

THE IMPACT OF BIG DATA ANALYTICS ON THE GLOBALIZATION OF HIGHER EDUCATION

Yuanyuan Liu

School of Management, Guizhou University, Huaxi, Guiyang, Guizhou, 550025, China

Abstract

As the concept and application of big data and big data analytics are accepted by many sectors, the globalization of higher education is being greatly influence by them. This study aims at presenting concepts related to big data and big data analytics as well as their applications, and the impact of big data analytics on the globalization of higher education from strategic managerial, human capital, and data security and privacy perspectives. This study is to fill the research gap between big data analytics and the globalization of higher education. Although there are studies connecting big data analytics and higher education, yet this one specifically concerns the globalization of higher education.

Key words: big data, big data analytics, globalization, higher education

1. INTRODUCTION

We are living in a world of increasing data generated by the increased amount of information technological devices, users, and information. Nowadays, almost all sectors have been powered by Information Technology (IT) (Varma et al. 2016) and generating big Data (Jagadish 2015; Jin et al. 2015; Varma et al. 2016). The topic of big data has rapidly developed into such a hotspot that attracts great attention from academia, industry, to even governments around the world (Mayer-Schonberger & Cukier 2013; Cuzzocrea 2014; and Thomson, Lebiere & Bennati 2014). Schonberger & Cukier (2013) regards big data as a mark of the beginning of a major transformation, and Jin et al. (2015) comments that big data can be regarded the new petroleum that will power the future information economy, and it has already been in the offing.

Jin et al. (2015) provides details on significance of big data in the areas to national development, industrial upgrading, scientific research, emerging interdisciplinary research, helping people better perceive the present and better predict the future. Jin et al. 2015 also predicts that big data as the main support of the second economy, the competence under the second economy will longer be that of labor productivity but of knowledge productivity. The great value of big data is beyond all doubt.

Higher education, like many other sectors, has been hugely influenced by the wave of big data and analytics. Various types of internet-based teaching activities have been developed by higher education institutions, such as online courses, curriculum database, credits database, etc. However, influenced by the violent competition of globalization of higher education, the institutions begin to be burnt by incentive to make sense of data. To know more about the students choice, preferences, and other information, to lead to better decision making, and uncover the truth behind phenomenon, higher education institutions begin to apply big data and analytics, and integrate them into their institutional strategic management process.

2. BIG DATA AND BIG DATA ANALYTICS

2.1 Big data

Big data, as a relatively new term, describes the large volume of data, both structured and unstructured. Big data is to assume the information or database systems as the main storage facility to store large quantities of data (Picciano 2012). Siemens & Long (2011) defines big data as datasets whose sizes are beyond the ability of typical database software tools to capture, store, manage, and analyze. While Chen, Chiang & Storey (2012) adds that big data is so large from terabytes to exabytes and so complex from sensor to social media data so that advanced and unique technologies are required not

only for storage, management, analysis but also visualization. Huang et al. (2015) expanded the definition of big data into 5 Vs: (1) Volume (big volume of data), (2) Variety (of data types), (3) Velocity (high velocity of data generation and updating), (4) Veracity (of acquired data), and (5) Value (big value). While the first four Vs are concerned about data collection, preprocessing, transmission, and storage, the last V focuses on value of data. To create value, volumes of data should be processed to determine the relevance within the datasets (Varma et al. 2016). Thus it is not the amount of data which is important, but how to do with the data matters. Varma et al. 2016 states that data are unorganized facts which need to be processed so that they will become useful. Insights need to be extracted from the massive data to lead to better decisions. Big data, as the key research subject, consists of the basis of future data driven decision making techniques and applications (Kolomvatsos, Anagnostopoulos & Hadjiefthymiades 2015, and Batarseh & Latif 2016).

2.2 Big data analytics

The initial idea of 'big data' was used to describe the volume of information which had grown so large that the quantity being examined no longer fit into the memory that computers use for processing, and as a result engineers needed to revamp their tools to analyze the data (Mayer-Schonberger & Cukier 2013). Huge amount of data are produced and stored requiring the appropriate management mechanisms, so Big data analytics is created (Kolomvatsos, Anagnostopoulos & Hadjiefthymiades 2015). Big data analytics provides basic understanding of how to make decisions and comprehend reality (Mayer-Schonberger & Cukier 2013).

Big data and big data analytics seem to be two interchangeable items, which both may refer to the analytical tools used to manage the flood of data and turning the flood into a source of productive and useable information (Maltby 2011). And big data analytics is to examine data to draw conclusions, to present paths or course of action, and to incorporate operations (Picciano 2012). The hidden patterns and correlations of the large amount of data will be uncovered by statistical and analyzing methods (Huang et al. 2015).

