

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 91 (2016) 855 – 861

Procedia
Computer Science

Information Technology and Quantitative Management (ITQM 2016)

A Study on the Application of Big Data to the Korean College Education System

Yeon Hee Kim and Jin-Ho Ahn*

Hoseo University, 79-20 Hoseo-ro, Asan and 336-795, Korea

Abstract

Big data are referred to bulk data which cannot be collected, saved, and analyzed with the traditional data analysis tools. The field of learning analysis, which has consistently appeared on the Horizon Report from New Media Consortium for the recent years, is receiving fresh attention with the proliferation of big data.

The purpose of this study is to examine the environment for the learning analytics, a branch field of big data, to be applied to the Korean education curricular and its possibility and find out how to improve them. First, as an application of a new technology involves side effects in most cases, it is desirable that potential problems be considered from the beginning and negative effects be minimized. Second, the learning analytics begins with securing sufficient data, and a data exchange system should be established for setting up the data ecology. Third, as big data are applied to public data and the corporate business areas and frequently mentioned in the media, the expectation for their potential growth is reaching its peak.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of ITQM 2016

Keywords: Big-Data, Learning analytics, Horizon Report, Predictive Analytics

1. Introduction

For the past 10 years, Korea, the world's highest Internet access rate holder, has witnessed its many fields digitalized with an explosive increase in the amount of the digital data, and it is expected that by 2020 the data amount around the world will be expanded to 35 ZB, which is beyond the capacity of the existing data

* Corresponding author. Tel.: +82-41-540-5667; fax: +82-41-540-5668.

E-mail address: jhahn@hoseo.edu.

technologies. Various values are being created using these digital data, and in particular, creating new information and values using 'big data' is emerging as an important factor for sharpening the competitiveness of the nation as well as of companies. It can be said that Government 3.0 as well is part of the Korean government's effort to preoccupy the future information values. The concept of 'big data' has been one of the most popular keywords since it was first used by McKinsey and Company, a consulting firm, in 2011. Big data are referred to bulk data which cannot be collected, saved, and analyzed with the traditional data analysis tools.[1] The field of learning analysis, which has consistently appeared on the Horizon Report from New Media Consortium for the recent years, is receiving fresh attention with the proliferation of big data. Against this backdrop, various attempts to use big data in education are being made around the world. For instance, many schools in the U.S. are making data-based decisions, where individual students' education data are utilized to improve their academic achievements. The result of the analysis of the data on the students' academic achievements at a public school in Atlanta, the U.S. showed that mathematics (algebra 1) was the most important subject that had the greatest influence on the rate of the graduation or their academic achievements and that those who did well in this subject also did well in the subject of the creative writing.[2] This learning analytics, which began to be studied also in the field of big data, can present the future career, the area of interest, and the education curricular for a specific student, and may provide an individualized learning given the fact that it provides individually customized learning methods and information. The purpose of this study is to examine the environment for the learning analytics, a branch field of big data, to be applied to the Korean education curricular and its possibility and find out how to improve them.

2. Learning Analytics

While with the rapid expansion of the learning activities through the online expansion thanks to the ICT (Information and communication Technology), currently used in Korean higher education, meaningful online data are consistently increasing, the number of the cases of using them are small. However, many countries and colleges around the world are building the infrastructures for big data to be analyzed, studying various analysis methods, and producing meaningful outcomes by using the data accumulated in and outside of schools. The data mining technique, which is the most important technology in utilizing big data in the field of education, is used to extract meaningful information from the vast amount of data.[3] For instance, an analysis of the weblogs using the mining technique helps establish customized advertising strategies by determining what tastes individual customers have and what products they are interested in. Likewise, the mining technique can be used in various ways in the field of education. In other words, customized education services can be provided by mining meaningful information suitable for each student. It is also important that as new data for each student are being constantly saved, they should be analyzed and the result should be presented in real time.[4] Siemens & Long, who said that the learning analytics was the conceptualization of collecting, processing, and analyzing these data, defined it as "measuring, collecting, analyzing, and reporting the data for learners and their contexts with the purpose of understanding and optimization of learning and the environment in which it occurs." [5] Chatti, et al. proposed as the possible purposes of the learning analytics monitoring analysis, prediction, intervention, tutoring/monitoring evaluation, feedback, individualization, recommendation, and retrospection.[6] In other words, the learning analytics is measuring, collecting, analyzing, and applying all the related data to understand and improve all the things related to learning. According to Siemens' UNESCO Policy Brief, major application plans for the learning analytics include learning analytics dashboard, prediction analysis, adaptive learning analysis, social network analysis, discourse analysis, and evaluation using the ICT.[7]

