Time Series Clustering of Moodle Activity Data

Ewa Młynarska, Derek Greene, Pádraig Cunningham

Insight Centre, University College Dublin
{ewa.mlynarska,derek.greene,padraig.cunningham}@insight-centre.org

Abstract. Modern computer systems generate large volumes of log data as a matter of course and the analysis of this log data is seen as one of the most promising opportunities in big data analytics. Moodle is a Virtual Learning Environment (VLEs) used extensively in third level education that captures a significant amount of log data on student activity. In this paper we present an analysis of Moodle data that reveals interesting differences in student work patterns. We demonstrate that, by clustering activity profiles represented as time series using Dynamic Time Warping, we can uncover meaningful clusters of students exhibiting similar behaviours. We use these clusters to identify distinct activity patterns among students, such as Procrastinators, Strugglers, and Experts. We see educators as the potential users of a tool that might result from this research and our preliminary analysis does identify scenarios where interventions should be made to help struggling students.

1 Introduction

The availability of log data from virtual learning environments (VLEs) such as Moodle presents an opportunity to improve learning outcomes and address challenges in the third level sector such as high levels of student dropouts [1,14]. Research has shown that certain activity patterns are potentially indicative of good student performance. At the most basic level, it is typically the case that higher levels of activity in learning environments correlates with good grades [4], work submitted close to the deadline is less likely to score well [2] and that evening activity is a better predictor of good performance than daytime activity [4].

Since this kind of correlation analysis entails aggregating Moodle activity data into counts and looking for correlations with respect to these counts, it is expected that some information is lost in this aggregation process. We previously discussed this issue in our preliminary work [6]. To explore it in more depth, we propose representing a student's efforts as a complete time-series of activity counts. We analyse Moodle data from Computer Science courses at University College Dublin (UCD), Ireland, and seek to identify patterns and relationships between more than one attribute that might lead to a student failing a course, with a view to addressing low student retention levels.

Acknowledgments: This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

A major potential benefit of this would be to introduce mechanisms identifying issues in the learning system early during the semester, supporting interventions and changes in the way in which a course is delivered.

Our Moodle system provides us with activity data, submission timestamps, and grade outcomes. Since we are not dealing with course delivery through MOOCs, activity levels online may be sparse and substantially different from course to course, depending on the nature of the material. Therefore, in order to facilitate the performance prediction on less structured systems, we need methods incorporating multiple features to deal with the sparsity problem.

As a solution, we present a method for mining student activity on sparse data via Time Series Clustering. This procedure reveals patterns characteristic of certain student behaviours in an assignment, which we can then relate to high or low grades. We explore the use of Dynamic Time Warping (DTW) [8,9] as an appropriate distance measure to cluster students based on their activity patterns, so as to achieve clustering indicating more structured activity patterns influencing students' grades. DTW allows two time series that are similar but out of phase to be aligned to one another. To gain a macro-level view regarding whether these patterns occur across all assignments, we subsequently perform a second level aggregate clustering on the clusters coming from each assignment. This results in seven prototypical behaviour patterns (see Figures 2 and 3), that we believe can lead to better understanding of the behaviour of larger groups of students in VLEs.

In the next section we provide a summary of previous research on educational data mining that is relevant to our work. Section 3 describes the Moodle data that is analysed in our paper. The time series clustering methodology is described in Section 4, and the corresponding results are presented in Section 5. The paper finishes with some conclusions and plans for future work in Section 6.

2 Related Work

In contrast to the majority of VLE-based research, in our work we do not target MOOCs. Rather, the courses analysed are in-class courses, which are partially supported by a Moodle system.

A large amount of previous research in this area relates to specific log types, which are most predictive for a single dataset [5,13,11]. This makes it difficult to generalise those methods to systems where the type and volume of Moodle activity can vary significantly. Therefore our focus is on developing a method that can provide insight into sparsely populated datasets that do not necessarily include wide range of activity types.