2.3 Big data analytics pipeline

Diamantoulakis, Kapinas & Karagiannidis 2015 states the big data analytics pipeline from Gartner whitepaper, and the major steps in the analysis of big data are shown in the top half of the figure below (Jahadish 2015). Note that the possible feedback loops at all stages. The bottom half of the figure shows big data characteristics that make these steps challenging. The first step is to decide what data to acquire, and then to make the best of data which is even imperfect. Then next step is to decide in which manner to present the data suitably after extraction, cleaning, and integration with other data sources. The next is to establish models and analyze, while the last step is to do interpretation.

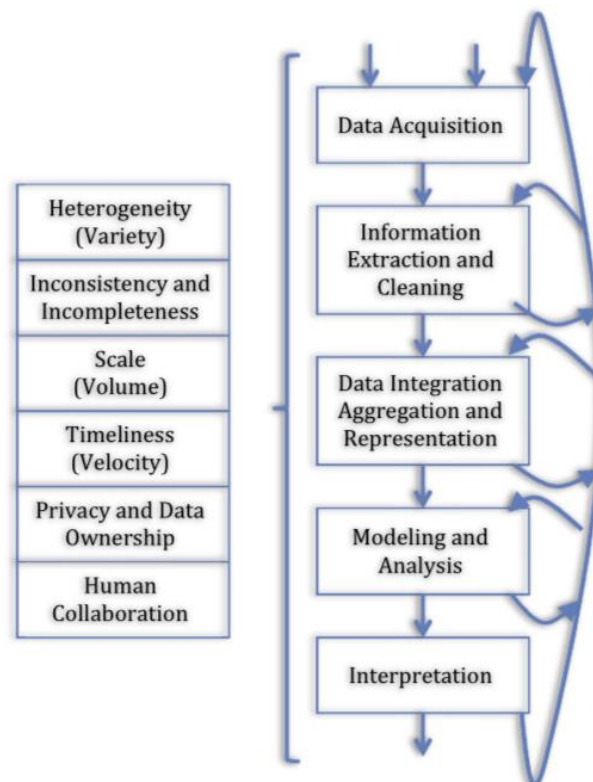


Fig 1. Big Data Analytics Pipeline (Jahadish 2015)

2.4 Techniques and technologies

Big data techniques are targeted to solve system-level problems which cannot be solved by conventional methods and technologies (Ang & Seng 2016), such as data mining, cluster analysis, crowd sourcing, text analytics (Maltby 2011). While technologies, such as EDWs (enterprise data warehouse), visualization products, Map-Reduce & Hadoop, NoSQL can be efficient way to store, visualize, organize and process the big data.

2.5 Applications

Big data has made a strong impact in almost every sector and industry today (Jin et al. 2015). Life sciences (Jagadish 2015, Huang 2015, and Batarseh & Latif 2016), engineering, natural sciences, art and humanities (Jagadish 2015) have adopted the application of big data and analytics. Companies adopt analytics in order to deal with critical applications like security, customers behavior, etc (Jin et al. 2015). Areas such as dynamic energy management in smart grids (Diamantoulakis, Kapinas & Karagiannidis 2015), open digital innovation (Brunswick, Bertino & Matei 2015), urban environment (Ang & Seng 2016), social network (Pääkkönen & Pakkala 2015, and Corea 2016), business intelligence (Fan, Lau & Zhao 2016), and genomic (Qin et al. 2015) are all examples of researches studying big data and analytics.

On the other hand, industry is more successful in applying big data and analytics as they need to process big data in real time and make better use of the data collected (Jin et al. 2015), for example, Alibaba disclosed in March 2014 about their data center has stored more than 100 PB of processed data, which amounts to 100 million high-resolution movies (Jin et al. 2015). During its annual shopping event named 'Double 11 Day', Alibaba pulled in CNY 9.3 billion in sales which corresponded to around 278 million orders. And to support this event, Alibaba developed its real-time data processing platform called Galaxy, which is able to handle 5 million transactions per second, and the total amount of data processed per day is about 2 PB (Jin et al. 2015).

During the wave of applying big data, Picciano 2012 states that big data concepts and analytics have been applied to higher education from both administrative and institutional perspective, including recruitment and admission processing, financial planning, donor tracking, and student performance monitoring.

However, there is still a gap between big data analytics and the globalization of higher education, which will be filled by the article.

3. GLOBALIZATION OF HIGHER EDUCATION AND THE IMPACT OF BIG DATA ANALYTICS ON IT

3.1 Globalization of higher education

Globalization is made both possible and necessary by the development of two transforming technologies - transportation and communication (Boyd & Mitchell 2005). Meanwhile globalization is also considered as the reality shaped by an increasingly integrated world economy, new information technology (Johnstone, d'Ambrosio & Yakoboski 2010). And higher education institutions are facing challenges and opportunities of cultivating world citizens (Johnstone, d'Ambrosio & Yakoboski 2010).

