APPLICATION OF BIG DATA FOR HIGHER EDUCATION: DATA MINING AND LEARNING ANALYTICS

Benson K.H. Hung

Department of Construction,
Hong Kong Institute of Vocational Education (Tsing Yi),
Vocational Training Council, HKSAR, China

bensonhung@vtc.edu.hk

ABSTRACT

Institutions of higher education are operating in an increasingly competitive and different environment. They are subject to competing demands nationally and globally, increasing pressure to respond to economic, political and social reforms and massive developments in new shape of higher education. In spite of the growing changes happening in higher education, the emergent field of research that uses big data in helping addressing contemporary challenges is often overlooked. As such, the purpose of this paper was to examine the evolving world of big data and learning analytics in the context of higher education. It examined the nature of these concepts, unlocked the value of the increasing data, and discussed the results of a research project (n=235) on big data, gauging with a purposive student survey of n=25. The results found that a similar behavioural pattern and meaningful trend were exhibited. This paper opened up new research areas that can be explored to enrich our understanding of the role of big data in future higher education.

KEYWORDS: *Big data, data-driven decision making, data mining, learning analytics, higher education.*

INTRODUCTION

Institutions of higher education are operating in an increasingly competitive and different environment to respond to economic, political and social reforms and massive curriculum developments. Although the emergent field of research that uses big data may help in addressing contemporary challenges, there is limited research into big data in higher education.

In view of the huge potential offered by big data, the purpose of this paper is to examine the evolving world of big data and learning analytics in the context of higher education. It examines the nature of these concepts, provides basic definitions, and discusses the implementation of a research project on big data.

This paper begins with an introduction about the big data in higher education and examines the relationships between data-driven decision-making, data mining and learning analytic. A research project on the use of big data to identify areas of weakness of students (n=235) from the Higher Diploma in Civil Engineering at the Hong Kong Institute of Vocational Education was conducted. Data was collected through a learning management system and was then proceeded to check the application of data analytics. Following a survey (n=25) on learning difficulties from the students' perspectives, this paper concludes how big data can be utilized to make pedagogical decisions and to inform teaching and learning. Based on the current research, it showed that the role of big data in understanding academic performance is significant and the students' learning experience can therefore be enhanced.

LITERATURE REVIEWS

Big data is a knowledge system that is already changing the objects of knowledge and social theory in many fields while also having the potential to transform management decision-making theory (Boyd & Crawford, 2012). Big data refers to any set of data that is so large or so complex that conventional applications are not adequate to process them (Manyika *et al*, 2011).

Data are available from student activities and data are also created by educational institutions which use applications to manage courses, classes and students. The amount of data made available in the above scenarios is so enormous that traditional processing techniques cannot be used to process them (Saptarshi, 2013). Big data make it possible to mine learning information for insights regarding student performance and learning approaches (Manyika *et al*, 2011).

Data mining and data analytic can provide immediate feedback to students and teachers about academic performance. This approach can analyse underlying patterns in order to predict student outcomes and it can identify pedagogic approaches that seem most effective with particular students (United States Department of Education Office, 2012).

Per the definition given by the Educause (2012), analytics is an overarching concept that that is defined as data-driven decision-making, which is the use of data analysis to inform courses of action. In particular, learning analytics is the use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals (Bach, 2010). By focusing on learning analytics, teachers can study learning in far more nuanced ways (Castro *et al*, 2007).

METHODOLOGY

The research methodology of the study consisted of a case study of students who studied the module CON4382 Highway Engineering from the Higher Diploma in Civil Engineering at the Hong Kong Institute of Vocational Education in the September 2017 – January 2018 period. A mixed method was employed to collect narrative data using ranking by a survey for students' opinions. On the other hand, a structured instrument was employed to collect numeric data of students' performance. Both results were compared to give in-depth description and identification of significant findings. Three stages were applied in this research to unlock the value of big data including: -

Step 1) Data Collection
Step 2) Data Analysis
Step 3) Visualisation and Application

The data was gathered and filtered for relevance and stored in a form that was useful. Once data had been rendered into a usable form, it had to be turned into actionable information. Finally, this information was used to help learners for progress and betterment or to enable teachers to make informed choices.