Activity data is often used in projects dealing with student performance prediction and intervention, such as in the Purdue Signals platform [10]. The main feature of the platform is a warning signal about potential future performance as well as "Check My Activity" and "Grade Distribution" tools [7], which show a student's progress in comparison to their peers. Such intervention systems can raise concerns about privacy and lack of effectiveness [12].

Table 1. Moodle data statistics showing the structure of the two analysed data sets, corresponding to two semesters of the same academic year

Sem.	*Courses*	*Assign.*	*Students*	*Subm.*	*Activity*
1	5	20	202	911	43.9k
2	8	32	317	1486	56.5k

Morris [11] discovered using T-tests and multiple regression that students engaging in online activity more frequently are likely to achieve higher grades. The most significant predictors revealed in the study were three variables: number of discussion posts viewed, number of content pages viewed, and seconds on viewing discussion pages. Brooks [3] modelled student interactions with learning resources on MOOCs using time series analysis, with a view to predicting student performance. However, in contrast to our work, the authors proposed constructing time series features analogous to n-grams as used in the field of text mining. Here these features correspond to patterns of accesses to multiple learning resources. They considered four different levels of time granularity: a single day, a three day period, a week, and a month. Their input data was limited to three types of learning resource: lecture videos, forum threads, and quizzes.

Sael [13] considered the task of clustering students into groups with different behaviour "profiles" by applying k-means to Moodle session log data. As opposed to us, their primary focus was on pre-processing methodologies and they did not consider the use of a time series approach to clustering. Cerezo [5] analysed Moodle log data using EM clustering and k-means to extract four different patterns of learning related to different academic achievement. The features they used for their model were: total time spent, time spent in every unit, number of words in forum posts, and number of relevant activities in the learning environment.

In this paper, we also consider clustering, however, we specifically attempt to address the issues caused by diversity in activity structure of the courses and absence of certain types of activity data. This brings us closer to the prediction of future student performance in any general VLE system.

3 Data

The data we gathered for the purpose of this research contains a year of Moodle activity logs for 16 Computer Science courses at UCD for which there was complete assignment submission information available. Student records were anonymised before any analysis was performed, and we considered only data related to grades, deadlines, submission times, and Moodle activity. The full dataset contains 95 assignments in total. Due to the differing nature of 1-week and 2-weeks assignments, these two sets should be treated separately. In the remainder of our analysis, we took into account only 2-weeks assignments due to their longer and richer time series. The final data analysed in this paper comprises 13 courses and 52 assignments, which are naturally split into two semesters (see Table 1).

All courses in our data are in-class courses, rather than online courses. Often Moodle is used primarily here as a supporting platform for assignment submission and delivery of lecture notes. Due to the fact that not all the courses use forums and each course has a different structure in terms of activity types, this data is far more sparse and limited in comparison to that which might be available for MOOCs. Therefore we choose to analyse all types of activity on Moodle, rather than focusing on a subset of activities.

4 Time Series Analysis

In our preliminary studies [6] we confirmed that features such as the timing of a submission, the level of Moodle activity, and the balance between daytime and evening activity are often predictive of a student's performance in a course. However, these aggregate statistics are static in the sense that they fail to fully capture the temporal signal present in Moodle activity data. We now consider an analysis of this temporal aspect of the data using a time series clustering approach. Our objective is to determine whether a clustering of students produced with this approach can reveal coherent groups of students sharing both common Moodle participation behaviour and similar grade performance. This can ultimately be predictive of bad or good performance.

4.1 Measuring Distance

In order to cluster time-series data, the challenge of appropriately measuring the similarity/distance between pairs of series must be addressed. In the context of our data, this challenge can be explained using the sample illustrated in Figure 1, where four activity time series S1–S4 are shown over 12 time steps (these are discretized versions of the real data shown in Figure 2). We require a distance measure that will correctly identify that for example S1, S2, and S3 are similar without being perfectly aligned – these three series all exhibit spikes in activity early and late in their time series.

