Exploring the Use of Big Data Analytics for Improving Support to Students in Higher Education

Si Fan[1], Saurabh Garg[2] and SoonjaYeom[2]

[1]Faculty of Education, University of Tasmania, Australia
Si.Fan@utas.edu.au
[2]School of Engineering and ICT, University of Tasmania, Australia
{Saurabh.Garg,Soonja.Yeom}@utas.edu.au

Abstract. In the past two decades, with the globalisation of education, there has been a continuous increase in the diversity of students in Higher Education. This diversity form a basis for a culturally rich environment, although, the cultural and language differences and the diversity in teaching and learning styles also bring challenges. From a university's perspective, providing the maximum support to overcome these challenges and achieving maximised student engagement would be in its best interest. Recent advances in Big Data and increase in electronically available education data can help in achieving these aims. This paper reports the findings of a preliminary study which applies Big Data analysis methods to analyse education data gathered from learning management systems. The aims was to understand ways to improve student engagement and reduce student dropout. This paper documents the experience gained in this early exploration and preliminary analysis, and thereby provides background knowledge for reporting of data from the formal data collection stage which will be conducted at a later stage of research.

Keywords: Big Data, learning analytics, higher education, learning management system, Australian higher education.

1 Introduction

Big Data is relevant to all organizations, including higher education institutions who produce large data sets and wish to benefit from the interpretation of these data. Through analysis of information flow, Big Data analytics looks for hidden threads, trends and patterns (Matteson, 2013). Influential examples of Big Data applications include Google Flu Trends, which provides predictions and estimates of influenza activity for more than 25 countries, through aggregation of Google search queries (Pervaiz, Pervaiz, Rehman & Saif, 2012); and the research of Aslam et al. (2014), which reported on influenza-like illness rates through the collection and analysis of 159,802 tweets. While Big Data analytics is having a wide impact on public health and businesses' commercialization and marketing, its application remains limited in the field of education. Even in the few emerging studies on Big Data and learning analytics, the majority are carried out in America (e.g. Picciano, 2012) and Canada (e.g. Ellaway, Pusic, Galbraith & Cameron, 2014).

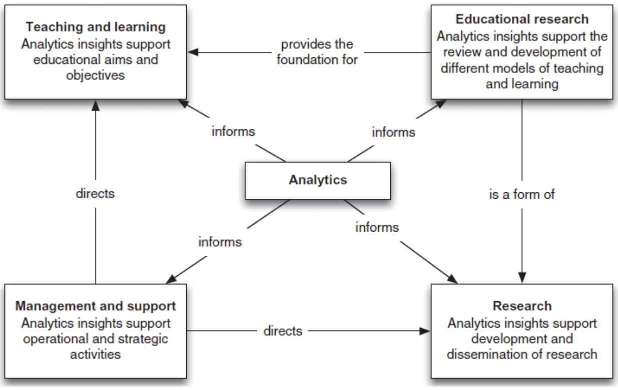

Figure 1. One model of educational data analytics (Ellaway et al., 2014)

In the field of higher education, an increase in information flow can be seen through the wide adoption of web-based learning systems by educational institutions (e.g. MOOCs). These learning systems often feature a wide array of materials, which build on a large resource of publicly available data. These data may reveal hidden patterns to "predict student outcomes such as dropping out, needing extra help, or being capable of more demanding assignments" (West, 2012, p. 2). Therefore, mining information from web-based learning systems can lead to insights regarding student performance and teacher pedagogical approaches (see Figure 1).

In this paper, we give the details of our project to harness this insight from electronic education data available through learning management systems using Big Data technologies. In particular, this research focuses on extracting insights in relation to student behaviour in learning a higher education course, and factors that can affect their learning. This paper introduces the preliminary findings from a small data set which we intend to further validate using large data sets from different courses at this university.

