Learning Behaviours and the Virtual Learning Environment: Unearthing Behaviour in a ‘Bricks and Mortar’ setting

Lucy Minford[1] and Simon Rudkin *[1]

[1]*School of Management, Bay Campus, Skewen, Swansea, SA8 1EN, United Kingdom*

September 30, 2018

Studies on the use of virtual learning environments (VLEs) fall into three distinct categories, those which capture user behaviours as a prediction of performance, those who take user behaviours to generate early warning systems for potential failure, and those which seek to uncover the effectiveness of particular features. We are motivated in this study by the way these learning behaviours are gathered, adopting a topological perspective to understanding the data bringing persistent homology of point clouds to the learning analytics literature for the first time. Ours is also a study which integrates online data with physical attendance in lectures and seminars to gain a fuller picture of student engagement, something only possible for the small subset of works on traditional ‘bricks and mortar’ universities in which this study sits.

Evidence on the link between learning analytics data and student outcomes is mixed, Iglesias-Pradas et al. (2015), Strang (2016) and Hernández-García et al. (2018) amongst the many to find a lack of statistical significance. Hernández-García et al. (2018) focuses particularly on success in group assignments, demonstrating how even specific measures related to a defined learning activity may not produce significance. Alves et al. (2017) identify access to VLEs as improving performance in a large study of over 6500 students. Gašević et al. (2016) find differences in association across disciplines, but do identify many engagement behaviours that produce significant uplift for performance. Our study does not demand association per se, but it is through the potential linkage with performance that the rationale for understanding more of VLE behaviour is found. Hernández-García et al. (2018) raises an important qualification on extant results; the confounding effects of related variables generating multicollinearity far beyond the levels credited in most studies.

Clustering learning behaviours can be highly informative, helping educators to identify behaviours and to gain a fuller picture within the early warning context Gašević et al. (2016). Cerezo et al. (2016) identifies those who are task-orientated and those who may require different learning approaches to engage; a role for personalised VLEs is posited. Common amongst all of these papers is the use of k-means clustering, an approach data topography

*Corresponding author. Tel: +44 07955109334, Email: s.t.rudkin@swansea.ac.uk

can significantly improve upon. Clustering has the advantage of producing measures that can then be put into second stage regressions combining the data and removing the confounding concerns of Hernández-García et al. (2018); trading a loss of detail with statistical validity we argue persistent homology can maintain more information whilst still delivering a meaningful measure for future association with performance.

Boulton et al. (2018) study of engagement and learning outcomes at the University of Exeter provides an excellent exposition of the challenges facing analysis of data from a 'bricks-and-mortar' institution. Developing the discussion from Agudo-Peregrina et al. (2014) and Gašević et al. (2016) it is argued that controls for departmental cultures, physical learning design, assessments and interaction with the pedagogy must be included in any study. Boulton et al. (2018) extracts these differences from modules across academic disciplines with different cohort sizes and learning designs. Identified significance of such cautions on the generalisation of results from small studies like ours; our proposition is that greater understanding of the data and combination with physical attendance data is an approach for wider adoption which can then be expanded to gain the generalisability achieved with less fine-grained approaches.

Data logging can inform of which content students open, how long they spend looking at it and when this activity takes place; such logs are highly detailed but also very noisy as clicking behaviour and attention are very different things; Lorigo et al. (2008) study of eye movement in webpage viewing demonstrates such. Viewing time aides identification of procrastination as a behaviour class in Cerezo et al. (2016). Strang (2016) takes a more pragmatic approach to attendance, recording presence for a student who logs into the VLE more than once during a given week; this abstracts from the viewing time challenge and informs the approach we take with our analysis. Breakdowns of access, time online and engagement with various activities allows us to produce a multidimensional rich data cloud which is evolving for each student over each week of the module.

Persistent homology, as adopted here offers a more noise robust clustering tool based on the data cloud. Our first demonstration is to show the additional clarity it delivers versus commonly adopted segmentation approaches such as k-means. An excellent exposition of the methodology is provided by Carlsson (2009), whilst opportunities in text mining applied widely in learning analytics can be found in Wagner et al. (2012). Our paper adapts these for the context of module engagement data. Persistent homology views data points as realisations from a larger surface of possibilities, removing any need for distribution assumptions of the type so heavily criticised in learning analytics studies. It is a stable representation robust to small peturbations within the data cloud (Chazal et al., 2013; Chacón et al., 2015).

We demonstrate two processes through which homology can identify student behaviours. Firstly we can use positions within the overall point cloud to construct distance based clusters, this has closest similarity to the k-means and geometric approaches of Cerezo et al. (2016); Gašević et al. (2016) and others. Point locations are described as being in dimension 0, robustness to noise is the primary benefit of persistent homology from this perspective. From observing the homology of individual behaviours through the course we create a persistence landscape for each student measuring it via the landscape norm. We focus on the zero and first dimension homologies; the latter capturing the stability of features within the data as well as the locations of points. A wealth of statistics enable delving deeper into many more dimensions but as we demonstrate basic dimension 0 analysis is sufficient for learning analytics. Rieck and Leitte (2016), Otter et al. (2017) and Assaf et al. (2018) discuss how persistent homology as a clustering technique is addressing complex problems from science and image recognition; we translate these gains and innovations to module engagement data.