Data Collection

Two assessments were converted in a way that students' weaknesses can be identified by using big data through an online form. As a result, 79,775 entries were retrieved that comprehensively covered all the intended learning outcomes. Providing that assessments in such a way that students' weaknesses and common mistakes can be addressed by using big data, the programme learning outcomes can be obtained by analyzing which parts of programme learning outcomes were underperformed, extra attention was paid to focus on particular areas of weaknesses. The design was achieved through this step-by-step approach in order to retrieve useful data as shown in Figure 1.

Data Analysis

Afterwards, the retrieved entries were proceeded as corresponding indicators with respective to the intended learning outcomes (ILOs) and as the weakness locators (see Figure 2). By calculating the percentage of correct responses, analyses of which parts of ILOs were underperformed were identified in big data. Hence, teachers can identify sections of under-performed (Question Step 4 in as shown in Figure 2) and then he/she can reconfirm suitable pedagogical practices for more effective learning and teaching

Visualisation and Application

To visualize and to apply, the analysed data should make available to users in a form that is interpretable and ultimately is used to guide decision making. In this case, the decision-making was to support the learning needs of different classes/students in order to improve overall students' performance. Students' weaknesses were identified and there was a demand to propose remedial measures including class revision, provision of relevant support, examining real-life case studies and additional exercises etc. Specific learning and teaching strategies were used to tackle such weaknesses and to fill the learning gap. Moreover, regular and timely feedback were provided so that they knew what they were doing right and, more importantly, where they can improve.

RESULTS AND DISCUSSION

The big data told the areas of weakness to be tackled and it brought out a discussion with specific learning focuses. The example showed step 4 was identified as one of the ILOs underperformed. This result called for a reconfirmation of suitable pedagogical approaches for an effective teaching and learning.

To validate the results, a subsequent survey was conducted to compare the big data research with the learning difficulties expressed by students. In this survey, students were asked to rate the order of difficulties of all steps. Gauging both orders, it found that results exhibited a similar pattern, with the highest rated items the same for the first four. Table 1 has the details.

CONCLUSIONS

Big data, which is a growing area in education, incorporates the emergent research field of learning analytics (Long & Siemen, 2011). In higher education, technological developments have certainly served as catalysts for the move towards the growth of analytics (Wagner and Ice, 2012).

The results showed that learning analytics can identify the same priorities as a questionnaire of students' opinions which might be translated as they perform much the same as they say they performed. The results found that a similar behavioural pattern as expressed by students and meaningful trend of students' performance as exhibited by the big data.

This paper demonstrated a case study on how big data and learning analytics can be related in higher education. It explained the connection between the use of big data and learning performance for enhancing teaching effectiveness. The results showed that big data brought new opportunities for institutions of higher education in understanding students' performance that can be utilized in helping the students' learning and improving the overall learning experience. The study putted forward issues of consideration in choosing big data as an appropriate method for conducting studies to enhance learning and teaching. It also shed light on the basic designs and implementation for teaching enhancement via big data. With its exploration, future directions relevant to make important pedagogical decisions and to inform teaching and learning were made possible.

Question 1

Step 1 $AADT_d = AADT_b(1+r)^m =$ (Ans) $\times (1 + 0.85\%)^{\text{(Ans)}} =$ (Ans) veh./day

Step 2 $C_e = P_s \times P_v \times D_s \times AADT_d =$ (Ans) $\times$ (Ans) $\times$ (Ans) $\times AADT_d =$ (Ans) veh./day

Step 3 $C_v = 365 \times C_e \times \frac{(1+r)^n - 1}{r} =$ (Ans) million commercial vehicles

Step 4 $D_f = \frac{\text{(Ans)}}{\text{(Ans)}} \times (1 -$ (Ans) $) =$ (Ans) veh/hr/lane and $K_p =$ (Ans), thus

$C_d =$ (Ans) million commercial vehicles

$\Longrightarrow$ checked (Ans) and take $C_v =$ (Ans) million commercial veh. for design

Step 5 $\overline{CVDF} =$ (Ans), $W_f =$ (Ans) and $C_n =$ (Ans) MSA / $C_a =$ (Ans) MSA

Step 6 $E_s =$ (Ans) MPa

Step 7 Provide (Ans) mm capping layer and (Ans) mm granular sub-base.

