Literature review on Big Data and Its Application Fields

Xiani Fan¹,*, Zeping Li² and Li Zhou²

¹School of information, Beijing Wuzi University, Beijing, China
²School of information, Beijing Wuzi University, Beijing, China

*Corresponding author e-mail: 1227656316@qq.com

Abstract. In recent years, due to information technology such as computer, Internet and the development of sensor technology, modern life has appeared "everything can be a digital" thinking, as a complex system test, test and condition monitoring of data series in growth, big data is becoming a hot research topic in the field of all kinds of industrial. This paper will start from the background of the era of big data, discuss the features of big data and its application in reality, and elaborate the trend of big data development from the perspective of data science, industry 4.0 and information physics system, and analyze the challenges of big data development.

1. The Introduction

With the continuous development of information technology, massive data has become the most valuable wealth in recent years. In today's extremely rapid information dissemination, various kinds of data permeate our lives, and they grow at an exponential rate. [1]The data explosion brings us into the era of big data. Big data starts to affect the development of our life, work, study and society. Therefore, relevant researches on big data are highly valued by all walks of life. The advent of the era of big data has upended the recognition of traditional data by industry and academia, and also led to the transformation of technologies such as data acquisition, storage, analysis, mining and visualization. Based on the detailed introduction of the meaning of it, this paper summarizes the technical system of big data processing and forecasts the development of big data.

2. The Connotation of Big Data

2.1. Definition of Big Data

According to wikipedia, big data refers to data sets that use common software tools to capture, manage and process data in excess of tolerable time limits. [2,3]The world-renowned management consulting company refers to data sets beyond the acquisition, storage, management and analysis capabilities of traditional database management software as big data. Relevant research institutions have classified big data into large-scale, high-speed growth and enrichment of information materials that require new deal modes to enhance decision making capacity and level of optimization process. [4]In his report to the xiangshan scientific conference, academician Xu Zongben believes that big data is a large and complex data set cannot be centralized storage and are difficult to be understood and
analyzed in the allowed time frame, in which individual or partial data present low value while the
data overall present high value”. Although more than about big data definition methods, the definition of Angle and different emphasis, but the message, namely, large data in the final analysis is a kind of data sets, its feature is compared with the traditional data management and processing technology to highlight, and under the different needs, it's time to deal with scope have differences, the most important point is the value of big data, not the data itself, but by "big decision", big data is a reflection of the "great knowledge", "big problem", etc.[5]

2.2. Generation of Big Data

The generation of data in today's era has been completely free from the limitations of time and place. Starting from the adoption of database as the main method of data management, the data production mode of human society has roughly experienced three stages, and it is the huge change of data production mode that finally leads to the production of big data.

2.2.1. Operational System Stage. The appearance of database greatly reduces the complexity of data management. In practice, database is mostly adopted by the operation system as the data management subsystem of the operation system. For example, supermarket sales recording system, bank transaction recording system, hospital patients' medical records. [6] The first big leap in the amount of human social data was built on operational systems starting to make extensive use of databases. The most important feature of this stage is that data is often generated and recorded in the database accompanied by certain operational activities. For example, for every product sold by supermarkets, a corresponding sales record is generated in the database. The way this data is produced is passive.

2.2.2. User Original Content Stage. The birth of the Internet has led to the second big leap in the amount of data in human society. However, the real data explosion occurred in the era of Web2.0, and the most important sign of Web2.0 is user-generated content. Such data have been growing explosively in recent years, for two reasons. [7,8] First of all, the emergence and rapid development of new social networks represented by blogs and micro blogs have made users more willing to generate data. Secondly, there is the emergence of new mobile devices, such as smart phones and tablet computers. These portable, all-weather mobile devices make it easier for people to express their opinions online. The generation of data at this stage is active.

2.2.3. Perceptual System Stage. The third big leap in the amount of human social data eventually led to the production of big data, and today we are at this stage. The fundamental reason for this leap is the widespread use of perceptual systems. With the development of technology, people have been able to make extremely small sensors with processing functions, and have begun to widely arrange these devices in all corners of society to monitor the operation of the whole society. These devices generate a stream of new data in an automated way. In simple terms, data generation has experienced three stages: passive, active and automatic. These passive, active and automatic data together constitute the data source of big data, but automatic data is the most fundamental reason for the generation of big data.

3. Features and Technical System of Big Data

3.1. The Role of Big Data

With the evolution of time, the industry has a deeper and more comprehensive understanding of big data. In addition to the scale, speediness and diversity of data structure, the specific characteristics of big data in different application fields are also different. For example, the Internet field needs to process and analyze users' purchasing behaviors in real time, so as to make push plans in a timely manner, which requires high precision and reliability. [9] The medical field needs to judge the patient's
condition based on information such as user cases and images. Since it is closely related to people's health, its accuracy and reliability requirements are very high.

