Time Series Analysis of VLE Activity Data

Ewa Młynarska, Derek Greene, Pádraig Cunningham

Insight Centre, University College Dublin

ABSTRACT

Virtual Learning Environments (VLE), such as Moodle, are purpose-built platforms in which teachers and students interact to exchange, review and submit learning material and information.

In our work, we examined a complex VLE dataset from a large Irish university in an attempt to characterize student behavior with respect to deadlines and grades. We demonstrate that, by **clustering activity profiles** represented **as time series using Dynamic Time Warping**, we can uncover meaningful **clusters of students exhibiting similar behaviors** even in a sparsely-populated system.

We use these clusters to identify **distinct activity patterns** among students, such as Procrastinators, Strugglers and Experts. These patterns can provide us with an **insight into the behavior of students**, and ultimately help institutions to exploit deployed learning platforms so as to better structure their courses.

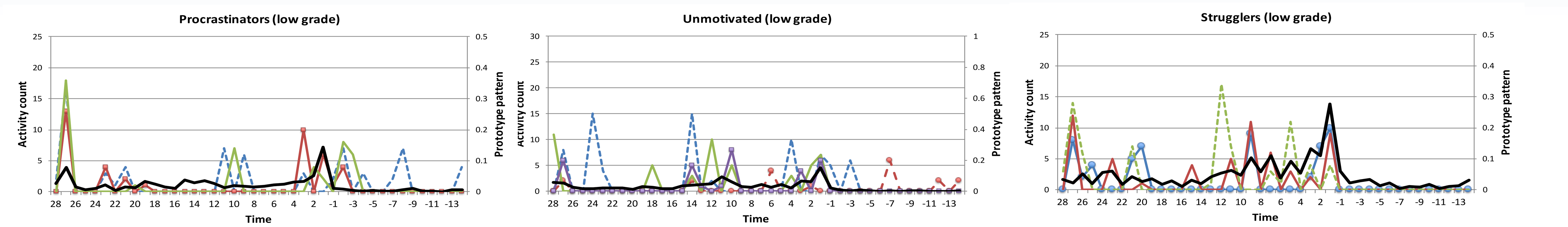
OBJECTIVES

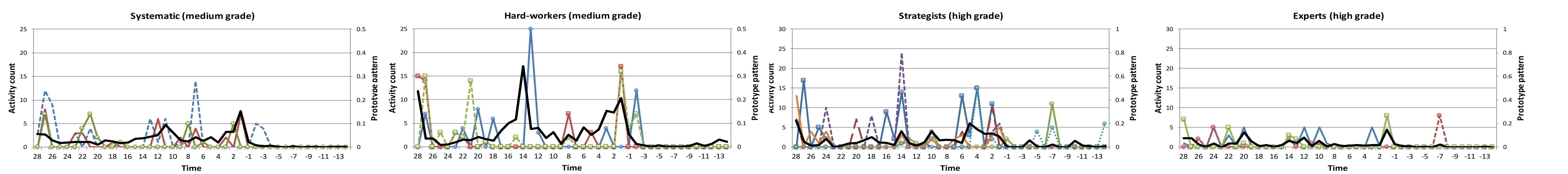
- Levels of activity in learning environments positively correlated with good grades
- Work submitted close to the deadline less likely to score well
- Evening activity a better predictor of good performance than daytime activity
- Loss of information when aggregating Moodle activity data into counts and looking for correlations with respect to these counts
- Sparsity of activity levels online
- Difference in activity levels from course to course, depending on the nature of the material
- Usage of only logs predictive for particular data set
- Methods incorporating multiple features needed to deal with the sparsity problem
- Concerns about privacy in intervention systems