Step 8 Modulus of subgrade reaction is $k =$ (Ans) MPa/mm

First reading from Design Chart (Ans) of (Ans) mm layer thickness,
if necessary interpolated with second reading from Design Chart (Ans) of (Ans) mm layer thickness, to give thickness (Ans) mm, rounded to (Ans) mm.
Provide bituminous layer thickness (Ans) mm as per Table 6.

Provide the following schedule:-

Bituminous wearing course material (20 mm)	-	40 mm thick
Bituminous base course material (37.5 mm)	-	65 mm thick
Bituminous road base material (37.5 mm)	-	(Ans) mm thick
Granular sub-base material	-	(Ans) mm thick
Capping layer	-	(Ans) mm thick
Mesh reinforcement	-	(Ans) kg/m^2
Main reinforcement	-	(Ans) mm^2/m
Cross reinforcement	-	(Ans) mm^2/m

Figure 1. A step-by-step approach to retrieve useful data through a learning management system^.

^ Students can input corresponding answers through an online form Individual answers refer to particular key concepts.

Time	Student N	Student N	Class	nswer	nswer	nswer	nswer	nswer	nswer	nswer	nswer	nswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer	ıswer
Answer				14340	6	)50-15:	0.65	33.5-34	0.55	300-185	31-32	1700	2	10	700-800	0.08	138-140	1	31-32	3.1	).9-0.94	89-91	0
2016	Ho King K\	1.41E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	CHAN DER	1.41E+08	EG524101	14340	6	15087	0.63	33.8	0.33	1823	31,59	1700	2	0.1	765	0.08	139.6	1	33.8	3.1	0.92	90	10
2016	So Hang W	1.41E+08	3PE01	14340	6	15087	65	33.74	0.55	1820	31.5	2800	2	10	1260	0.05	368	1	31.5	3.1	0.92	90	0
2016	Liu Ka Leo	1.41E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	Law Kin Kc	1.21E+08	3PE01	14340	6	15091	0.65	0.337	0.55	1820	31.53	1700	2	0.1	765	0.08	139.6	1	31.53	3.1	0.92	89.92	0
2016	TAM Ho Yi	1.41E+08	EG524101-	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	So Chi Wa	1.31E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.47	2800	2	10	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	CHAN Yin	1.31E+08	EG524101	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	139.6	31.54	3.1	0.92	90	0
2016	So Yi chun	1.31E+08	3PE01	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	1	31.54	3.1	0.92	90	97.77
2016	TUNG KIN	1.31E+08	3PE01	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	0	31.54	3.1	0.92	90	97.77
2016	Chan Ka C	1.31E+08	3PD01	14340	6	15087	0.65	33.74	0.55	1820	31.5	2800	2	10	1260	0.08	139.6	0	31.54	3.1	0.92	90	0
2016	CHAN DER	1.41E+08	EG524101	14340	6	15087	0.65	0.337	0.55	1820	31.49	2800	2	0.1	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	Sung Man	1.31E+08	EG524101	14340	6	15087	0.65	55	0.337	1820	31.49	2800	2	0.1	1260	0.08	230	1	31.49	3.1	0.92	89.8	0
2016	cheng ku	1.21E+08	3pe01	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	1	31.54	3.1	0.92	90	90
2016	lau wai hc	1.41E+08	EG524101	14340	6	15087	0.65	0.337	0.55	18.2	31.49	2800	2	0.1	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	YIM WAI L	1.31E+08	3PE01	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	1	31.54	3.1	0.92	90	97.77
2016	Tsang Chu	1.31E+08	EG524101	14340	6	15087	0.65	38	0.55	1823	31.54	1700	2	10	765	0.08	139.6	1	31.54	3.1	0.92	90	0
2016	Tang Wing	1.31E+08	EG314101-	13860	4	15087	0.65	33.8	0.55	1823	31.54	2800	2	0.1	1260	0.08	230	1	230	3.1	0.92	89.95	0
2016	Chow Cho	1.41E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.05	367.9	1	31.49	3.1	0.92	89.81	0
2016	Kwok Po V	1.41E+08	3PE01	14340	6	15087	0.65	33.8	0.55	1823	31.54	2800	2	10	1260	0.08	230	1	31.54	3.1	0.92	90	0
2016	Lai Wai Ch	1.31E+08	3PD02	14340	6	15087	0.65	33.8	0.55	1823	31.54	1700	2	0.1	765	0.08	139.6	0	31.54	3.1	0.92	90	100
2016	He Bingbi	1.41E+08	EG524101-	14340	6	15087	0.65	0.337	0.55	1820	31.49	2800	2	0.1	1260	0.05	36792	1	31.49	3.1	0.92	89.81	0
2016	He Bingbi	1.41E+08	EG524101-	14340	6	15087	0.65	0.337	0.55	1820	31.49	2800	2	0.1	1260	0.05	36792	1	31.49	3.1	0.92	89.81	0
2016	KWONG C	1.41E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.08	230	1	31.49	3.1	0.92	89.81	97.62
2016	KWONG C	1.41E+08	3PE01	14340	6	15087	0.65	33.74	0.55	1820	31.49	2800	2	10	1260	0.08	230	1	31.49	3.1	0.92	89.81	97.62
			Count of Correct Responses	**374**	**376**	**380**	**391**	**373**	**392**	**388**	**384**	**344**	**386**	**253**	**339**	**370**	**344**	**358**	**374**	**393**	**388**	**371**	**325**
			Percentage of Correct Responses	**94%**	**95%**	**96%**	**99%**	**94%**	**99%**	**98%**	**97%**	**88%**	**98%**	**64%**	**86%**	**94%**	**87%**	**94%**	**95%**	**99%**	**98%**	**94%**	**86%**
			Question Step	**1**	**1**	**1**	**2**	**2**	**2**	**2**	**3**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**5**	**5**	**5**	**5**

