

Application of Big Data in Higher Education for Learning Analytics

Zhi-jun Pei, Lei Han and Jin-qing Gu
School of Electronic Engineering
Tianjin University of Technology and Education
Tianjin, China
peizj@tute.edu.cn

Abstract—Institutions of higher education are facing contemporary challenges. And big data can be used for the higher education practice. Apart from the introduction of the relevance of big data to higher education, analytical approaches such as learning analytics is provided for the analysis and exploitation of big data in higher education. It is helpful for big data to provide insights to support student learning needs. And the potential of learning analytics for quality improvement in learning process is also discussed in detail.

Keywords—big data; higher education; learning analytics; teaching; learning process

I. INTRODUCTION

Nowadays, big data is used as a term to describe the recent emergence of the vast amounts of different formats data from different sources. Data possessed in a system or a specific domain is considered as big data when the volume, variety and velocity are simultaneously high. Higher education is just one of domains that volume, variety and velocity coexist in the generated data. On a daily basis, large amounts of educational data are captured and generated from different sources and in different formats in the higher educational ecosystem. The educational data may be produced from student interaction with learning management systems and platforms. These data can also be generated from learning activities and courses information consisting a curriculum. Big data is related to the research field that uses data analysis to inform decisions. And it is currently being explored mostly in business, government and health care, due to the growing plethora of data collected and stored in these environments. There is only limited research into big data in higher education, despite growing interests in exploring and unlocking the value of the increasing data within higher education environment. Recently, big data analysis have shown promise in promoting different actions in higher education, such as administrative decision-making and organizational resource allocation, prevention of students at risk to fail by early identify them, learning activities and the curriculum [1]. This paper will contribute to the conceptual and theoretical understanding of big data in higher education and its application potential to learning analytics.

II. BIG DATA AND HIGHER EDUCATION

A. Big Data

Many organizations are currently using data to make better decisions about their strategic and operational directions. Most businesses today run on structured data, such as numbers and categories. However, there also exist a large amount of unstructured data, which come in a variety of formats such as text, video, audio, diagrams, images and combinations of any two or more formats. Recent developments in database technologies made it possible to collect and maintain these data generated in many forms and from multiple sources. In addition, there are analytical tools available to tap the hidden business value of those corporate data, referred to as big data.

The notion of big in the term big data does not reflect only data size but complexity. Big data refers to emergent suit of technologies that can process mass volumes of data of various types at faster speeds than ever before. Generally, big data has come to be identified by a number of fundamental characteristics [2]. For example, volume means that large amount of information is often challenging to store, process, transfer, analysis and present; velocity is related to increasing rate at which information flows within an organization; variety refers to data in diverse format both structured and unstructured; and most importantly, value implies that the data should be utilized to generate value of the insights, benefits and business processes.

There are three stages required to unlock the value of big data in any organization, which includes data collection, data analysis, visualization and application, as shown in figure 1. Once data have been rendered into a usable form through data collection, it has to be analyzed to generate actionable information. However, with the growing diversity in the nature of data, managing and analyzing diverse data set is becoming a very complex process. Visualization and application is the last stage where the analyzed

data is made available to users in a form that is interpretable and integrated into existing processes, and ultimately used to guide decision making.

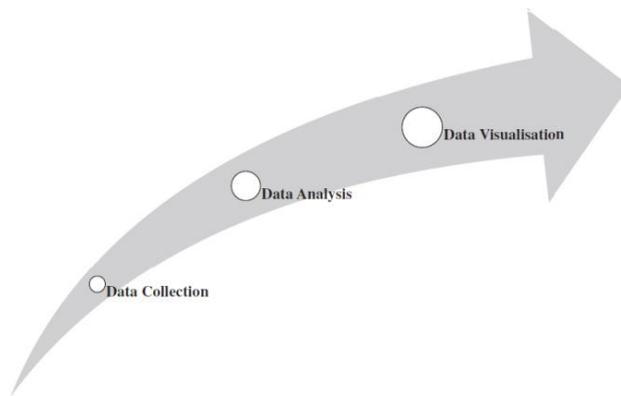


Fig. 1. Three essential stages of big data.