2.1. Learning Analytics Dashboards

Most of the educational institutions provide the one-way dashboard related to learning through the learning management system. Recently, one-way learning information began to be added to the graphs and tables regarding the log information analysis and others of the users. However, the log data arising in the platforms and software for learning are very technical formats of data, which are the types of data not available to the general users. Usually, log data were used in performing analysis works with the software for limited data mining but were not the information provided to the final users like teachers and learners after being processed in the visualized formats such as tables, graphs, etc. The learning analytics dashboard has the function of visualizing the log data to make it easy for individual users or groups to understand them.[8]

In addition, data can be collected and expressed from various sources; for example, individual students can find out their relative standings in comparison with those of other students by collecting the information of their own log data, grades, forum activities, attendances at seminars, use of the library, etc.

2.2. Predictive Analytics

Google's prediction technology was possible to made by the development of Big Data, Learning analytics is possible to predict such as Prediction of the student's academic performance through analysis of the data, the grades of specific learning process, Prediction of learning method through learning style analysis. For example, It can be alert to learners of the dangers step closer through predictive analytics and could guide the activities trajectory for entering into the average or excellent step.

Now, The most reliable way to determine whether or not to pass through the overall evaluation was to determine how to take the test at the time of starting the study. However, this method has limitations that do not reflect the status information and learning. Therefore, it can be predicted that more effective analysis method to determine based on the real-time learning status and level.[9]

2.3. Adaptive Analytics

This technology has for its object to provide a most appropriate level and attention to the subject information. It provides to necessary skills and information (the number of days that the person is accepted levels) for an understanding of the particular subject, and it provides the following steps as consider learning process of existing. These adaptive learning analytics will provide the student-directed learning will enhance the results of the study.

In which case, determining the difficulty level in consideration of the state of the learner and It becomes possible to skip the learning on the part. Therefore, it is necessary to implement a real-time status of the learner and the learner can decide adaptive technology.

2.4. Social Network Analytics

Social network analysis could see the human relationship and could understand how to develop and maintain the human relationship. People form a variety of "relationship" from a wide range of an indirect relationship to the direct interaction reflecting the important relationships. Interaction of the student is also similar. It is to be formed in the most school, takes place in the LMS, or SNS, such as e-mail and discussion, information exchange. Social network analysis is to maximize the effectiveness of learning as the collect and analyze these types of collaborative learning. The conduct of student's cooperative learning promote to understanding of the learning content, the fact has been reported that a student degree of social network connections with fellow students is a correlation with the overall learning outcomes of students.[9]

2.5. Discourse Analytics

Analysis of log data covered by the learning analytics are the most easy to analyze data such as whether students participated in the login records or information found in any discussion and what paths. For a more accurate analysis of the information should be more analysis such as contents, comments or thinking, text of students. That is, the text is the opinion of the individual, writing assignments, using sentence, email, SNS, etc., this technique is referred to as text mining based on the natural language processing technology. The text mining is the main topics that were long research in the field of pattern recognition and parsing technology, it is an effort to apply the learning analysis is processing. In distance education agencies Milton Keynes UK universities conducted a study to identify the informative conference sessions to provide useful information by analyze the contents of an online chat conference participants. Researchers sure how many participants are involved in exploratory studies to analyze the conversations in an online chat text chat, after a lot of analyzing to try how much exploratory conversations per person, and compares these figures to each session and to date. As a result, This chat discourse analysis technique can be considered necessary research in order to analyze the quality and performance of online education In a large online classes such as measuring and assessing the lessons MOOC participants, and to assess the degree of interaction between participants.