2 Big Data Analysis for Australian Higher Education

Big Data is a concept that emerged with the rapid growth of web-based technologies and computer and mobile devices. It indicates "large pools of data that can be captured, communicated, aggregated, stored, and analysed" (Manyika et al., 2011, p.4). Big Data in education is also gaining a growing attention from scholars (Eynon, 2013). Data that are produced by university learning systems can be seen as one type of Big Data. Learning analytics is a term researchers often refer to when discussing the use of Big Data in the field of education. Analysis results can be generated either from build-in analytics functionalities in learning management systems, or through analysis of data gathered using data mining techniques. Researches (Clow, 2013; Timms, 2015) summarize the differences between learning analytics and educational data mining that, educational data mining has a greater emphasis on technical challenges, in that educational data mining more often develops new methods and models for data analysis, and the later tends to apply existing models.

This research examines university lecturers' pedagogical approaches and student engagement in online learning, through the use of Big Data and learning analytics, using one Australian university as a starting point. The study collects archived data from the web-based learning system used at this university, which is based on the Desire2Learn (Brightspace) platform. The objectives of the project were set to examine correlations between:

- lecturers' pedagogical support and student engagement;
- types of teaching materials and resources provided and student engagement; and
- student evaluation and satisfaction and their level of engagement.

3 Methodology

The research was approved by the Social Science Human Research Ethics Committee (HREC) Tasmania Network (reference H0016064). This project is being conducted in the following three stages.

Stage 1 of the project has been completed as a literature review. Relevant studies that have been conducted in higher education contexts using learning analytics have been reviewed and summarized.

Stage 2 of the project is a pilot study which will be conducted on approximately eight units from the disciplines of computing and education. This stage of the project will involve the data extraction from this specific learning management system and data cleaning. Archived data of student online discussions, lecturers' feedback, news items, unit materials and resources, and student engagement statistics generated by this learning management system are extracted, for the three year period of 2013-2015. After the initial data collection, invalid data not serving the aims of the study are filtered and removed.

Stage 3 involves the analysis of the collected data. To understand the students' interaction with learning materials, different features are identified. These features include: the number of times a student accesses a particular learning module, dates when the student visited the learning module, dates when the student submitted assignments, time when s/he accesses the assignment details, his/her achievement in terms of marks, and how frequently the student posted to the discussion board. To understand the teaching team's interaction, features were collected, such as: when teaching materials were uploaded, teaching strategies, number of assignments collected, and dates assignments are recorded. Data mining techniques such as clustering and classification are utilised to detect different patterns in the data. Big Data frameworks such as Hadoop will be utilised to analyse the data using machine learning techniques. Pearson's correlation coefficients for the association between student engagement and participation and other three variables: lecturers' pedagogical support and interactions; types of teaching materials and resources; and student evaluation and satisfaction, will be calculated in ***R*** Foundation for Statistical Computing.

4 Preliminary Study and Results

We conducted the preliminary study using Excel's statistical tool to guide what features we need to extract from future larger data sets and which factors we should study to generate insights. The small scale study consisted of two master level ICT units: UNIT1 (Web Development) and UNIT2 (Data Management Technology) offered at the chosen university. The enrolled cohorts in the two units overlap to a large extent. One of these units (UNIT1) is a lower level unit and the other (UNIT2) is the higher level unit. For the study, we used a set of 20 students in UNIT2, consisting of low achieving (PP = Pass), medium achieving (CR-DN = Credit or Distinction), and high achieving students (DN-HD = Distinction or High Distinction).

The aim is to study the behaviour or study pattern of the students in these units, thus we chose three factors: 1) frequency of access to each learning component; 2) date of access; and 3) their achievements in the units. To summarise and further analyse the data, we evaluated the minimum number of accesses, maximum number of accesses, and average number of accesses, in all the learning contents.

Figure 2 and Figure 3 shows a summary of data. It can be seen clearly from the data, the study pattern of the students is similar in both units. This may indicate a high level of similarities in the patterns of teaching activities. At the same time, it can be noticed that on average the students in UNIT1 accessed learning materials less than those in UNIT2. This may indicate that more advanced units requires more effort from the students to understand the contents than lower level units. This can be evidenced further by the marks comparison between the two units. The data reveal that UNIT1 students got overall higher marks than in UNIT2. It is also interesting to notice that there are some high achieving students who enrolled in both units and achieved similar marks in both units. A difference can be seen in some medium achieving students who enrolled in both units. There is a one grade point difference in their achievement in the two units. This may be a reflection of the fundamental difference in the nature of these units. Both units have a portion on programming; however, UNIT2 has more difficult programming contents than UNIT1. Among low achieving students there is no clear pattern, however, two of these students achieved almost 20% more marks in UNIT1. This may indicate that these students have specific interest in a particular area as included in UNIT1, due to which they were able to achieve significantly higher results.

