

Information Analysis and Research on Classroom Teaching in Colleges and Universities in Big Data Era

Jixia Song¹ Xueqing Song^{2,*}

¹Jilin Engineering Normal University, Changchun, Jilin 130052, China

²College of Humanities & Sciences of Northeast Normal University, Changchun, Jilin 130117, China

*Corresponding author. Email: 1285516@qq.com

ABSTRACT

The data information presents new value in big data era. New information technology and analysis method promote the application of big data analysis in various fields so that the potential value behind the data is explored infinitely. Big data analysis also shows unique advantages in references retrieving, collecting, collation, analyzing and making conclusion. The participation of big data method helps people effectively dealing with classroom teaching information, exploring the new knowledge and cognition behind the information, and creating a new knowledge production mechanism. This method is of great significance to the current research-oriented classroom teaching in colleges and universities, and can improve the teaching quality and truly realize the idea of combining teaching and researching to promote teaching.

Keywords: big data, information technology, information analysis, classroom teaching

I. INTRODUCTION

Since the beginning of the new century, data has been growing at an unprecedented rate, with the popularity and development of the Internet, as well as the rise of forum/BBS, Weibo, WeChat, Twitter and other online communities from the media. The McKinsey Global Research Institute forecast that global data usage would be expected to reach 35ZB in 2020 [1]. The massive data cloud storage and cloud computing solved the technical problems in data centralized storage and centralized computing, which made it possible for big data to enter the practical application level as a technology, method and concept. Nowadays, Big Data has become the key word of this era, which affected all aspects of social life, from business technology to health care, government, education, economy, humanities and other fields of society [2]. The application of big data in the field of education has achieved remarkable results, but the application in classroom teaching is still insufficient, which should attract more attention.

II. THE NEW CHARACTERISTICS OF DATA AND INFORMATION IN BIG DATA ERA

As early as 1980, the American futurist Alvin Towler predicted the emergence of Big Data in his book *The Third Wave*. In 1997, Mike Lesk noticed that the Internet made data information at a rate of 10 times a year in *How Much Information in the World*. In 1998, the *Journal of Science* published big data processing

and put forward the concept of big data again. It's pointed that the quantitative growth of data had exceeded the ability of typical database tools for collecting, storing, managing, and analyzing [3]. In September 2008, Nature launched the Big Data column, *Big Data: Science in the Petabyte Era*, put forward the concept of big data in the modern sense for the first time, and summed it up as a new idea and method of data information processing. [4] The year of 2011 was a big year for big data. Science launched the "Dealing with Data" column and made great efforts to promote the research and application of big data. In May of that year, the 11th EMC World Congress was held with the theme of Cloud Meets Big Data. the McKinsey Global Research Institute (MGI) released the report named *Big data: The Next Frontier For Innovation, Competition, And Productivity*, which foresaw the arrival of the big data era. Big data quickly attracted the attention of HP, IBM, Microsoft and other companies, as well as countries all over the world, from the state to the civil society are actively involved in the development and utilization of big data.

An important feature of big data is variety in the data structure itself, that is, the heterogeneity of data information. The data information of big data era is no longer a single structured data in the Small Data Age in the references, literature and database. In addition to the structured data, a large number of semi-structured and unstructured data, such as documents, pictures, images, audio, video, web pages, which cannot be represented by digital or unified structure, began to appear. Xu

estimated that 75% of the data in the world is unstructured in his book [5], and now it could dominate even higher percentage since the amount of unstructured data was growing significantly. In this case, the construction of unstructured database has become an important task. The unstructured database adopted the way of hanging files to image, sound and other multimedia information, and realized the full-text retrieval and efficient management of TXT, HTML, XML, DOC, PPT, PDF and some multimedia data [6]. Semi-structured and unstructured data cannot be processed directly by big data. While the high information value motivated to seek new technology to convert them into structured data that can be processed. How to retrieve and analyze the information has become an important topic given to us by the times.

III. FORMAT CONVERSION BETWEEN INFORMATION TECHNOLOGY AND UNSTRUCTURED DATA

At present, the main data formats that big data can recognize directly are TXT, DOC, RTF and so on. For HTML, XML, PPT, PDF, PDG, and image, photocopying information and other data formats, there are still difficulty in identification. There are two ways to identify these unstructured data: one is to speed up the format transformation of unstructured data structure, the other is to improve the information recognition technology of unstructured data.