Table 2. The Euclidean distance matrix exhibits that series S1 and S3 are most similar, while S1 and S2 most dissimilar. The Pearson correlation similarity matrix indicates that series S1 and S3 are identical. The DTW distance matrix shows that S2 and S3 are identical, while S1 and S2 are the next most similar series.

	Euclidean distance				**Pearson correlation**				**DTW distance**			
	S1	***S2***	***S3***	***S4***	***S1***	***S2***	***S3***	***S4***	***S1***	***S2***	***S3***	***S4***
S1	0	5.1	1.4	4.7	1	-0.2	1	-0.14	0	1.4	1.4	4.7
S2	5.1	0	4	3.5	-0.2	1	-0.2	-0.14	1.4	0	0	2
S3	1.4	4	0	3.5	1	-0.2	1	-0.14	1.4	0	0	3.5
S4	4.7	3.5	3.5	0	-0.14	-0.14	-0.14	1	4.7	2	3.5	0

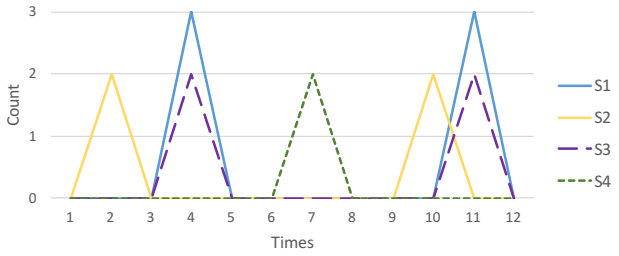


Fig. 1. Sample time series S1–S4 showing activity spikes over 12 time steps.

In Table 2 we see how Euclidean distance and Pearson correlation respectively perform on this task. Euclidean distance misses the similarity between S1 and S2 because the activity spikes are offset in time. The Pearson correlation measure captures the fact that S1 and S3 are perfectly correlated, but still misses the similarity between the pair S1 and S2.

The Dynamic Time Warping (DTW) measure has been proposed for quantifying the similarity between pairs of temporal sequences [8,9]. DTW maps one time series onto another and outputs the cumulative distance between them. The advantage is that it can help to detect similar patterns, even if the activity signatures of those patterns are not perfectly aligned over time.

A standard DTW implementation requires the use of a parameter, the *window size*, that controls the extent of *warping* allowed. A small *window size* will limit the degree to which time can be warped to match two activity spikes. For example in Figure 1, we use DTW with a *window size* of 2 time steps to produce the results shown in Table 2. The pairwise distance values indicate that DTW successfully identifies the similar activity patterns of S1, S2, S3, while series S4 remains dissimilar to the other time series.

4.2 Clustering Methodology

To explore the effectiveness of time-series clustering we run the experiment on all available assignments. We applied time series clustering to group the individual submissions for each assignment into clusters of activity timelines (see sample in Figure 2). We included late submissions in our analysis due to their high importance in the process of understanding students behaviour. Our expectation for the clustering task is that individual clusters will be characterised by different activity patterns, which are indicative of either good or bad grades.

We choose to use all types of logs to generate more activity signal and give better results (lower variance of grades). The graphs show much more activity going on in all cases, which gives richer signal to DTW. This does not have to be the case for all the systems, our Moodle system is not heavily used for online activity, which makes the activity signal quite sparse.

To perform clustering, the Moodle activity data was first transformed into a series of equispaced points in time. In our case, a time series is a three week

timeline – from two weeks before a given assignment submission date until one week after the deadline. These timelines were divided into 12 hour buckets of activity counts. Choosing a small number of hours lead to weaker signals due to insufficient numbers of activities. Following the work in [5,13], we applied k-means clustering using DTW as a distance measure to cluster the timelines for each assignment. For a given number of clusters k, the algorithm was repeated 10 times and the best clustering was selected (based on the fitness score explained below). Due to the fact that DTW is not a true metric, k-means is not guaranteed to converge, so we limited each run to a maximum of 50 iterations.