Table 1. Characterization of big data

| Characteristic | Connotation                      |
|---------------|---------------------------------|
| Volume        | Large amount of data            |
| Velocity      | Fast data analysis and processing|
| Variety       | Diverse data types              |
| Value         | Value sparsity                  |
| Veracity      | Data reflect objective facts     |
| Variability   | Big data has multi-layer structure|

3.1.1. Big Data is the Power to Transform Values. In the future, the core meaning criterion to determine whether China has great wisdom is that national happiness is reflected in people's livelihood. Secondly, it is reflected in the ecology. Through big data, meaningful things can be made clear. In short, let us move from the previous chaotic era to the future explicit era of meaning.

3.1.2. Big Data is the Power to Transform the Economy. Producers are valuable and consumers are what value is. What is meaningful is valuable. If the consumers do not agree with it, they will not sell it and will not realize the value. Only if the consumer agrees, can sell out, can achieve the value. Big data helps us identify meaning from the source of consumers, thus helping producers realize value. That is how domestic demand is kick-started.

3.1.3. Big Data is the Power to Transform the Organization. With the development of data infrastructure and data resources with semantic web characteristics, organizational change becomes more and more inevitable. Big data will drive the network structure to produce disorganized organizational power. The first to reflect this structure is a variety of decentralized WEB2.0 applications, such as RSS, wikis, blogs, and so on. The reason why big data has become the power of time change is that it gains wisdom by following the meaning.[10]

Table 2. Specific features of big data in different fields

| Field                  | Number of users | The response time | The data size | Reliability requirement | The precision requirement |
|------------------------|-----------------|-------------------|---------------|-------------------------|----------------------------|
| Scientific computing   | small           | slow              | TB            | general                 | very high                  |
| Financial              | big             | very fast         | GB            | very high               | very high                  |
| Medical                | big             | fast              | EB            | high                    | very high                  |
| The Internet of things | big             | fast              | TB            | high                    | high                       |
| The social network     | a very large    | fast              | PB            | high                    | high                       |
| Mobile devices         | a very large    | fast              | TB            | high                    | high                       |
| Multimedia             | a very large    | fast              | PB            | high                    | general                    |

3.2. Technical System of Big Data

The big data processing technology system mainly involves four parts: big data collection technology, storage technology, analysis and mining technology and visual presentation technology.

3
Big data collection: big data from different fields has different characteristics, data quantity and number of users. According to the structure characteristics, it can be divided into three types: structured data, semi-structured data and unstructured data. The challenge of big data acquisition is high concurrency and fast streaming data speed.

Storage of big data: the improved lightweight database can be used to complete storage of big data and respond to users' simple query and processing requests; However, when the data volume exceeds the storage capacity of lightweight database, large-scale distributed database or storage cluster platform is needed. [11]With the development of Internet technology and cloud computing technology, cloud storage based on distributed storage has become the main trend of big data storage. The main challenges of big data storage are heterogeneous data, diverse structure and large scale.

Analysis and mining of big data: the analysis of big data involves simple statistical analysis and classification summary. The challenge lies in large amount of imported data and multiple query requests. Big data mining involves data classification, clustering, frequent item mining, etc.[12]

Big data visualization: the mining and analysis results of big data will be presented in the display terminal in a friendly, image and easy-to-understand form for the accuracy of the results of professional analysis or to provide users with decision-making information support. The challenges presented by big data lie in high data dimensions and diversified demands.

The mutual use of various technical functions in big data processing can provide a technical foundation for the effective realization of the value of big data.

4. Development Trend of Big Data

4.1. Change the Way of Thinking

In 2007, Jim Gary put forward scientific paradigm - 4 "data-intensive science", before the three kinds of experimental science, scientific paradigm theory science and computing science, the fourth paradigm proposed marks the data for the ascension of the importance of scientific research, its essence is a scientific research will be from computing centered to data-centric, namely the arrival of the data of thinking. "Data intensive science" has been widely concerned by researchers in the field.

Big data not only changed the development of scientific thinking, also will inevitably lead to enterprise and government, the change of personal way of thinking, for big data era, should give up for causal care, and pay more attention to relationship, in the era of big data, people wake up every day, is to want to so much data can be used to do, its value can be reflected in which aspects, and if I can find a people never involve things makes ideas and thoughts become important assets. [13]Therefore, it can be seen that the big data era will inevitably lead to the transformation of thinking, and the faster the transformation of thinking, the more it can seize the opportunity in today's highly competitive society.

4.2. Change Survival Mode

In the 21st century, with the rapid development of information technology, the development of Internet of things, embedded technology, sensing technology, etc. provides the foundation for human beings to perceive the physical world with objective existence more comprehensively. The development of the Internet, other information technologies and cloud computing has changed the way humans communicate and manage information.[14] With the development of technology and the upgrading of tools, humans also put forward higher survival demands. Big data focuses on "data", focuses on actual business, provides data collection, analysis and mining, and focuses on information accumulation, namely data storage capacity. Cloud computing focuses on "computing", focuses on solutions, provides infrastructure, and focuses on computing power, or data processing power. Without the information accumulation of big data, the computing capacity of cloud computing is no stronger, and it is difficult to find a place to use it: without the processing capacity of cloud computing, the information accumulation of big data will be no more than a picture in the end. Technically, big data is rooted in cloud computing. Massive data storage technology, massive data management technology and other key technologies in cloud computing are the foundation of big data technology.
4.3. Change of Production Mode