Figure 2. Indicators in percentage of correct responses with respective to the ILOs and as the weakness locators

	Average % of Correct Responses	The order of difficulties evaluated by big data	The order of difficulties rated by students
Step 1	95	5	7
Step 2	97	7	8
Step 3	97	7	6
Step 4	88	**3**	**2**
Step 5	95	5	5
Step 6	76	**1**	**4**
Step 7	92	**4**	**2**
Step 8	81	**2**	**1**

Table 1. The orders of difficulties evaluated by big data and rated by students showed how aggregated massive volumes of data can correlate students' views to identify a recurring behavioural pattern and meaningful trend.

References

Bach C. (2010). *Learning Analytics: Targeting Instruction, Curricula and Student Support.* Office of the Provost, Drexel University.

Barneveld, A., Arnold, K., and Campbell, J. (2012). Analytics in Higher Education: Establishing a Common Language. *Educause*.

Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Communication & Society*, *15*, 5, pp. 662–679.

Castro, F., Vellido, A., Nebot, A., and Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems. *Studies in Computational Intelligence*, Volume 62, pp. 183-221.

Long, P. & Siemen, G. (2011). Penetrating the fog: analytics in learning and education. *EDUCAUSE Review*, *46*, 5, pp. 30–40.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. *et al* (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.

Saptarshi, R. (2013). Big Data in Education. Gravity, *the Great Lakes Magazine*, pp. 8-10.

United States Department of Education Office. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics*. Educational Technology.

Wagner, E. & Ice, P. (2012). Data changes everything: delivering on the promise of learning analytics in higher education. *Educause Review*, *July/August*, pp. 33–42.