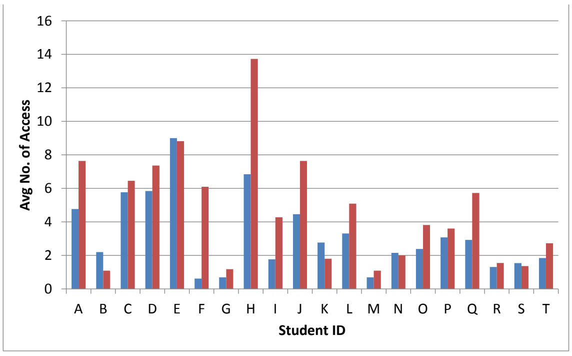

Figure 2. Comparison Based on Number of Access (Blue: UNIT 1; Red: UNIT2)

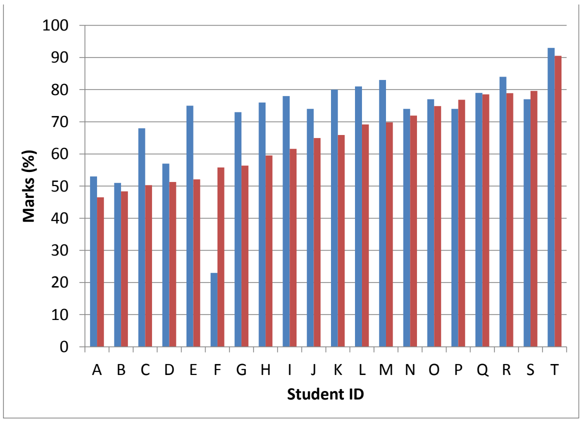

Figure 3. Comparison Based on Student's Achievement (Blue: UNIT 1; Red: UNIT2)

4. Concluding remarks

This paper presents the methodology and some preliminary analysis results from a small data set. Archived data were harvested from the learning management system used at one chosen Australian university. The aim was to study the innovative ways in which data management and analytics can benefit and contributes to Australian higher education. The preliminary study conducted using two master's level ICT units at one Australian university. The analysis validates that features, such as the number of visits to learning modules and their achievements in the units, are important indicators of users' interaction and learning in the unit. There are two key findings that emerged from this preliminary analysis. First, student interaction pattern remains the same from one unit to another. Second, students would access the teaching material more often if the unit is difficult or of higher level. These findings will be further tested and validated with larger data sets. Other features that are relevant to lecturers' pedagogical support and student engagement will also be added in later stage of the research. These features will be examined, using Big Data analytics techniques and skills that to date remain an unexplored area in Australian higher education. Starting with one university, the study will demonstrate relevance and impact for higher education nation wide, with the potential to contribute to online education at an international level.

Acknowledgement. This project is funded by the Research Enhancement Grant Scheme (REGS) at the University of Tasmania.

References

1. Aslam, A. A., Tsou, M. H., Spitzberg, B. H., An, L., Gawron, J. M., Gupta, D. K., ...& Lindsay, S. (2014). The reliability of tweets as a supplementary method of seasonal influenza surveillance. *Journal of medical Internet research*, *16*(11).
2. Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683-695.
3. Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical teacher, 36*(3), 216-222.
4. Eynon, R. (2013). The rise of Big Data: What does it mean for education, technology, and media research? *Learning, Media and Technology, 38*(3), 237-240.
5. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. U. S.: McKinsey Global Institute.
6. Matteson, S. (2013). Big data basic concepts and benefits explained. *Web blog*. Retrieved from *http://www. techrepublic. com/blog/big-data-analytics/*
7. Pervaiz, F., Pervaiz, M., Rehman, N. A., &Saif, U. (2012). FluBreaks: Early epidemic detection from Google flu trends. *Journal of Medical Internet Research, 14*(5).
8. Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20.
9. Timms, M. (2015). Big data in education: A guide for educators. *Centre for Strategic Education Occasional Paper* (No. 139), 1-10.
10. West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance Studies at Brookings*, 1-10.