students pass/fail performance of the course is another area researcher studied through EDM [17]. Romero et al. used supervised classification algorithms and unsupervised clustering algorithms as EDM method to identify the best algorithm to predict performance[17]. Romero et al. noted that ability to display the groups of students by clustering algorithms provides advantages, and it would be nice to be able to cluster accurately. With the data set used in this study (students online discussion forum activities), only the Expectation Maximization algorithm [15] provided similar accuracy in predicting student’s success compared to supervised classification algorithms [17].

Ahadi, Lister, Haapala, and Vihanvainen, [1] [1] reported between 70 and 80% percentage of accuracy in predicting low-performing and high-performing students by using classifier based on decision trees. In the study conducted by Sisovic at all, researchers used data extracted from a learning management system called Mudri which is based on Moodle. Another group of researchers using K-means clustering algorithm reported that general lack of activity is the main characteristics of students with tendency to fail, both in terms of attention and participation [18]. The course in study is offered as the combination of online e-learning activities and face to face classroom activities. Researchers study only the first half of the semester data. Mwalumbwe and Mtebe [16] develop a tool to predict students’ performance using learning management data. Their study finds supporting evidence that peer interaction is the most significant factor in predicting student’s performance.

We like to mention that, all above summarized studies used blended course learning management data to provide analysis using either classified or clustering EDM methods. Thus, both the course, which is fully online course, and the EDM method used above are different than the study reported in this paper. Fallowing, in the rest of this section we will be more focusing on studies investigation fully online course as a time series data.

Unlike courses used in above studies, course in the study conducted by researchers Hussein at all was a fully online course, we like the note that this public Moodle data also be used by this study [13]. Thought performing different EDM methods, Hussein at all concluded that there is a strong relationship in between discussion forum view and students excellent score along with workshop view and workshop participation. Although, identifying higher performing, lower performing students via using EDM methods are popular, scientist are becoming more and more interested using the learning management data to identify the learning behaviors as well. Carezo, Esteban, Santillan and Nunez [5] study students procrastinating behaviors in computer-based learning environments. Researchers used Predictive Apriori algorithm as the study analysis method. Carezo at all study highlighted the importance of time management in student’s performance.

Observing students activity as a time series important especially to gain insight in students, study patterns, higher order skills [9]. Study by Cerezo et al. [4] found 4 different clusters from the data collected on Moodle learning management system. Identifying the three variables as the most related with the final marks; time task, followed by days “hand in” and words forums; researchers, to make sense of the results, name the groups as Cluster 1:Non Task or Theory Oriented Group (non-procrastinators),Cluster 2:Task Oriented Group (socially focused),Cluster 3:Task Oriented Group (individually focused),and Cluster 4:Non Task Oriented Group (procrastinators). Cerezo at all used Expectation Maximization clustering and k-means to extract these four clusters.

Another learning pattern analysis study performed by Fatahi et al. [10]. By employing Generalized Sequential Pattern Mining (GPS) algorithm [2], Fatahi et al. successfully identified frequent behavioural sequences unique and can separate learners with different learning style in each dimension of MBTI with high accuracy [7].

Greene and Cunningham Dynamic Time Warping (DTW) as an appropriate distance measure to cluster student’s behavior an an appropriate distance measure to cluster student’s behaviour patterns [14]. Greene and Cunningham choose 3 weeks period around each assignment. These timelines were divided into 12 hours of buckets activity counts. Greene and Cunningham successful identified 7 activity patterns that can be observed around each assignment as: procrastinations- unmotivated- strugglers-steady- hard workers- strategist and experts as per these groups relations with performance.

Based on our review, most scholars preferred using Moodle data as static data in their analysis. Again, most scholars prefer performing the analysis in pass/fail dimension. Also, most of these studies are performed on hybrid course offering, where lots of interaction also goes on in face to face part of the
Therefore, online interaction might be limited on capturing the student full course related activities. We believe it is important to understand the role of each component in learning process in order to provide timely customized feedback.

3 STUDY DESIGN

3.1 Data Set

Digital nature of used LMS in today’s educational world making it possible for educational organizations to capture the data that was not possible in face to face learning environments. Moodle captures wealth of data to provide information to educational organizations to understand the need and help students to improve performance.