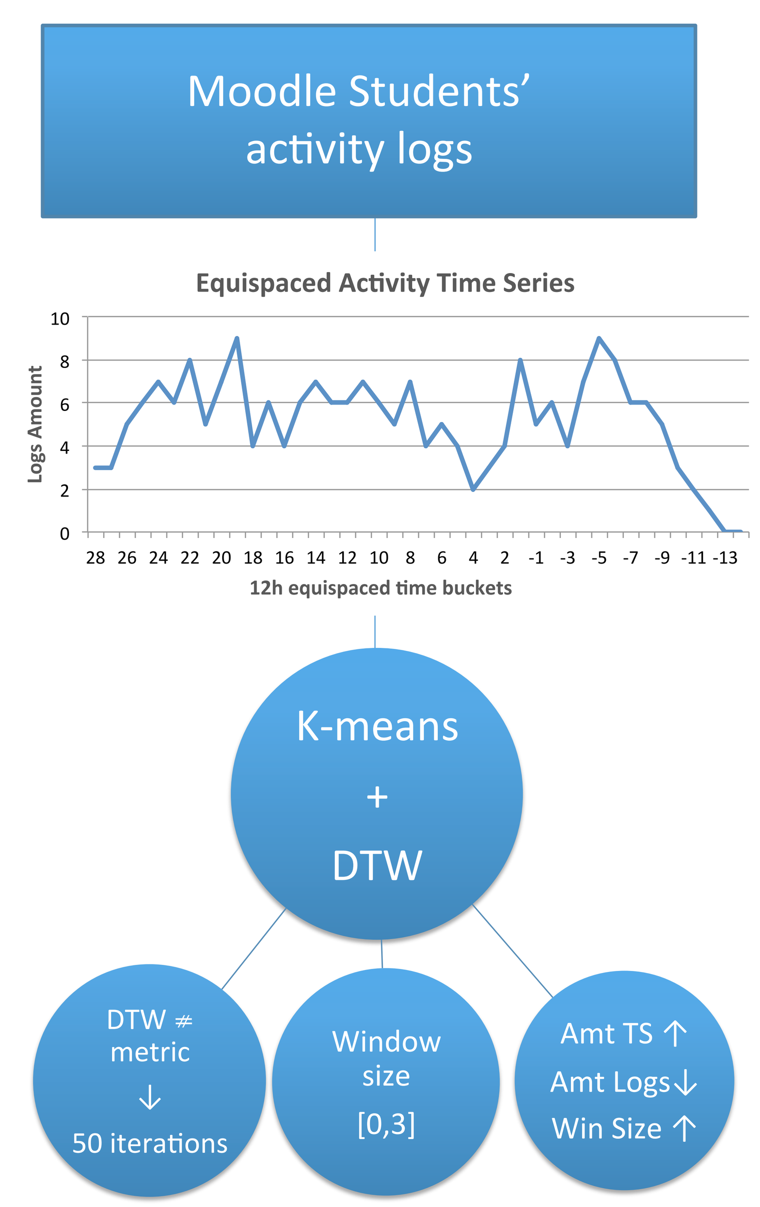
Solution:

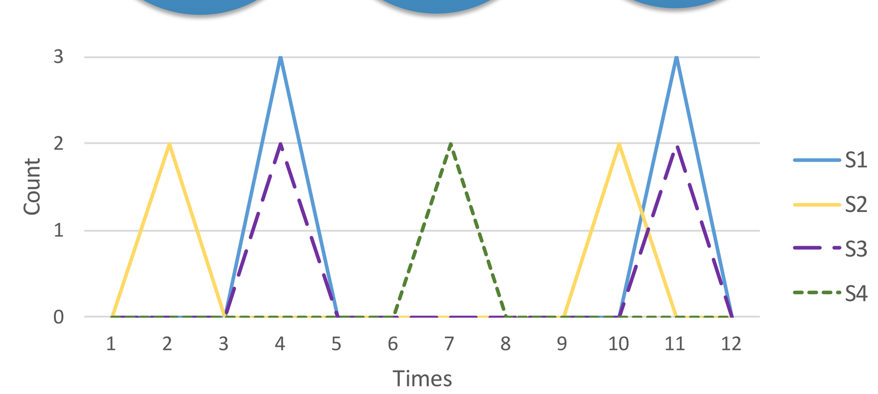
- Representing student's efforts as a complete time-series of activity counts
- Mining student activity on sparse data via Time Series Clustering

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









- **The Euclidean distance** - S1 and S3 are most similar, while S1 and S2 most dissimilar.
- **The Pearson correlation** - series S1 and S3 are identical.
- **The DTW distance** - S2 and S3 are identical, while S1 and S2 are the next most similar series.