Fitness measure An important issue is how to construct the fitness function so that it will help in the selection of the best clustering. Such a function needs to take into consideration a number of factors:

1. Two clusters of different sizes might have the same variance value; this issue can to be solved by applying a penalty to smaller clusters.
2. We aim at having the smallest amount of clusters possible to avoid unnecessary partitioning and overfitting.
3. However, we also aim to avoid the situation where our clustering consists of one large cluster and a selection of tiny clusters. In other words, we would like a "balanced clustering" where the variance of the cluster sizes is as small as possible

Based on the requirements above, the fitness score calculation for a clustering generated by k-means consists of three steps:

1. The mean variance of the k-means clustering is calculated using the weighted average of all the clusters' variances, where the weight is based on the size of the cluster. This way the clusterings containing larger clusters with lower variances will be awarded better scores.
2. It is crucial to test the difference between a baseline clustering and actual results to define the significance of the clustering. For that purpose we run multiple random assignments of time series to calculate the expected score which could be achieved by chance for a given number of clusters.
3. To incorporate the baseline comparison in the score, the weighted average variance score from Step 1 is normalised with respect to the random assignment score from Step 2. A good clustering should achieve a low resulting score.

Parameter selection We would expect that the number of clusters k will differ for each assignment, depending on the activity levels of students during the lead up to the assignment deadline. This parameter value will depend on factors such as the number of students taking the course and the level of Moodle activity during the time period under consideration. Therefore, we automatically select k for each assignment as follows. We test a range of candidate values of $[2, k_{max}]$, where k_{max} is defined as total number of submissions for the assignment divided

by 3. This upper limit is based on the fact that we do not consider in our analysis clusters of size < 3. For each candidate value k in this range, we run k-means 10 times and select the run resulting in the minimum value for our fitness score. We then select the final value of k for the assignment which results in the overall minimum grade variance value.

The are two parameters which have to be specified for the DTW distance measure when clustering: 1) the size of the time buckets in the time series (the number of hours or activities which define a single point on the timeline); 2) the *window size* (how much warping is allowed). To check whether there is a bucket size that could be used as a fixed parameter for all assignments, we applied k-means to activity data for the 12 assignments with bucket sizes of 6 hours and 12 hours. For 60% of cases, the 12 hour buckets resulted in clusterings with lower grade variances. Consequently, we use this parameter value for the remainder of our experiments.

To choose the size of the DTW time window, we ran k-means for *window sizes* $\in [0, 3]$. The results did not conclusively indicate that any single *window size* lead to a significant decrease in cluster grade variance, which is unsurprising. In the case of assignments with a small number of time series (student submissions) and a high volume of activity, a small warping area will be appropriate. In cases where there are many time series exhibiting little activity, it will be difficult to differentiate between the series and so a larger window size will be more appropriate. Based on this rationale, we believe that *window size* selection should be run for each assignment separately when applying this type of analysis in practice.

5 Discussion

5.1 Prototype Pattern Selection

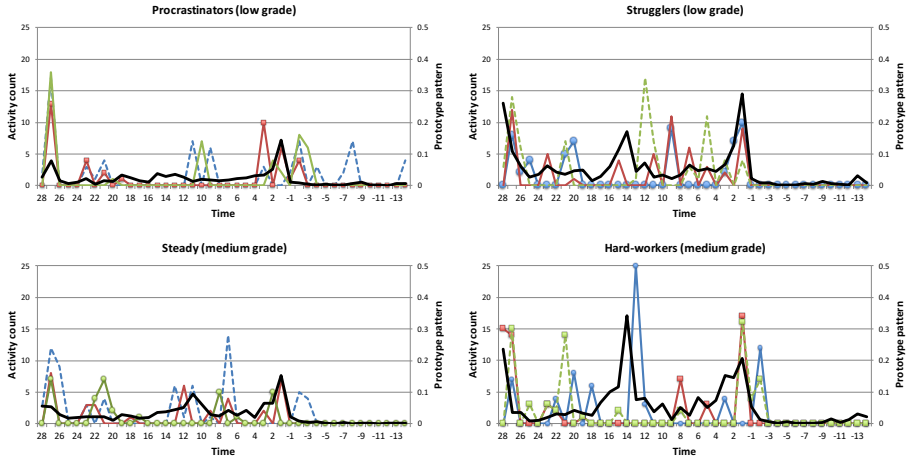
We applied the time series clustering methodology described previously in Section 4.2 to the activity data for each of the assignments in the Semester 1 dataset. These clusterings appeared to show a number of frequently-appearing patterns across different courses. To gain a deeper insight into these patterns, we applied a second level of clustering – i.e. a clustering of the original clusters from all assignments. To support the comparison of clusters originating from different modules, the mean time series for each cluster was normalised. Based on the associated assignment scores, these normalised series were then stratified into low, medium, and high grade groups. We subsequently applied time series clustering with $k = 4$ and *window size* 1 to the normalised series in each of the stratified groups. Grade group names chosen by us were motivated by the behavioural pattern of students and some of them were inspired by previous research [5].