The importance of digitizing data information is noticed by both the government and the experts. The so-called digitization is mainly to re-enter the text, image, table, number and other information electronically, and convert it into measurable data information, which is convenient for computer information recognition, retrieval and processing. Especially, many libraries began to dig out the documents in their collections and invested a lot of manpower and material resources to digitize the documents. But currently, the digitization of library documents is mainly to convert the documents into electronic version, and there are many problems in the process of digitization, such as the selection of the version of the documents, the suspicious words in the text, different words and various errors, especially those ancient inscriptions, manuscripts and transcripts, damage, blur, handwriting and so on which caused difficulties in identification. It is precisely because of excessive literature, difficult to identify, and insufficient manpower and material resources, so more use of photography, scanning, photocopying and other techniques to store part of the literature data into the database. Frankly speaking, This method is not the best but a good way of transforming the valuable documents into data information as soon as possible, which can expand the utilization rate of documents through

network communication and avoid the waste of resources.

By far, the digitization of literature resources cannot meet the requirements of the current data format of information analysis to a large extent, it needs further data conversion, especially those huge amount of unstructured data. Because of the difficulty in digitizing data and information, it not only strengthens the digitizing of documents but also accelerates the upgrading of computer information technology, and develops increasingly more effective information technology to improve the identification, retrieval and analysis of data. For example, the current DjVu color image compression technology used in document image processing improves the quality of image scanning, improves the clarity and recognition of the image, and most importantly, the scanned image data can be directly and easily transmitted to the Internet, which can be read by readers and retrieved by researchers. At the same time, in order to enhance the recognition ability of unstructured data, the optical character recognition (OCR) system was developed in the computers, which can use the technology of electronic equipment or software to recognize the unstructured data such as image, audio, video and so on into computer text. OCR technology greatly improves the efficiency and the volume of information retrieval, but also provides technical support for information analysis.

IV. BIG DATA ANALYSIS IN RESEARCH-ORIENTED CLASSROOM TEACHING ACTIVITIES

At present, the main methods of big data analysis are quantitative analysis, multi-source data fusion and related relations method.

Quantitative analysis is an important method for big data analysis, and it can be said that quantitative thinking is an important way of thinking in the age of big data. And quantitative statistical analysis of all the data is carried out on all the data. The computer information technology is used to re-classify and integrate the massive data, and the big data is transformed into small data, and the data mining, statistical analysis, and the literature measurement and other quantitative analysis methods are adopted to carry out the analysis and treatment, and the potential value behind the data is mined. The quantitative analysis method mainly includes cluster analysis, association rule mining, time series analysis, social network analysis, risk analysis, prediction analysis and so on. [7]

The multi-source data fusion is a special method of big data analysis. The information analysis in the big data age is more complex and the data source facing is more diversified. It is often necessary to deal with different types of data sources, and to complete the data analysis on the basis of data fusion. In order to obtain a

more accurate result. Therefore, multi-source data fusion refers to the process of bringing together data of different structures to form a dataset with a uniform format and facing a plurality of applications by using a plurality of acquisition modes. [8] One aspect of multi-source data fusion requires the progress of information technology, and on the other hand, it also puts forward higher requirements for the main body of information analysis. It is necessary to pay attention to the integration of various types of data sources, and to realize the processing and judgment of the data, and finally to obtain the reasonable and correct conclusion.

The relevant relation is a basic principle of the big data analysis method. The relevant relationship principle is insist on using no matter in the information retrieval and the information analysis. All relevant data are collected and sorted out so the comprehensiveness of the data retrieval is guaranteed. The analyzing deviation of the results is avoided as a result of one-sided data loss. In the process of data analysis, more attention is paid to the mutual relationship among all kinds of data. By finding the relation between the change rules of the data value and the ones hidden, the change trend of the hidden value is deduced and the prediction function of the big data is realized. This is also one of the most noticeable features of big data.