Since 1960s, the term 'globalization' and its impact have been developed and discussed. Mitchell & Nielsen 2012 discussed that the term is used either to characterize international spatial awareness or to highlight a transformation in the process of interaction among individuals and groups, the world as a single place, a global village but the other dimension for interaction, a networking society. It is also discussed that from the social process transformation perspective, globalization is having a transformative effect on the core functions of institutions of higher education. As a result, they are developing a consumerist mentality which transforms higher education into product in an open market. Thus higher education institutions began their competition for monetary purposes and human resources globally to benefit their institutions, and turn knowledge into a commodity, students into customers. Neo-liberal is the mechanism of competition for resources in the context of globalization of higher education, which allowing a new order in the world of higher education (Mitchell & Nielsen 2012).

The global race for talents, branching out of universities, the wanted 'world class' universities, the world wide university rankings, the global expansion of for-profit institutions and the free-trade in minds are the major themes for higher education globalization (Wildavsky 2010).

3.2 The impact of big data analytics to globalization of higher education

Big data analytics will foster higher education institutions to adopt top-down management model and integrate data analytics into strategic planning during their process of globalization. As higher education institutions will be able to apply big data concepts and big data analytics to discover the situation and preference of global students, to uncover the correlations between factors, to explain reasons behind the phenomenon, higher education institutions may make better decision of targeting. On the other hand, higher education institutions should summarize and discover their strength, weakness, opportunities and threats by applying data analytics and thus strategically create their absolute competitive advantages, and integrate them into strategic management of the institutions. The most powerful equipment that data analytics will bring to higher education institutions is the transparent information all around the world so that they can better understand their positions in the world and the competitors' as well. Once the institutions collected information needed, they should adopt a top-down management model to strategically make plans for development, positioning and targeting, instead of a bottom-up approach although it may solve some niche problems, the isolated solutions often cannot be put together into a complete solution (Jin et al. 2015).

Big data analytics will foster higher education institutions to cultivate and store human capital of cross-discipline and multi-discipline. Although cross-discipline and multi-discipline talents cultivation had been set as one of the major aims of higher education institution, big data analytics and its

application for globalization of higher education asks for human resources who grasp knowledge of big data technique and technologies, computer science, software tools, marketing and statistical skill, and education as well. In order to well apply big data and analytics and make it effectively and efficiently work for higher education institutions, human resource will become the first and most crucial problem because higher education institutions need talents who have strong skills in interpreting factors and analysis, taking appropriate managerial actions after discovering the facts, and having adequate computer knowledge to support their ideas. During the process higher education globalization and application of big data analytics, human resources has become the biggest obstacle. However, because of the particular function of higher education institutions, they can cultivate talents by themselves, and those talents will become the institutions' human capital. For one thing, they can satisfy their own needs, and for the other, those talents are also needed by other organizations, and the corresponding majors will become widely accepted and highly required by students and organizations.

Big data analytics will foster higher education to pay greater attention to protect data security and privacy. As the world is becoming more and more transparent because of big data and analytics, both students and staff will be more worried about their data security and privacy. Those who adopt big data and analytics are collecting and storing data, while they are using and analyzing data. These are the concerns from students and staff. Thus higher education institutions should protect data and use data illegally.

4. DISCUSSIONS

Higher education has become a violent arena during its process of globalization. Thus understanding needs of the students, the organizations, and the society is becoming more and more important. Establishing the higher education institution which is able to meet the needs has become the crucial part of higher education institutions' survival during its process of globalization. Big data and analytics will be able to help higher education institutions to better develop and plan themselves from the 'customer' side and thus to lead to better decisions. Those higher education institutions who adopt big data and analytics will win the competition of globalization.

Meanwhile, there is a wide range of applying big data and analytics to higher education and its globalization. From promoting the higher education institutions by social media to collect data from the promotion, the institution may get to know which part of the world the students interested in their program are coming from by checking their clicks and absolute clicks per day per visit, and the institution may also know if their promotion program is effective by checking how many students enrolled are affected by the promotion and how much their decisions choosing the institution are influenced by the promotion. The institution may also manage students online learning process by checking their clicks of the webpage, their stay at the webpage, their post and comments on the webpage, etc. Thus manage those at-risk students to better perform at the beginning not simply failing them after the exam.

Batarseh & Latif 2016 states the data analytics lifecycle and its six steps including identifying and formulating the problem, preparing the data, data exploration, data transformation and collection, statistical modeling, and evaluating and monitoring the results. Besides, any data analytical project requires four main human roles: the data manager, the system manager, the analyst and the data miner.