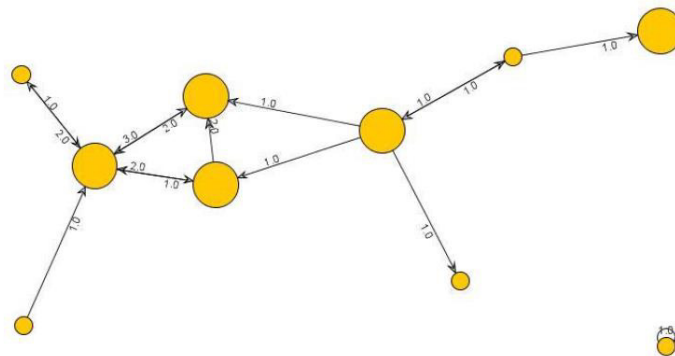


Fig. 1. Discourse Analytics

3. Contents of the learning analytics applicable in the college education curricula

Collecting data for the learning analytics is referred to the entire process of collecting and analyzing all the data caused by students' learning activities, in which teaching and school activities related to their learning activities should be included. These diverse types of data include atypical data, numerical data, text data, etc. most of which are based on data, and understanding of the data to be collected is needed. An international standardization organization, IMG Global Learning Consortium (hereafter called IMG Global) classified the data that can be collected and analyzed in the field of education into 5 types as seen in the list below and is now actively pushing ahead with the clarification of the data for learning analytics.

- Learning Content Data
- Learning Activity Data
- Operational Data
- Career Data
- Profile Data

On this wise, Learning analytics noted that the data of the five type’s data is Leverage content and data are generated in connection with activities. IMS Global is presented based on the data matrix for learning analytics as <Table 1> with the various stakeholders. Learning Activity Metrics is the data group generated during the learning activity (or learning), Foundational Metrics listed the data group previously used (that were mostly recorded in the database and management) in order to online service or learning outcomes measure.

Table 1. Learning analytics Data Matrix of IMS Global

<i>Learning Activity Metrics</i>					
Item	content	Item	content	Item	content
Reading	annotations	Quiz	scores	Homework	scores
	page/block use		attempts		attempts
	media use		remediation		remediation
	lookups		assoc refs		assoc refs
Lectures	frameset use	Pro Ject	deliverables	media	media type
	scrub marks		structure		frameset use
	view time		milestone pref		scrub marks
	weblink refs		group profile		view time
			patterns		usage context
Tutoring	topics	Research	sreaches	Assessment	scores
	assoc. context		patterns		patterns(item)
	frequency		citations		time utilization
	feedback		topics		attempts
					completion
Collaboration	connections	Annotation	highlights	Gaming	progress
	assoc context		notes		cognition
	message profile		marks		attempts
	frequency		tags		hints
			attachments		collaboration
Social	connections	Messaging	assoc context	Scheduling	assoc context
	assoc context		outbound pool		event patterns
	message profile		inbound pool		event profile
	frequency		attachment		time utilization

Discussions	post mark				
	frequency				
	participation				
	collaboration				
<i>Foundational Metrics</i>					
Context	institution	Engagement	activity usage	Performance	grades
	course/section		time on task		progress
	learner profile		session time		rubrics - course goals - topic objectives - qualitative evaluation - quantitative scores
	course context		last access		
	path/sequence		activity affinity		
	usage context		content affinity		
			task patterns		patterns
	correlation	correlations			

However, the data on this line is different from the present situation as the ultimate goal in the study analysis. That is, on the current complex studied form of the on and off-line is not accurate learning analytics to analysis content of the on-line. Therefore, there is a need to further analysis to learning activity of off-line as follows.

- Reading major books or those of interest at the library
- Reading books or talking with friends on the school bus
- Having interviews with professors
- Having talks with seniors
- Having discussions with friends: at coffee shops or snack bars, or on campus, etc.
- Having talks at the restaurant in front of the school

Like big data, learning analytics needs ecological standardization system consisting of generating, collecting, saving, processing, analyzing, and visualizing data which are normalized with the standardization process based on the developed data eco-system. So far, the overall process from generating to collecting to saving data of the ecological standardization system has been performed through IMS Global, private organizations, and the public standardization organizations such as ISO/IEC JTC 1/SC 36 Information technology for learning, education and training. IMS Global presents the following objects to be standardized through learning data measurement framework called IMS Caliper.