This second level of clustering revealed seven distinct prototypical patterns, which are present across multiple assignments and courses:

Table 3. The proportion of prototype patterns among 1st and 2nd Semester submissions after stratification into each grade group, where m_1 is the mean assignment grade and m_2 is the average of the mean and maximum assignment grade.

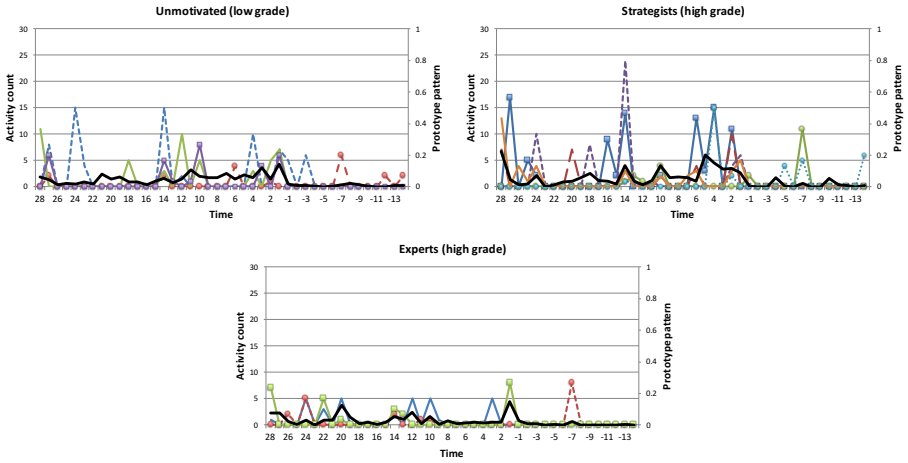
Grades group	Activity patterns	1st Sem	2nd Sem
LOW $grade < m_1$	Procrastinators	16%	6.2%
	Unmotivated	15%	12.8%
	Strugglers	13.5%	23%
	Outliers	*0.6%*	*10.5%*
MEDIUM $m_1 \leqslant grade < m_2$	Steady	23.6%	12.3%
	Hard-workers	13.5%	9.8%
	Outliers	*11%*	*15.6%*
HIGH $grade \geqslant m_2$	Strategists	2.3%	1.8%
	Experts	2.1%	3.4%
	Outliers	*2.4%*	*4.6%*

Fig. 2. The samples of 4 out of 7 prototype activity patterns that occurred in Assignment #1. The black trend-line represents the prototype pattern. The group of time series represents activity of students who submitted their work for Assignment #1. Negative labels on the Time axis symbolise period after the deadline.



1. *Procrastinators (low grade)*: Students achieving low grades, barely active on Moodle and showing high activity only close to assignment deadlines.
2. *Unmotivated (low grade)*: Students active on Moodle during the practical sessions and engaging less than more successful peers, getting poor results.
3. *Strugglers (low grade)*: Students doing some work, the level of activity rises towards the deadline

Fig. 3. The samples of 3 prototype patterns found in Assignment #2. The black trend-line represents the prototype pattern and time series represent activity of students who submitted their work for Assignment #2. Negative labels on the Time axis symbolise period after the deadline.



