ELECTRIC DREAMS
30th ascilite Conference, 1-4 December 2013
Macquarie University, Sydney

Improving retention in first-year mathematics using learning analytics

Yasmin Erika Faridhan
Faculty of Engineering and Industrial Sciences
Swinburne University of Technology
Birgit Loch
Faculty of Engineering and Industrial Sciences
Swinburne University of Technology
Lyndon Walker
Faculty of Life and Social Sciences
Swinburne University of Technology

Despite the importance of mathematical skills in quantitative disciplines, high failure rates in first-year university mathematics subjects have been observed in many parts of the world. Mathematics support provisions are established in many tertiary institutions in order to assist at-risk students to master and pass mathematics subjects. However, while a significant amount of data is being collected on students (e.g. entry scores, backgrounds), their behaviour (e.g. access of support services, engagement with online resources) and their performance (e.g. in assignments, tests), not much analysis is currently done with this data to predict a student's chances of success, and to better guide the services of mathematics support centres and target intervention procedures. This paper reviews relevant literature and describes a proposed research project to improve retention in first-year mathematics using a learning analytics approach.

Keywords: first-year mathematics, mathematics support, retention, learning analytics

Introduction

The lack of mathematics prerequisite skills at tertiary level has been recognised as an issue since the late 1970s and is known as the 'mathematics problem' (Rylands & Coady, 2009). It is a serious problem even in developed countries. Many first-year university students are struggling to pass mathematical subjects, especially those studying in quantitative areas such as engineering and science (Wilson & MacGillivray, 2007). High failure rates in mathematics subjects and lower retention in disciplines with mathematics-intensive subjects have prompted tertiary institutions to set up some form of mathematics support in order to assist these students from failing these subjects. Studies have been undertaken to identify these students with weak mathematics skills and refer them to available mathematics support and intervention schemes. However, regarding mathematics support, so far only a limited number of academic performance variables have been used to determine which students are to be classified as "at-risk" (see, for example, Croft, Harrison, & Robinson, 2009; Lee, Harrison, Pell, & Robinson, 2008).

In the United States and some European countries, learning analytics and educational data mining approaches have been used to predict student performance, identify at-risk students, and set up intervention schemes in order to help students passed their subjects. While many studies have been done in this area, none of them integrated mathematics support variables in their research (e.g. Arnold & Pistilli, 2012; Garcia & Mora, 2011). This concise paper will introduce a project that attempts to close this gap. It will incorporate mathematics support aspects in a learning analytics approach to improve student retention, and to achieve this aim, a new intervention strategy to assist at-risk students in first-year mathematics will be configured. This paper is organised as follows. Firstly, it reviews the relevant literature on mathematics support and learning analytics, and critically evaluates the papers to justify the need for this project. Secondly, it describes the project plan; and finally, it concludes with expected outcomes of the project.

Evaluating mathematics support and its role to increase student retention

Mathematics support provision was initially set up to assist at-risk students to ease their transition into university. As a general term, mathematics support provision normally ranges from introductory mathematics courses offered before the semester begins, drop-in learning/help centres, help desks, pre-booked individual appointments, exam revision support, or peer-assisted support (Parsons, 2008), to online support that provides materials (lecture notes documents as well as videos) that can be accessed from anywhere via the internet in the students' own time (Loch, Gill, & Croft, 2012). Mathematics support is now common practice in many universities in many parts of the world. Drop-in learning centres in particular have been established mostly in the UK, Ireland, and Australia, and they now cater for students with a range of different mathematical abilities (Gill, Mac an Bhaird, & Ní Fhloinn, 2010).

Many studies have been undertaken to measure the impact of mathematics support services on students' grades and retention rates. Factors such as past examinations, student grades, diagnostic tests, and whether or not students make use of a range of available mathematics support were used. The last factor was found to be important as it seems students who make use of mathematics support tend to perform better in mathematics, as evidence provided by, for example, Croft et al. (2009) and Mac an Bhaird, Morgan, & O'Shea (2009).

Studies into the effectiveness of mathematics support in improving student progression and retention involved more complex issues than academic performance variables alone. Patel & Little (2006) and Lee et al. (2008) concluded that in order to improve student progression and retention, diagnostic tests should be followed by mathematics study support provision.

Evaluating learning analytics and its applications

The 21st century has seen the collection of data expanding as a result of extensive uses of the web for learning, and with this the term "analytics" is now widely used in many areas. The application of analytics in education is often referred to as educational data mining and learning analytics (Romero & Ventura, 2007; Siemens et al., 2011). These terms are similar in many ways, and overlapping research studies in the two disciplines were observed (e.g. Romero-Zaldivar, Pardo, Burgos, & Delgado Kloos, 2012). However, as a rule of thumb, Siemens & Baker (2012) explained that educational data mining focuses on technical aspects of computing algorithms and automated discovery, while learning analytics focuses more on the educational side, i.e. empowerment of human resources (instructors and learners).