TIME SERIES ANALYSIS

Fitness measure:

1. Mean variance of the *k*-means clustering calculated using the **weighted average of all the clusters' variances**, where the weight is based on the size of the cluster

   →Better scores awarded to the clusterings containing larger clusters with lower variances

2. Multiple random assignments of time series run to calculate the expected score which could be achieved by chance for a given number of clusters

   →The difference between a **baseline clustering** and actual results tested to define the significance of the clustering.

3. **Normalization of the weighted average variance** score from Step 1 with respect to the random assignment score from Step 2.

   →A good clustering characterized by a low resulting score

	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

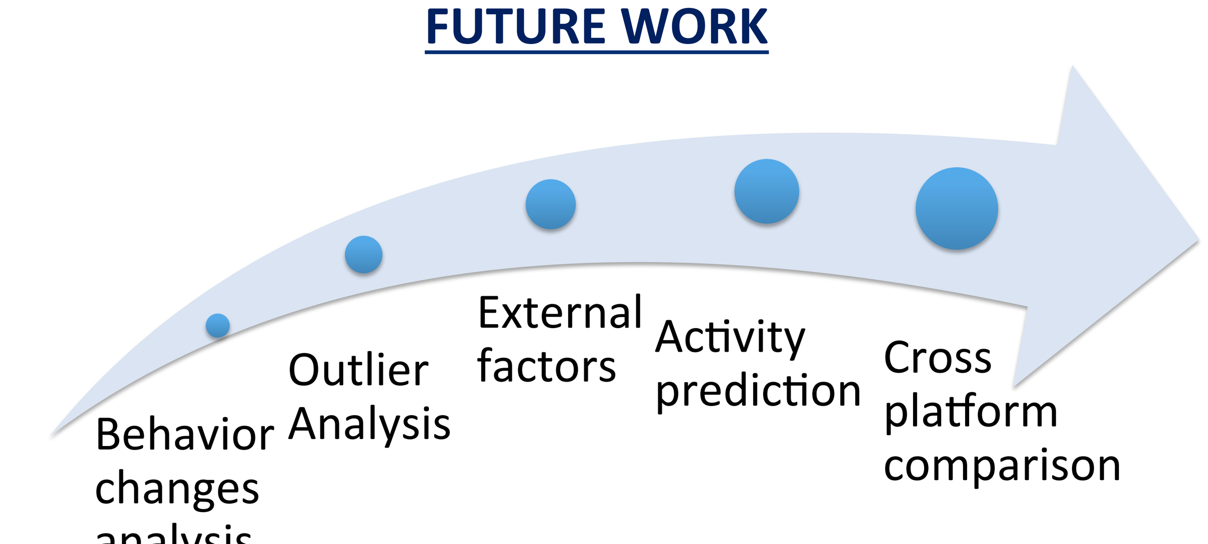
Table 1. The structure of two analyzed data sets, corresponding to two semester of the same academic year.

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

Table 2. 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.

Grade 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≤grade<m_2*	Systematic	23.6%	12.3%
	Hard-workers	13.5%	9.8%
	Outlier	11%	15.6%
HIGH *grade≥m_2*	Strategists	2.3%	1.8%
	Experts	2.1%	3.4%
	Outliers	2.4%	4.6%

FUTURE WORK



CONTACT

Ewa Młynarska

Email: ewa.mlynarska@insight-centre.org

ACKNOWLEDGEMENTS

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