4. *Steady (medium grade)*: Students getting satisfactory grades, exhibiting regular and decent amount of activity earlier before the deadline with small spike at the deadline.
5. *Hard-workers (medium grade)*: Students who are active during practical sessions, working a lot not only during the deadline but regularly before and getting quite good grades.
6. *Strategists (high grade)*: Students achieving the highest grades, who are characterised by relatively high and regular spikes. This patter indicates that regular, strategic work will pay off.
7. *Experts (high grade)*: This group exhibits low-level frequent activity with no high spikes around deadlines. These students are likely to understand the subject well and follow their own path.

The students rewarded with low grades were the second largest group of submissions having the smallest average activity per submission. The first out of 3 largest clusters was a group barely active on Moodle, performing submission activity at the deadline only (See Figure 2). As mentioned by Cerezo [5], these could be labelled as Procrastinators. The black trend-line on the graph depicts prototype activity pattern and group of time series represents activity of students from the sample cluster. The second largest group consists of Unmotivated students, appearing during the practical sessions and engaging less than their successful peers, achieving low grades as a result (See Figure 3). These students may need extra assistance to increase their efficiency during the practical sessions. The times of practical sessions were obtained from the school office and used as an external factor for deeper understanding of students behaviour. The

third biggest group contains those students doing the minimum amount of work and showing larger activity towards the deadline. These patterns were observed in the two example cases of two individual assignments shown in Figures 2 and 3 respectively.

Across all assignments, the largest group consisted of those achieving medium grades. Here the most common pattern showed students exhibiting a reasonable amount of activity earlier before the deadline. These so-called Steady students achieve satisfactory grades. The second most common pattern for students at this grade level, the Hard-workers, are very active during the practical sessions, working regularly before the deadline. These students are motivated to tackle problematic material and get awarded with quite good grades (See Figure 2). Overall, the medium grade group exhibits the highest amount of activity per student.

At the high grade level, the most common pattern is characterised by regular, relatively high spikes in activity. This indicates that regular, strategic work will pay off. Another popular group are Experts, with small frequent activity spikes and no high spikes around deadlines. This group of students knows the subject well and follows their own path.

5.2 Analysis of Semester 2 Data

When we examine the dataset from the second academic semester, the courses mostly exhibit similar clusters from the first semester. The percentages reported in Table 3 indicate that for the Low Grade group, the Strugglers were most common and Procrastinators were less common. The most significant outlier cluster however, included quite a large number of students from different courses (7 courses and 13 assignments) working hard during practical sessions avoiding spikes at the deadline (see Figure 4). This is likely to be a case where intervention is required, due to the high number of students achieving poor results.

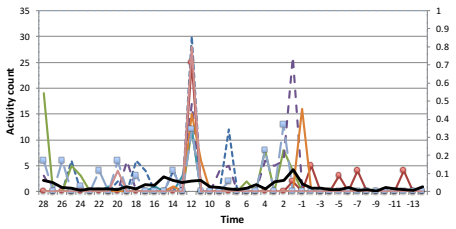
At the medium grade level, the same common patterns appeared as in Semester 1, which are Steady students and Hard-workers. The other two large outlier groups of students worth mentioning here are students highly active a week before the deadline and students with a significant activity spike at the deadline. At the high grade level, the Strategists are less prevalent than in the first semester. However, Experts still appear frequently.