With the further development of science and technology, the progress of science and technology is bound to bring about the change of production mode. Therefore, Germany put forward the "4.0" industry, namely the fourth industrial revolution, dominated by intelligent manufacturing implementation manufacturing integration, "4.0" industrial herald the birth of revolutionary production mode is put forward, and is the basis of realization of "industrial 4.0" big data analysis and the expansion of the CPS, it marks the manufacturing must shift to centered on data analysis. It can be seen that the development of big data will play a key role in the change of production mode.[15]

The industry 4.0 plan emphasizes the fourth industrial revolution with the Internet of things and and large production industries service. Although there is a great controversy about the arrival of the fourth industrial revolution, many countries have invested a lot of money and energy to promote the process of "industry 4.0". [16]The goal of "industry 4.0" is to realize the intelligent factory through the Internet of things system, that is, every product and component will contain a large amount of information, including when it is produced, how long it can be used, and whether it needs to be replaced, etc., through the intelligent way of non-human intervention. It can be seen that big data will play an important role in changing the mode of production.[17]

5. Conclusion

Big data as an important resource of present and future, has appeared in several areas of production and life, caused the extensive concern of various departments, and will become the future market competition and the importance of science and technology innovation competition for resources, but its value embodies need to break through the limitation of traditional data analysis and processing, attach importance to the relationship between data, in the case of meet the accuracy requirement rapid response analysis requirements.

Based on the connotation of big data, this paper summarizes and analyzes the key technologies involved in each link of big data analysis. From the perspective of data science, industry 4.0 and information physics system, it forecasts the important role of the development of big data in changing human thinking mode, production mode and life mode. We should grasp the key technologies in the process of big data processing, establish a continuous research system, grasp the development opportunity of big data, and make full use of big data to create nice value.

Acknowledgments

This research is led by the China Logistics Association's 2018 key project “Research on Interconnection of Modern Logistics Network System” (2018CSLKT2-004), National Natural Science Foundation Project (71501015), Beijing Great Wall Scholar (No.CIT & TCD20170317), Tongzhou Canal Project Leaders and The Intelligent Logistics System is funded by the Collaborative Innovation Center.

References

[1] MANYIKA J, CHUI M, BROWN B, et al. Big Data: The next frontier for innovation, competition and productivity[R]. Mckinsey Global Institute, 2016.
[2] TOFFER A. The third wave[M]. New York: Bantam Books, 2011.
[3] KATALA, WAZIDM, GOUDARRH. Big data: issues, challenges, tools and good Practices[C]. 2013 Sixth International Conference on Contemporary Computing ( IC3) , Noida, INDIA, 2017: 404-409.
[4] LIUZY, YANGP, ZHANGX. A sketch of big data technologies[C]. 2016 Seventh International Conference on Internet Computing for Engineering and Science( ICICSE 2016) , Shanghai, China, 2016: 26-29.
[5] KAGERMANNH, WAHLSTERW, HELBIGJ. Securing the future of German manufacturing
industry recommendations for implementing the strategic initiative INDUSTRIE4.0[R]. Germany: Federal Ministry of education and research, Final report of the Industrial 4.0 working group, April 2017.

[6] Big Data[EB/OL].[2012-10-02]. http://en.Wikipedia.org/wiki/Big_data.

[7] IBM. What is big data[EB/OL]. [2012-10-02]. http://www-01.ibm.com/software/data/big data/what-is-big-data.html

[8] EMC (2017) Big data: big opportunities to create business value. EMC Corporation, Hopkinton

[9] Fang JQ (2015) Network science and engineering faced with a new challenge and developing opportunity under the wave impact of big data. Chin J Nat 35:345–354 (in Chinese)

[10] Guo HD (2012) China’s earth observing satellites for building a digital earth. Int J Digit Earth 5:185–188

[11] BRANTNER M, FLORESCU D, GRAFD, et al. Building a database on S3[C]. Proceedings of the ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada, 2008:251-263.

[12] GHEMAWAT S, GOBIOFFH, LEUNG S-T. The Google file system[J]. ACM, 2013, 37(5) : 29-43.

[13] DEAN J, GHEMAWAT S. Map Reduce: Simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1) : 107-113.

[14] TOSHNIWAL A, TANEJA S, SHUKLA A, et al. Storm @ Twitter[C]. 2014 ACM SIGMOD International Conference on Management of Data, Snowbird, UT, United States, 2014: 147-156.

[15] ZAKI MJ, AGGARWAL CC. X Rules: An effective algorithm for structural classification of XML data[J]. Machine Learning, 2006, 62 (1-2) : 137-170.

[16] GUO HD, WANG LZ, CHEN F, LIANG D. Scientific big data and digital earth[J]. Chinese Science Bulletin, 2014, 59(35) : 5066-5073.

[17] ARMES T, REFERNM. Using big data and predictive machine learning in aerospace test environments[C]. IEEE AUTOTESTCON, 2013: 16-19.