In this study, we propose to understand students need by analyzing Moodle data from, specifically two Moodle tables: i) mdl_grade_grades_histor and ii) mdl_logstore_standard_log. First table stores students’ grades received from all activities, quizzes, workshops, forums, discussion with the timestamp, along with max grade student can received from this activity. Second table captures all the activities students performed in the system; viewing a course, checking the grade, viewing a discussion in forum, etc.

The data set used for this study is collected from the free online course (Teaching with Moodle) [11][12]. course is given between 7th of August to 4th of September 2016. Data is anonymized and made publicly available for research purposes. Each student is given a unique anonymized username and each action on every unique instance of the interaction captured accordingly. Number of students enrolled the course is 6119, however only %12 of the student fully completed the course. We choose to perform analysis only on active students – students who participated at least one graded activity.

For this study, students’ grade, and activities are examined. Activities performed on Modules such as Mod_quiz, Mod_page, Mod_Data, Mod_forum, Mod_lesson, etc are classified as per their action identifier. If action is only view it is classified as “Attention”. All other actions such as; uploaded, updated, submitted are classified as “participation”.

3.2 Analysis flow and outcomes

Two branches of analysis have been performed in analyzing data. Activity pattern analysis on Attention and Participation dimension on one branch (Identified as “a” path in the steps detailed below). Grade based analysis on another branch (Identified as “b” path in the steps detailed below). Figure 3 summarizes steps and analysis performed.

Step 1- Selecting active students: Students who are taken at least one graded activity is selected as active students and included in the analysis. Table mdl_grade_grades_history table is used for this step and final grade value is checked.

As a result; 1202 students are chosen as active students out of 6119 students.

Step 2a- Select log entries for active students: From the log file active student entries are selected for further analysis. mdl_logstore_standard_log file.

Step 3a: Grouping activities into “Attention/Participation dimension Entries selected from log files belong to active students are grouped into two main groups: “Attention” and “Participation” All the action entries in “viewed” category by students regardless of the module activity taken upon, forum, book chapter, viewing the grade etc are grouped under attention activity. The rest of the activities are grouped under participation.

To verify correlation of attention and participation activity series, individual case analysis performed on selected students. Below are sampled time series plots of individual students on Attention and Participation dimension.
Visual time series analysis on selected students provided insight on attention and participation activity series being related. Next step is to explore if separated cluster analysis providing complimentary insight.
Step 4a: Applying time series clustering learning to identify emerging patterns: It is mentioned in analysis techniques section that deciding on the distance measure and number of the clusters are important decisions in time series learning. With three distance measure mentioned; DTW, SBD and DTW_basic clustering performed with cluster numbers 3 to 7. Clustering validation indices are examined to find the best possible values out of the nine possible combination. Based on our analysis following are the combination giving the best CVSs:

- For Attention Time Series: DTW_basic clustering with cluster number 5
- For Participation Time Series: DTW clustering with cluster number 6

Below are the cluster groups and centroids for student’s attention and participation dimensions for selected best cluster validation indices (CVI) combinations and their centroids:
Figure 3 Students Attention Based Clusters and Centroid Plots.
Step 2b: Capture active students’ grades: After completing clustering activity on student’s attention & participation dimensions, we performed analysis on students grades to compare and gained performance-based insight on emerging pattern. Based on active students found Step 1, active students’ grades are captured from table mdl_grade_grade_history table.

Step 3b: Calculate Final Grade and Performance Measurement: We decided to perform grade-based clustering on two variables:
Students sum of final grade that is received by graded activities that attended
Students final grades / total raw grade that student can receive as performance

**Step 4b:** We decided to perform grade-based clustering on two variables. Clustering based on students Final Grade & Performance.

Following are the cluster size and means after K-means clustering applied: i) Cl 1: S=432 FGM=432 PM=0.71 ii) Cl 2: S=348 FG=29 P=0.74 iii) Cl 3: S=6 FGM=1165 PM=0.81 iv) Cl 4: S=174 FGM=314 PM=0.60 v) Cl 5: S=123 FGM=583 PM=0.74 vi) Cl 6: S=138 FGM=175 PM=0.52

**Step 5:** The last step of our analysis is comparing all emerging clusters and interpreting them. To do that we grouped all clusters in the following order: I) performance ii) attention ii) participation. Following are the groups and their sizes emerged.