There have been a few implementations of learning analytics projects on campus. The most recognized project is Course Signals at Purdue University, which was automated in 2009. The Course Signals features real-time feedback, early intervention, as well as frequent and on-going feedback, which are essential in identifying at-risk students, both for the faculty (lecturer, tutor, retention coordinator) and the students themselves. For students, it is very simple to comprehend; each student receives 'signals' similar to traffic signals (red, yellow, or green) in their Blackboard site regarding to each course s/he is currently taking. Lecturers can track students with yellow or red signals as early as from the second week of the semester, and can therefore decide early on what kind of intervention is suitable to help a particular student (Arnold & Pistilli, 2012).

In particular, the Course Signals project is very much in line with our study, as it aims to predict on-campus student performance on an individual basis and attempts early intervention for at-risk students. The system's ability to involve thousands of students across many disciplines and courses is outstanding. Nonetheless there are gaps in their project that can be filled by our study, which will be explained below.

The first difference is that Course Signals does not target any particular course or discipline to reach its goal. On the other hand, our study targets specific students, i.e. engineering students taking mathematics subjects, whose circumstance might be different from other disciplines. Secondly, Course Signals only use variables around academic performance and mainly students' online engagement with their courses. In contrast, our study will also use students' socio-demographic variables as well as their secondary school academic performance, in addition to their academic performance in the tertiary level. Lastly, in predicting student performance, Course Signals does not involve any variables regarding learning support, in particular mathematics support, which can be accessed by students in order to help them with their study. On the contrary, our study will focus on the mathematics support aspect, whether and how students engage with the available support, not only to pass but also to improve their understanding of mathematics subjects.

Learning analytics to improve retention with integration of mathematics support aspects

So far there has been no research on identifying at-risk students or predicting student performance using learning analytics and educational data mining that integrates the mathematics support aspects. Focusing on data mining techniques, Garcia & Mora (2011) mined data of over 6500 engineering students with 57 independent variables, but did not use mathematics support variables. On the other hand, Lee et al. (2008) incorporated one mathematics support variable, but only used a small data set (133 observations in one semester, 14 independent variables). Studies of such small size do not quite belong to learning analytics research as per the current definition.

In Australia, Loch & Elliott (2012) carried out a preliminary study to investigate the effectiveness of the current retention strategy in terms of mathematics support provision. This study extracted one cohort of 77 civil engineering students taking a single mathematics subject, analysing six variables in a descriptive manner. This study will extend the previous study, by analysing large data sets of several cohorts of engineering students with different majors taking different mathematics subjects.

By using data from a range of various sources instead of only academic performances, it is hoped that a new intervention strategy can be proposed to improve retention in these mathematics subjects, to the benefit of the faculty and the university. It will also add to current research on best practice of mathematics learning support and retention in mathematics education in general.

Proposed Methodology

In this study, we will learn from **past student data** to understand the **present challenge** we are facing in tackling the 'mathematical problem'. **Future performance** of current students will be predicted within the semester in order to prevent them from failing mathematics subjects, thereby improving retention.

This study will take a learning analytics approach to first-year student data sets at Swinburne University of Technology. These students have access to mathematics supports provided by Swinburne's Mathematics and Statistics Help (MASH) Centre. It will include four cohorts (2011 to 2014, i.e. past, present, and future data) of all first-year engineering students taking one of two core mathematics subjects with large enrolments, which is expected to be around 5,000 students in total.

Datasets will incorporate demographic, socio-economic and student academic performance variables, as well as data from the MASH Centre visits and access to 'Mathscasts', i.e. online support materials (Loch, Gill, & Croft, 2012) provided by the MASH Centre. Secondary school data such as VCE mathematics scores as well as university entry scores will also be included. Past student data will be evaluated and statistical models will be built on this data to predict the likelihood of student success in these subjects. These models will be verified with current student data, constantly updated on a regular basis to improve the models as the students are progressing through the semester.

At-risk students will be identified during the semester using predictive models that will employ all available variables. Certain behaviours of students regarding their engagement or non-engagement to the MASH Centre and its online support will also be taken into account, in order to configure triggers for an intervention strategy. These at-risk students will be referred to the faculty retention strategy coordinator, who will implement support services for these students as suggested from the data analysis. At this stage, ethics approval has just been granted and access to the data will commence shortly.

Statistical approach

Data mining techniques such as regression, decision trees, and support vector machines, as well as ensemble techniques will be applied to identify significant variables that contribute to student success as well as student failure in the mathematics subjects evaluated. Different predictive models, one for each past cohort and subject, will be built to identify which student is likely to be at risk of failing the subject based on all data available.

The models will be evaluated to assess their accuracy. In order to do this, data on each cohort/ subject regarding each model will be split in two groups, i.e. a training set to train the model, and a test set to test the model to determine the model's accuracy. These models will be regularly updated with the addition of current data, and evaluation of the models will be performed accordingly. Ensemble models will also be exercised as they usually yield more robust predictions compared to individual models (Delen, 2010). In this way, higher prediction accuracy is expected to identify at-risk students and predict student performance.