5. CONCLUSION

There are also challenges and barriers that big data analytics will bring to the globalization of higher education, such as the privacy of students and staff, the security of data, legal right and legal use of data, and human resources and capital to support the strategic management of the higher education institutions during the process of globalization. On the other hand, the data complexity, computational complexity, and system complexity are challenges of big data. The appropriate management of huge amount of both structured and unstructured data is also a key issue for higher education institutions.

ACKNOWLEDGEMENTS

This work is one of Yuanyuan Liu's research results of the project 'The Research of Higher Education Globalization of Guizhou Province' of Guizhou Provincial Education Department, partly supported by Karst Region Development Strategy Research Center, and the project number is JD2014028.

REFERENCES

- A. Cuzzocrea 2014. Privacy and security of big data: current challenges and future research perspectives, in: Proceedings of the First International Workshop on Privacy and Security of Big Data PSBD'14, 2014.
- Anthony G. Picciano 2012. 'The revolution of big data and learning analytics in American higher education', *Journal of Asynchronous Learning Networks*, Vol. 16: 3.
- Ben Wildavsky 2010. *The great brain race: How global universities are reshaping the world*, Princeton University Press, Princeton & Oxford, 2010, xiii, pp. 1-240.
- D. Bruce Johnstone, Madeleine B. d' Ambrosio & Paul J. Yakoboski 2010. *Higher Education in a Global Society*, Edward Elgar Pub, 2010.
- Dylan Maltby 2011. 'Big Data Analytics' ASIST 2011.
- Feras A. Batarseh & Eyad Abdel Latif 2016. 'Assessing the quality of service using big data analytics with application to healthcare', *Big Data Research* 4 (2016) 13-24, ELSEVIER.
- Francesco Corea 2016. 'Can twitter proxy the investors' sentiment? The case for the technology sector', *Big Data Research* 4 (2016) 70-74, ELSEVIER.
- G. Siemens & P. Long 2011. 'Penetrating the fog: analytics in hearing and education', *EDUCAUSE Review*, 46 (5), 30-32.
- H. Chen, R.H.L. Chiang & V. C. Storey 2012. 'Business intelligence and analytics: from big data to big impact', *MIS Quarterly* 36(4): 1165-1188.
- H.V. Jahadish 2015. 'Big data and science: myths and reality'. *Big Data Research* 2 (2015) 49-52, ELSEVIER.
- Kostas Kolomvatsos, Christos Anagnostopoulos & Stathes Hadjiefthymisdes 2015. 'An Efficient time optimized scheme for progressive analytics in big data', *Big Data Research* 2 (2015) 155-165, ELSEVIER.
- Li-Minn Ang & Kah Phooi Seng 2016. 'Big sensor data applications in urban environments'. *Big Data Research* 4 (2016) 1-12, ELSEVIER.
- Panagiotis D. Diamantoulakis, Vasileios M. Kapinas & George K. Karagiannidis 2015. 'Big data analytics for dynamic energy management in smart grids', *Big Data Research* 2 (2015) 94-101, ELSEVIER.
- P. China Venkanna Varma, K. Venkata Kalyan Chakravarthy, V. Valli Kumari & S. Viswanadha Raju 2016. 'Analysis of a network IO bottleneck in big data environments based on Docker Containers', *Big Data Research* 3 (2016) 24-28, ELSEVIER.
- Pekka Pääkkönen & Daniel Pakkala 2015. 'Reference architecture and classification of technologies, products and services for big data systems', *Big Data Research* 2 (2015) 166-186, ELSEVIER.
- R. Thomson, C. Lebiere & S. Bennati 2014. Human, model and machine: a complementary approach to big data, in: Proceedings of the 2014 Workshop on Human Centered Big Data Research, HCBDR'14 2014.
- Sabine Brunswicker, Elisa Bertino & Sorin Matei 2015. 'Big data for open digital innovation - a research roadmap', *Big Data Research* 2 (2015) 53-58, ELSEVIER.

Shaokun Fan, Raymond Y.K. Lau & J. Leon Zhao 2015. 'Demystifying big data analytics for business intelligence through the lens of marketing mix', *Big Data Research 2* (2015) 28-32, ELSEVIER.

Tao Huang, Liang Lan, Xuexian Fang, Deng An, Junxia Min & Fudi Wang 2015. 'Promises and challenges of big data computing in health sciences', *Big Data Research 2* (2015) 2-11, ELSEVIER.

V. Mayer-Schonberger & K. Cukier 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.

Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng & Yuanzhao Wang 2015. 'Significance and challenges of big data research', *Big Data Research 2* (2015) 59-64, ELSEVIER.

Yiming Qin, Hari Krishna Yalamanchili, Jing Qin, Bin Yan & Junwen Wang 2015. 'The current status and challenges in computational analysis of genomic big data', *Big Data Research 2* (2015) 12-18, ELSEVIER.