**Figure 5 Emerging Cluster Groups Alignments and Counts.**

4 RESULTS

In this section we like to label the clusters emerged in performance-based clustering to better interpret the study outcomes.

Cluster # 1: “Comfortable”: Cluster grade mean is 432, size is 413.

Best Alignments: Attention clusters 3 & 5 & 2 and Participation clusters 5 & 2. Both these Attention and Participation clusters show stable high-volume activity counts on cluster centroids.

Clusters #2: “Failing”: Cluster grade mean is 29, size is 348

This group is not engaged and failing, unfortunately they are the second biggest group.

Best Alignments: Attention cluster 5 and Participation cluster 6. These cluster centroids shows the least number of spikes and smallest activity volume on their centroids.

Cluster #3: “Experts”: Cluster grade mean is 1165, size is 6

This groups seems to find this course way below their level.

Best Alignments: Attention cluster5 and Participation cluster5, with the size of 6. It is hard to make a generalization on them.

Cluster # 4: “Managing”: Cluster grade mean is 314 size is 174.

Considering class mean is FG is 314. This group is managing.

Best Alignments: Attention cluster 5 and Participation cluster 6. We may speculate that this group with more attention and activity actions, this group may easily increase their grades.
Cluster # 5: “Doing very well”

Cluster mean is 583, size is 123

This group is doing very well with cluster mean is well above class mean.

Best Alignments: Attention clusters 2 & 3 and Participation clusters 2& 5. These Attention and Participation clusters shows evenly distributed multiple spikes with steady on-going activities over the course duration.

Cluster #6: “Struggling”. Cluster grade mean is 175, size is 138.

This is a group that instructors may like to provide customized help. 109 students from this in trouble grade-based group.

Best Alignments: Attention cluster 5 and Participation cluster 5. Participation cluster 5 is showing not bad study activity counts, also align with strong groups Attention cluster 5 however, shows very little no none activity during early phase of the course.

4.1 Discussion

We have identified cluster 2, 6 and potentially 4 as cluster in need based on their grade performance. Performance cluster 4 “experts” are exceptions. Attention clusters 4 and 3 centroids show short term high but the highest level of attention volume. Participation cluster 1 seems similar characteristics.

Attention cluster 5 is noticeable as the cluster deserves more investigation. It is aligning with grade-based clusters identified as in need of help; 2 & 6. Participation clusters 6 and 5 are aligning with those groups.

One unique characteristic of Attention cluster 5 centroid is having minimum to no activity in the early period of the course. Participation clusters 5 and 6 do not demonstrate such difference compare to other participation clusters. This observation suggests that “Attention” activities in the early period of the cluster may provide stronger indication of student’s engagement and performance of the course, more than participation activities.

5 CONCLUSIONS

To give direction and to foster EDM, US Dept. of Education published a report “Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief”. This report looks into how educational data mining and learning analytics can be used to improve teaching and learning [8]The report states that “Educational data mining and learning analytics have the potential to make visible data that have heretofore gone unseen, unnoticed, and therefore unactionable.”

The study reported in this paper is in line with the goal of many educational organizations [19] to establish insight and enable educators to take supportive action through EDM. Our study suggests that educational organizations might identify attention and participation patterns using EDM and take note of students only participating in mandatory activities, without paying attention to course administrative information nor the content, discussions. This pattern might be treated as an indication of students lack engagement in the course. As a result, necessary actions can be taken to foster student’s engagement.

Analysing students “Attention” data, in terms of identifying content, administrative information or discussion might also provide valuable information in terms of providing information on content efficiency and overall students interest. In this way, instructors may choose to follow students’ interest in real-time and provide more similar content or bring students attention to the ones that do not get much attention.

We are hoping that through providing a new perspective of interpretation of students “Attention activities”, this paper provides a novel contribution to the EDM research field.

REFERENCES