B. Big Data in Higher Education

New technologies continue to have a significant impact on the domain of higher education. Over the last decades, associated with developments in new technologies such as ubiquitous computing devices, flexible class room design and open online courses, a digital revolution is radically reshaping the mode and accessibility of learning and teaching. In addition, many institutions are embracing new class formats and technologies, which are designed to meet either evolving student needs or as mechanisms to reduce operational costs. The interaction with these technologies generates large amounts of data that range from an individual access log file to an institutional level activity, which can be easily captured and made available for subsequent analysis. Trillions of data are being collected and stored in various institutional databases. Many institutions of higher education are delivering learning online. Subsequently, there is a widespread availability of online repositories, educational digital libraries, and their associated tools. Data is stored in student information systems, student social media usage, learning management systems, student library usage, individual computers and administrative systems. In addition to growing abundance in data, the data come in different formats of audio, video, text and pictures. However, the educational systems are still not fully prepared to cope with and exploit them for continuous quality improvement purposes. Institutions of higher education are operating in an increasingly complex and competitive environment. In spite of the growing changes happening in the environment of higher education, the role of data in helping addressing contemporary challenges is often overlooked.

In fact, the variety of stored data could be explored with modern analytic techniques. Analytics are attractive approaches in education. The attention to analytics is driven by advances in computation, especially on various platforms. For instance, smartphones currently exceed the computational power of desktop computers and are more powerful than mainframe computers today, which supports researchers in analyzing large quantities of data. There is also a growing number of systems and tools that are intended to leverage big data and analytics. For instance, systems such as Apache Hadoop, Hortonworks, MapReduce and Tableau Software are designed to support the use of analytics tools without advanced technical knowledge. Furthermore, SAS and IBM SPSS address the substantial challenges of managing data at the scale of the Internet [3]. And currently, the best platforms for harnessing the power of big data are open and flexible.

So as learning technologies continue to penetrate all facets of higher education, vary kinds of generated data traces can be used to inform institutions of higher education to adapt better in response to changes happening within and outside their environments. Therefore, attention has been paid to different approaches such as big data that could be useful in investigating and exploiting educational data too. Big data have proven to be an extremely useful approach for learning analytics.

III. LEARNING ANALYTICS

The sources that big data in higher education are derived from are distinguished in different levels. And analytics may be classified into five levels, including course, department, institution, region and national or international. There are also other terms to define the different levels more specifically, as shown in figure 2. The nano-level indicates activities in a course; the micro-level points anentire course in an education programme; the meso-level includes many courses in a specific academic year; and the macro-level concerns study programmes in an educational institution. When the focus is on decision-making concerning achievements of specific learning outcomes, then all included actions are governed by learning analytics, which refers to operations at the microlevel and nanolevel. When the focus is on decision-making regarding procedures, management and matters of operational nature, then it is governed by academic analytics, which relates to the applications to the other two levels, macrolevel and mesolevel. The different levels of analytics overlap and complement each other. Despite the analytics actions in different levels, the generated data enter the same analytics loop which is defined in five steps:capture, report, predict, act, and refine.

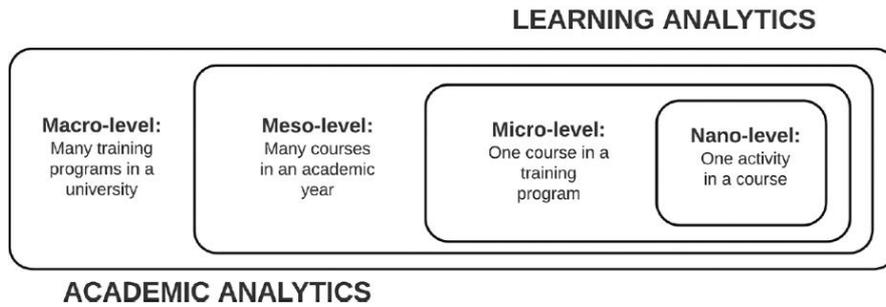


Fig. 2. Overlapping of analytics levels in higher education

A. Learning Analytics

In order to predict future performance and identifies potential issues related to academic program, research, teaching and learning, a conceptual framework is proposed to describe big data in higher education along four components: institutional analytics, information technology (IT) analytics, Academic analytics, and learning analytics[4], as shown in Figure 3. Institutional analytics refers to a variety of operational data that can be analyzed to help with effective decisions about making improvements at the institutional level. Information technology analytics aim at integrating data from a variety of systems, such as student information, learning management and alumni systems, as well as systems managing learning experiences outside the classroom. Academic analytics reflects the role of data analysis at an institutional level, which combines large data sets with statistical techniques and predictive modelling to improve decision making.

As for higher education institutions, learning is happening more and more within online environments and platforms. The educational data mining communities have already explored ways to track student behaviors. Those behavior specific data adds to an ever-growing repository of student-related information. Learning analytics is an emergent research area that intends to access and understand these data and adds a new dimension to the learning process. The term learning analytics is defined as the measurement, collection, analysis and reports of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs and affects actions and operations at the micro-level and nano-level. Through learning analytics, we can detect similarities in behaviors or detect anomalous patterns. It can function as a bridge between past and future operations by inserting data concerning past events into a learning analytics engine and analyze them to determine the probable future outcomes. It can thus synthesize big data in higher education and create a set of predictions to suggest different decision options revealing each time the implications of each decision option. Learning analytics can be further enhanced through visuals to amplify insight, increase understanding and impact decision making.