Expected outcomes

Expected outcomes of this study are:

4. "At-risk" criteria in first-year mathematics are well defined based on suggestions from the data analysis.
5. Applying these criteria on identification of at-risk students means that more students can be assisted to master and pass mathematics subjects, particularly if they do not avail themselves of mathematics support provision out of shyness and reluctance.
6. More proper and targeted intervention strategy to assist at-risk students from failing mathematics subjects.
7. Understanding students' behaviour in engaging with mathematics support provision, in order to configure improvement of the provision according to each student's needs.
8. Improved retention in first-year mathematics.

Conclusion

This proposed study will apply a learning analytics approach to identify how retention in first-year mathematics could be improved where mathematics support provision is available. It is an endeavour to combine two areas of research which previously seemed to have been investigated separately, i.e. learning analytics and mathematics support. This study will contribute to ways of improving retention in the light of available mathematics support provision.

References

Arnold, K. E. & Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). https://doi.org/10.1145/2330601.2330666

Croft, A. C., Harrison, M. C. & Robinson, C. L. (2009). Recruitment and retention of students-an integrated and holistic vision of mathematics support. *International Journal of Mathematical Education in Science and Technology*, *40*(1), 109–125. https://doi.org/10.1080/00207390802542395

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498–506. https://doi.org/10.1016/j.dss.2010.06.003

Garcia, E. P. I. & Mora, P. M. (2011). Model Prediction of Academic Performance for First Year Students. *Artificial Intelligence (MICAI), 2011 10th Mexican International Conference on* (pp. 169–174).

Gill, O., Mac an Bhaird, C. & Ní Fhloinn, E. (2010). The Origins, Development and Evaluation of Mathematics Support Services. *Irish Math. Soc. Bulletin*, *66*, 51–63. https://doi.org/10.33232/BIMS.0066.51.63

Lee, S., Harrison, M. C., Pell, G. & Robinson, C. L. (2008). Predicting performance of first year engineering students and the importance of assessment tools therein. *Engineering Education: Journal of the Higher Education Academy Engineering Subject Centre*, *3*(1), 44–51. https://doi.org/10.11120/ened.2008.03010044

Loch, B. & Elliott, D. (2012). Prototype of an intervention strategy with a focus on mathematics support for first year civil engineering students. *Profession of Engineering Education: Advancing Teaching, Research and Careers, The: 23rd Annual Conference of the Australasian Association for Engineering Education 2012* (pp. 565–572).

Loch, B., Gill, O. & Croft, T. (2012). Complementing mathematics support with online MathsCasts. *ANZIAM Journal*, *53*, C561–C575. https://doi.org/10.21914/anziamj.v53i0.4984

Mac an Bhaird, C., Morgan, T. & O'Shea, A. (2009). The impact of the mathematics support centre on the grades of first year students at the National University of Ireland Maynooth. *Teaching Mathematics and its Applications*, *28*(3), 117–122. https://doi.org/10.1093/teamat/hrp014

MacGillivray, H. (2009). Learning support and students studying mathematics and statistics. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 455–472.

Parsons, S. (2008). Overview of the provision of mathematics support to students in a University College. *MSOR Connections*, *8*(2), 29–32. https://doi.org/10.11120/msor.2008.08020029

Patel, C. & Little, J. (2006). Measuring maths study support. *Teaching Mathematics and its applications*, *25*(3), 131–138. https://doi.org/10.1093/teamat/hri031
Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005
Romero-Zaldivar, V.-A., Pardo, A., Burgos, D. & Delgado Kloos, C. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, *58*(4), 1058–1067.
Rylands, L. & Coady, C. (2009). Performance of students with weak mathematics in first-year mathematics and science. *International Journal of Mathematical Education in Science and Technology*, *40*(6), 741–753.
Siemens, G. & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). https://doi.org/10.1145/2330601.2330661
Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K. & Baker, R. S. J. d. (2011). Open Learning Analytics: an integrated & modularized platform- Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques. *SoLAR Concept Paper*. http://solaresearch.org/OpenLearningAnalytics.pdf
Wilson, T. M. & MacGillivray, H. L. (2007). Counting on the basics: mathematical skills among tertiary entrants. *International Journal of Mathematical Education in Science and Technology*, *38*(1), 19–41. https://doi.org/10.1080/00207390600819029

Author contact details:
Yasmin Erika Faridhan, yfaridhan@swin.edu.au
Birgit Loch, bloch@swin.edu.au
Lyndon Walker, lwalker@swin.edu.au

Please cite as: Faridhan, Y. E., Loch, B., &Walker, L. (2013). Improving retention in first-year mathematics using learning analytics. In H. Carter, M. Gosper and J. Hedberg (Eds.), *Electric Dreams. Proceedings ascilite 2013 Sydney*. (pp.278-282). https://doi.org/10.14742/apubs.2013.1506