- IMS learning data matrix profile
- IMS learning sensor API and events
- IMS standardization (LTI/LIS/QTI1) leverage and its expansion

IMS Caliper is a project to define the information model for collecting in the standardized formats learning activity data which are generated in various learning environments and to develop a standardized API to collect data. While big data are open to the issues of the low analytic accuracy as they collect and analyze the raw data as they are, IMS Caliper can increase the efficiency of the analysis by making it possible to collect qualified data through the standardization. So far, IMS Caliper has not had the process of analysis and visualization

which is performed following the stage of collecting data as the object to be standardized, but Ministry of Science, ICT and Future Planning's learning analytics standardization project currently performed by Korea Education Research Information Service (KERIS) has the entire process from collecting to processing to analyzing to visualizing data as the object to be standardized while also utilizing the standardized development strategies based on the learning analytics reference model composed of open source software.

4. Conclusion

The learning analytics is a very appealing method as it can drastically improve the existing education environment and increase the efficiency of college management and teaching/learning. However, a successful realization of the learning analytics system requires the following three issues to be well resolved. First, as an application of a new technology involves side effects in most cases, it is desirable that potential problems be considered from the beginning and negative effects be minimized. At current issue are the 'neutrality of data', 'possible changes in the evaluation system due to the learning analytics', 'ethics required from the people accessing the data', etc. Second, the learning analytics begins with securing sufficient data, and a data exchange system should be established for setting up the data ecology. However, with no specific standardized learning analytics determined yet, a severe competition are expected to arise between organizations and between companies for preoccupying and standardizing the platforms in the future. If the selection of the learning analytics system is too heavily dependent on one specific platform, the resultant risks should also be considered. Third, as big data are applied to public data and the corporate business areas and frequently mentioned in the media, the expectation for their potential growth is reaching its peak. Furthermore, the report "Hype Cycle for Emerging Technologies" recently published by Gartner put big data on the list of the technologies whose expectations reached their peaks. Those in data science should keep in mind that the key to success lies in the big volume of big data but in how to draw out some insight even from its small volume and convert it into some value. To achieve a successful realization of the learning analytics system. we should be able to minimize the negative factors that the above-mentioned problems will bring about and flexibly respond to the needs of the rapidly changing education environment. Specifically speaking, using an incremental model which is based on the feedback of data is thought to be the most appropriate for developing a software process. In addition, we should establish short- and long-term strategies for using the learning analytics and make efforts with consistent interests for undertaking research and development of the learning analytics-related theories and system.

References

- [1] Manyika, J. Big Data: The Next Frontier for Innovation, Competition, and Productivity. Executive Summary, Mickinsey Global Institute, 2011.
- [2] Lavelle et al., MIT Sloan Management Review, 2011.
- [3] Linoff, G. S. and M. J. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Wiley Computer Publishing, 2011.
- [4] kwon yongcock, *Data Analytics in Education : Current and Future Directions*, journal of Intelligent Information, 2013.06.
- [5] Siemens, G, and Long, P, *Penetrating the fog-analytics in learning and education*. Asynchronous Learning Networks, 2011.
- [6] Chatti, M, A, Dychhoff, A, L, Schroeder, U., and Thus, H. A reference model for learning analytics, *International journal of Technology Enhanced learning*, 2012.
- [7] Simon Buckingham Shum "Learning analytics" UNESCO Policy Brief, UNESCO IITE.
- [8] Jo yongsang, *Learning analysis techniques utilize and prospects - use cases and service models*, Journal of the Korea Communications, vol. 31, no. 12, 73-80, 2014.
- [9] Bakharia, A., and Dawson, S. SNAPP: A bird's-eye-view of temporal participant interaction. *Proceedings of the First International Conference on Learning Analytics and Knowledge-LAK'11*, pp. 168-173, 2011.