5.3 Changes in activity across different assignments

As a further test of the meaningfulness of these time series signatures we checked to see if the behaviours persisted between assignments. To do this we looked at the distance between activity patterns between successive assignments. If the patterns are meaningful, patterns for the same student should be more similar than those for different students. We expect:

$$\frac{1}{n}\sum_{i=1}^{n} Dist(a_{i1}, a_{i2}) < \frac{1}{n(n-1)}\sum_{i=1, i\neq j}^{n} Dist(a_{i1}, a_{j2}) \quad (1)$$

Fig. 4. Example time series representing students achieving low grades likely requiring intervention.



where $Dist(a_{i1}, a_{j2})$ is the DTW distance measure between signatures for assignment 1 for student i and assignment 2 for student j. The results of the analysis proved the hypothesis to hold true for all the 28 cases with an average value on the left hand side of the equation equal to 20.5 and one on the right hand side equal to 22.5.

The students who change their behaviour from one assignment to another are expected to exhibit a change in grade, too. The results of the correlation analysis between the change in grade (g_{i1} - g_{i2}) and DTW distance between time series in successive assignments $Dist(a_{i1}, a_{i2})$ confirms that the change in the activity pattern influences the grade.. Over 23 out of 28 pairs of consecutive assignments got positive correlations with an average value of *0.15*. The findings indicate existing potential in the analysis of students' activity changes across assignments and modules.

6 Conclusions

In this work we examined a large VLE dataset for which we performed a time series clustering of students achieving low and high grades, in order to observe the different behavioural patterns characteristic of these different groups. The clustering of activity data revealed several distinct behavioural patterns that highlight the relationship between VLE activity in relation to assignments and their final grades. Our study also shows that using VLE data can be used to predict successful students and unmotivated ones at an early stage.

While we did observe significant numbers of outliers, the relevant courses should be considered using a separate analysis to determine whether external factors are at play (e.g. continuous assessment rather than discrete assignments, lack of material provided on Moodle for a specific course). Finally, it is worth exploring anomalous clusters in the context of activity outside that assignment or course. Can this kind of analysis uncover bad interactions when assignment deadlines overlap? We are currently in the process of extending our research to address the behavioural patterns of knowledge seekers in alternative, more complex learning environments.

References

1. L. Agnihotri. Building a student at-risk model: An end-to-end perspective. In *Proc. 7th International Conference on Educational Data Mining*, 2014.
2. D. Arnott and S. Dacko. Time of submission: An indicator of procrastination and a correlate of performance on undergraduate marketing assignments. In *Proc. 43rd EMAC Conference*, 2014.
3. C. Brooks, C. Thompson, and S. Teasley. A time series interaction analysis method for building predictive models of learners using log data. In *Proc. 5th International Conference on Learning Analytics And Knowledge.*, ACM, 2015.
4. K. Casey and P. Gibson. (m)oodles of data: Mining moodle to understand student behaviour. In *Proc. 3rd Irish Conf. on Engaging Pedagogy*, 2010.
5. R. Cerezo, M. Sanchez-Santillan, J.C. Nunez, and M.P. Paule. Different patterns of students' interaction with moodle and their relationship with achievement. In *Proc. 8th International Conference on Educational Data Mining*, 2015.
6. E. Młynarska D. Greene, P. Cunningham. Indicators of good student performance in Moodle activity data. arXiv:1601.02975, 2016.
7. J. Fritz. Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education*, 14.2:89–97, 2011.
8. T. Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.
9. T. Giorgino. Computing and visualizing dynamic time warping alignments in R: the DTW package. *Journal of Statistical Software*, 31(7):1–24, 2009.
10. M. D. Pistilli K. E. Arnold. Course signals at purdue: using learning analytics to increase student success. In *Proc. 2nd International Conference on Learning Analytics and Knowledge. ACM*, 2012.
11. L. V. Morris, C. Finnegan, and S. Wu. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8.3:221–231, 2005.
12. M. Sharkey R. Barber. Course correction: Using analytics to predict course success. In *Proc. 2nd International Conference on Learning Analytics and Knowledge.ACM*, 2012.
13. N. Sael, A. Marzak, and H. Behja. Multilevel clustering and association rule mining for learners' profiles analysis. *International Journal of Computer Science Issues (IJCSI)*, 10(3), 2013.
14. G. Siemens and P. Long. Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review 46 (5)*, 2011.