Our data is taken from a final year undergraduate economics course which jointly covers theoretical understanding of contemporary issues in microeconomics and macroeconomics; the authors are responsible for one half of the module each. In the studied operation of the module a new approach to the microeconomic component is adopted with a greater use of blended learning techniques, increased online preliminary material, and a new assessment which asks students to create digital media content. For the macroeconomic component operation mirrors the previous year building on positive student feedback and engagement with the physical classes. Within this backdrop the opportunity to observe multiple behaviours is great, and hence we tease out patterns which would be masked in a single delivery style module. Deliveries include online quizzes as studied in Juhaňák et al. (2017) and group-work most recently evaluated in Hernández-García et al. (2018)

The module itself has 130 students, with all having a strong background in the application of economic concepts from their time studying with Swansea University. For many mathematics remains a weakness and the delivery seeks to accommodate these students through the additional support of online materials and a tailored class delivery style. Within the ten week course there are ten two-hour lectures in which the students are both given detailed coverage of topics and interactive opportunities, including competitive quizzes, problems with feedback and the opportunity to contribute to the class discussion. Whilst the number of students is nominally large, sector-wide issues of low attendance percentages and the conditioning of the cohort mean that engagement in lectures is above that which might be expected of an Economics group of this size. These lectures are supported by a series of 6 small group seminars and 4 larger "masterclass" sessions where the students are guided to the literature frontier in one hour exercise based lectures. Seminars are operated ahead of the lectures regularly asking students to reflect on their understanding of material and their weaknesses for the coming lecture[1]

Contributing to the literature on learning analytics we make three key advancements. Firstly we combine VLE data with physical attendance to obtain a stronger measure of engagement that recognises the 'bricks-and-mortar' setting and the complementarity between attendance and use of many of the learning resources on the VLE. Secondly we show how persistent homology increases noise robustness and identifies behaviours within the data with the potential to fine tune success prediction and clustering. Finally through applying persistent homology across the whole dataset, and measuring differences between the topology of individual behaviours, we demonstrate the potential of persistent homology to help leverage the early warning system, success prediction or simply to understand the ways students combine VLE engagement with class attendance in a 'bricks-and-mortar' university environment. Our lessons then spark a research agenda for enhancing understanding across the sector.

[1]Our rationale for this approach follows the learning cycle introduced and promoted in Willey and Freeman (2006); Lawson et al. (2015) and others.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., and Hernández-García, Á. (2014). Can we predict success from log data in vles? classification of interactions for learning analytics and their relation with performance in vle-supported f2f and online learning. *Computers in human behavior*, 31:542–550.

Alves, P., Miranda, L., and Morais, C. (2017). The influence of virtual learning environments in students' performance. *Universal Journal of Educational Research*, 5(3):517–527.

Assaf, R., Goupil, A., Vrabie, V., Boudier, T., and Kacim, M. (2018). Persistent homology for object segmentation in multidimensional grayscale images. *Pattern Recognition Letters*, 112:277–284.

Boulton, C. A., Kent, C., and Williams, H. T. (2018). Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar'university. *Computers & Education*, 126:129–142.

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., and Núñez, J. C. (2016). Students' lms interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, 96:42–54.

Chacón, J. E. et al. (2015). A population background for nonparametric density-based clustering. *Statistical Science*, 30(4):518–532.

Chazal, F., Guibas, L. J., Oudot, S. Y., and Skraba, P. (2013). Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):41.

Gašević, D., Dawson, S., Rogers, T., and Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28:68–84.

Hernández-García, Á., Acquila-Natale, E., Chaparro-Peláez, J., and Conde, M. Á. (2018). Predicting teamwork group assessment using log data-based learning analytics. *Computers in Human Behavior*.

Iglesias-Pradas, S., Ruiz-de Azcárate, C., and Agudo-Peregrina, Á. F. (2015). Assessing the suitability of student interactions from moodle data logs as predictors of cross-curricular competencies. *Computers in Human Behavior*, 47:81–89.

Juhaňák, L., Zounek, J., and Rohlíková, L. (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*.

Lawson, R., Scheepers, M., Taylor, T., and Glavas, C. (2015). Five years on: What has changed in assurance of learning? In *38th Higher Education Research and Development Society of Australasia (HERDSA) Incorporated Conference Book of Abstracts*. Higher Education Research and Development Society of Australasia (HERDSA).

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F., and Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59(7):1041–1052.

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., and Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17.

Rieck, B. and Leitte, H. (2016). Exploring and comparing clusterings of multivariate data sets using persistent homology. In *Computer Graphics Forum*, volume 35, pages 81–90. Wiley Online Library.

Strang, K. D. (2016). Do the critical success factors from learning analytics predict student outcomes? *Journal of Educational Technology Systems*, 44(3):273–299.

Wagner, H., Dłotko, P., and Mrozek, M. (2012). Computational topology in text mining. In *Computational Topology in Image Context*, pages 68–78. Springer.

Willey, K. and Freeman, M. (2006). Completing the learning cycle: The role of formative feedback when using self and peer assessment to improve teamwork and engagement. In *AAEE-Annual Conference of Australasian Association for Engineering Education*. School of Engineering, Auckland University of Technology, Auckland, New Zealand.