In the process of college classroom teaching, we need to focus on the related relationship in the teaching activities. Through comparing and analyzing the differences of the learning ability and the learning habits among the college students in different-level colleges as well as in the similar leveled colleges and in the same major, the learning situation of the contemporary college students are really understood and then the teaching contents and methods for the college students is adjusted. The change of teachers' behavior drives that of the students'. [9] At the same time, we should also pay attention to the pluralistic data sources to search for the current political, economic, cultural, geographical and other relevant databases. It's useful to find out the reasons behind the problems of college students classroom learning on the basis of multi-source data fusion. In the case of full data collection and arrangement, the data information analysis process is carried out by a quantitative analysis method, the geographic information system (GIS) technique can be used to collect the learning differences among college students from different places of origin, as well as to quantitatively analyze the trail of college students learning experience, and finally figure out the relationship between the change of college students learning trajectory and their learning ability.

At the same time, we can use the clustering statistical analysis method in the element statistical analysis tool, e.g. Spss17.0, to analyze the learning effect of the college students and the teaching effect of

the professors. For example, in the process of the four-year learning, the information sequence is organized according to the order of time and space, and the time series has a certain dynamic characteristic. With the increasing of the years and the transformation of space, the learning situation of the university students is increasing and changing. At the same time, the regression analysis method is used, which is a derivation method of the big data correlation analysis method, the computer modeling can be used to present the relationship between the learning situation of students in a college and in the others, and to obtain the quantitative relation between the two information indexes. These research results will enrich the content of the teaching material and give the students the most comprehensive and three-dimensional data analysis.

V. CONCLUSION

In summary, information analysis mainly uses modern information technology and methods to collect, organize, analyze and demonstrate data information and other deep processing activities to provide valuable information services and products. [10] As a new information analysis method, big data analysis can get complete data of teaching activities of professors and learning activities of students, help us to discover problems, improve teaching quality and realize effective reform of Chinese university education. As some researchers have found, there is no subsequent information and knowledge without data processing, not mention much higher-level wisdom and thought.^[11] Big data analysis is a tool with the ultimate goal of helping us realize the improvement of higher education level and providing new ideas for the classroom teaching in colleges and universities.

References

- [1] Cohen J, Dolan B, Dunlap M, et al. MAD skills: New Analysis Practices for Big Data [J]. *PVLDB*, 2009, 2(2): 1481-1492.
- [2] Viktor Mayer-Schönberger, Kenneth Cukier. Translated by Yangyan Sheng and Tao Zhou. *Big Data-A revolution That Will Transform How We Work, Live and Think*. Hangzhou, Zhejiang People's Publishing House. 2013. p15.
- [3] Bill Franks. *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics* [M]. Huang Hai, Che Yuyang, Wang Yue, etc. Beijing: The People's Posts and Telecommunications Press, 2013:4.
- [4] Big data: Science in the petabyte era [J]. *Nature*, 2008, 455: 1-136.
- [5] Zipei Xu. *Big Data: The Data Revolution Coming, and How It Reforms the Government* [A]. *Business and Our Life* [C]. Guilin: Guangxi Normal University Press, 2012: 57.
- [6] Yong Li. Construction, sharing and service of information resources under iBASE and network environment [J]. *China Science & Technology Resources Review*, 2001, (8): 62-64.
- [7] J Manyika, M Chui, B Brown, et al. Big data: the Next Frontier for Innovation, Competition, and Productivity [EB/OL]. [2014-05-06].

[http://www.mckinsey.com/Insights/MGI/Research/Technology and Innovation/ Big data: The next frontier for innovation.](http://www.mckinsey.com/Insights/MGI/Research/Technology%20and%20Innovation/Big%20data%20The%20next%20frontier%20for%20innovation)

- [8] Guangjian Li, Bolin Hua. Relationship between Big Data Analysis and Intelligence Analysis [J]. *Journal of Library Science in China*. 2014 (5): 14-22.
- [9] Su Mu, Jing Dong, Dongmei Tang, Yanjie Zhou. The Influence of Teacher's Behavior on Student Activities in Informatizationized Classroom Teaching [J]. *China Educational Technology*. 2019(8): 91-98.
- [10] Changhuo Bao, Qihao Qi, Xinzhou Xie. Research on the understanding and countermeasures of intelligence research in China [J]. *Information Studies: Theory & Application*, 1997(3): 133-135.
- [11] Yongxiao Zheng. Accelerate the Change from Digitization to Datamation-Theory of "Mass Data", "Cloud Counting" and Research of Classical Literature [J]. *Literature Heritage*, 2014(6): 141-148.