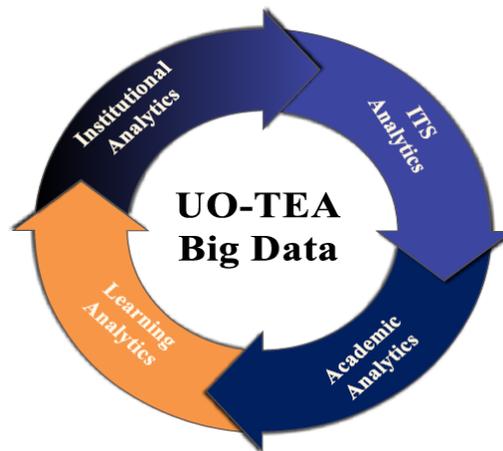


Fig. 3. Conceptual framework.

Learning analytics is undertaken more at the teaching and learning level of an institution and is largely concerned with improving learner success. Teachers, usually based on their experience, use their own gut feeling to translate student behavior and suspect if a student might drop out of a course or even abandon the studies. This can be proven to be either true or false. But there is low level of certainty in decisions that are based only on experience. Learning analytics has the capacity to add confidence to the decisions, where students at risk could be identified very early in order that the institution still had the time to react and take preventive actions.

B. *Quality Improvement of Learning Process*

Big data brings new opportunities and challenges for institutions of higher education, which incorporates the emergent research field of learning analytics. However, research in learning analytics has largely been limited to examining indicators of individual student and class performance. Big data presents the most dramatic framework in efficiently utilizing the vast array of data and ultimately shaping the future of higher education. Big data analytics could be applied to examine student entry on a course assessment, discussion board entries, blog entries or wiki activity, which could generate thousands of transactions per student per course. These data would be collected in real or near real time as it is transacted and then analyzed to suggest courses of action. It has been indicated that learning analytics are a foundational tool for informed change in education [5]. One practical use of learning analytics in the quality improvement of a course may be described as following.

In the preparation phase of a course, the instructors can use curriculum mapping tools to discover actual gaps precisely. They can thus recognize which learning objectives are not properly addressed by teaching or learning activities. Then recommendations for more proper and motivational teaching activities are needed to be included into their schedule. With the available analytics tools, they are able to analyze further the class and predict what it needs, such as student demographics, different learning approaches, and the group dynamics. This type of data is processed by a number of algorithms and predictive models that can develop the characteristics of the class. Visualization tools can be used for the following round to give alternative proposals for designing suitable activities fitting this particular class. The effects of each of the options can also be illustrated. The course director can control the activities and observe students related progress during the ongoing course. They can zoom in and out from the whole class to one working group or one individual student. They can judge the overall commitment and identify students at risk. In an extensively adopted platform, they can also compare particular indicators from other classes, or from a different department. They may even compare against data from related programs in other universities. The results and the produced experiences can be an important element of the quality development and academic research.

IV. SUMMARY

Institutions of higher education may benefit from an analytics of data sets from large volumes of student information. Big data and analytics in higher education can be used to alter the existing processes of administration, teaching, learning, and academic work, which is helpful for address contemporary challenges facing higher education. Learning analytics provide researchers with opportunities to carry out real time analysis of learning activities. By performing the analysis of student data, predictive models can be created to examine students at risk and provide appropriate intervention, which hence enable instructors to adapt their teaching.

ACKNOWLEDGMENT

This work is supported by Research Foundation of Tianjin University of Technology and Education under Grant No. KJY1312.

REFERENCES

- [1] S. Jones, "Technology review: the possibilities of learning analytics to improve learner-centered decision-making," *Community College Enterprise*, 2012, 18(1), pp. 89-92.
- [2] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges," *British Journal of Educational Technology*, 2015, 46(5), pp. 904-920.
- [3] A. A. Rupp, J. P. Leighton, R. S. Baker, et al, "Educational Data Mining and Learning Analytics," *Learning Analytics*, Springer New York, 2014, pp. 379-396.
- [4] B. K. Daniel, R. Butson, "Technology Enhanced Analytics (TEA) in Higher Education," *International Conference on Educational Technologies*, 2013, pp.89-96.
- [5] L. D. Roberts, J. A. Howell, Seaman K, et al, "Student Attitudes toward Learning Analytics in Higher Education: The Fitbit Version of the Learning World," *Frontiers in Psychology*, 2016, 7, pp.1-11.

