Application and Prospect of Big Data

Wenguo Liao[1], Guangping Liao[2,*]

[1]Dean's Office, ABA Teachers University, Wenchuan County 623002, Sichuan Province, China

[2]College of Resources and Environment, ABA Teachers University, Wenchuan County 623002, Sichuan Province, China; The Tibetan Qiang Institute of Medicine, ABA Teachers University, Wenchuan, Country 623002, Sichuan Province, China

*Corresponding author E-mai: 648705594@qq.com

Keywords: Big Data Technology; Information Technology; Big Data Ecosystem

Abstract: The rise of big data technology is completing the subversion of traditional fields. Globally, it is a trend to use big data to promote economic development, improve social governance, and enhance government service and regulatory capacity. Countries have successively formulated and implemented strategic documents of big data to vigorously promote the development and application of big data. From the trend of global big data development, big data industry promotes the network sharing, intensive integration, collaborative development and efficient utilization of social production factors, changes the traditional production mode and economic operation mechanism, and can significantly improve the level and efficiency of economic operation. China is a big data producer. At present, China's Internet and mobile Internet users rank first in the world, with rich data resources and application market advantages. If we can make a breakthrough in the research, development and application of big data management and analysis technology, we will continue to promote the rapid growth of Internet innovative enterprises and innovative applications.

1 Development trend of big data at home and abroad

1.1 International big data strategic decision

Throughout the world's big data strategies, there are three common points: first, to promote the application of the whole industry chain of big data; second, to attach equal importance to data opening and information security; third, the government and social forces jointly promote the application of big data. This paper takes the United States, Britain, Japan and Germany as examples.

1) The United States. In 2009, the U.S. government launched a public service platform, which fully opened 400000 federal government raw data and geographic data. By analyzing and extracting information from massive data, we can improve the prediction ability of social and economic development. NSF, NIH, DOD, DOE, Dara, USGS 6 federal departments and agencies announced an investment of US $2 billion 200 million to jointly improve the advanced technology of core technologies needed to collect, store, retain, manage, analyze and share massive amounts of data, and work together to strengthen the input and development of information technology to promote the development of supercomputing and the Internet. In 2013, the U.S. issued the administrative order on government information disclosure and machine-readable, which requires the disclosure of data in seven key areas, including education and health, and puts forward clear requirements for the data opening time of various government agencies. In November 2013, the information technology and innovation foundation of the United States issued the technology and policy to support data-driven innovation, which pointed out that the government should not only vigorously cultivate the required skilled labor force and promote the research and development of data-related technology, but also develop a legal framework to promote data sharing, and improve the public's awareness of the great significance of data sharing [1]. In May 2014, the United States released the white paper "big data: seize the opportunity, guard the value", which focused on the current

situation, policy framework and improvement suggestions of big data application and management in the United States. In April 2016, MIT launched the "data America" online big data visualization tool, which can analyze and display the open data of the US government in real time.

2) Britain. In November 2011, the UK government issued a strategic decision to study public data, and established the "UK data bank" data.gov.uk .The website hopes to further support and develop the development of big data technology in science and technology, commerce, agriculture and other fields by fully releasing government data. In May 2012, the British government invested 100000 pounds to support the establishment of the world's first open data Institute (ODI). ODI Research Institute will provide data background support for those business activities beneficial to the public, which not only releases new business potential, but also promotes new forms of economic development and personal income growth. In May 2013, the UK government and the Li Ka Shing Foundation jointly invested £ 90 million to establish the world's first medical and health research center that comprehensively uses big data technology at Oxford University. Through the collection, storage and analysis of a large number of biomedical data, the center will work with the industry to define the direction of new drug research and development, handle the bottleneck in the process of new drug research and development, and provide clues for the discovery of new treatment methods. In August 2013, the British government released the British agricultural technology strategy. The strategy points out that UK investment in agricultural technology will focus on big data in the future, with the goal of commercializing UK agricultural technology [2]. In 2014, the UK government invested £ 73 million in the development of big data technology, including the application of big data technology in 55 government data analysis projects, and established big data research centers, such as Turing big data research institute, relying on institutions of higher learning. In 2015, the UK government promised to open a core public database on transport, weather and health.

3) Japan. In June 2012, Japan's IT strategy headquarters issued a draft of the E-government open data strategy, which took a key step in government data disclosure. Our goal is to focus on the development of smart technologies such as social media, traditional industry it innovation, new medical technology development, traffic congestion alleviation and other public domain applications needed for big data application.

This is the official announcement of the new IT strategy - the creation of a state-of-the-art it manifestoes. This paper comprehensively expounds Japan's new it national strategy of developing open public data and big data during 2013-2020, and proposes to build Japan into a society with the highest level in the world and extensive use of information industry technology. To this end, the Japanese government launched a data classification website. The purpose is to provide data from different government departments and institutions for use and to open data to data providers and data users. The data includes various white papers, geospatial information, crowd movement information, budget, year-end financial and process data, etc. In July 2013, Mitsubishi Institute of comprehensive research led the establishment of the "open data circulation promotion alliance", which aims to promote the open application of public data in Japan by combining industry, government and education. In August 2014, the Japanese cabinet government decided to use the "big data" accumulated on the Internet as a new economic judgment indicator in the monthly economic report published every month. The cabinet will analyze the real-time consumption trend based on the search results of Internet users for products and services and the posts posted on twitter. Japan's defense ministry will also formally discuss the application of "big data" to overseas situation analysis from 2015. As a new program in the context of the expansion of the overseas activities of the self defense forces, this initiative aims to strengthen the intelligence collection capacity.

4) Germany. In 2010, Germany formulated the "ICT strategy of digital Germany 2015" and adopted modern ICT technology to realize intelligent networking in traditional industries such as energy, transportation, health care, education, leisure, tourism and management. In April 2013, the German government proposed the concept of "industry 4.0".Bitkom, the German IT Industry Association, reported in early 2014 that big data business is developing rapidly in Germany and is expected to reach 13.6 billion euros by 2016.[3]

1.2 Transformation of international big data industry

In June 2013, Snowden, a former CIA employee, unveiled the tip of the iceberg of "data war". In fact, the prism program of the United States includes all countries and individuals under the supervision of the National Security Agency (NSA). The companies involved in prism plan include Google, Yahoo, Facebook, and Microsoft, apple, Cisco, Oracle, IBM and other technology giants. It can be seen that in the era of big data, the strength of IT industry has become one of the most critical factors directly determining whether a big country becomes a powerful country. The industry needs to change, and the industry needs to be interconnected. The so-called "big data +" is to graft big data thinking into different industries and promote the implementation of big data in all walks of life. Big data is not only related to the IT industry, but also many industry leading companies have realized the huge impact of new big data thinking. Internet, finance, telecommunication, medical treatment and government are the key areas of big data operation. However, the development and application of big data in most fields are still in the primary stage.[4] In the practice process of big data application, there are also challenges such as unclear data assets, uncertain application needs, platform construction, technical route, security and privacy issues. However, big data application has made some useful exploration in various fields and made some achievements.

In the telecommunication industry, the utilization of big data by telecommunication operators in some developed countries, on the one hand, improves service quality and internal management, including customer maintenance, precision marketing and network operation and management. The representative enterprises are France Telecom, UK O2, NTT DoCoMo and Vodafone. France Telecom carried out big data analysis and evaluation for user consumption, improved service level and user experience with big data; UK O2 launched free Wi-Fi service in the UK to accumulate more users, so as to collect more user data for accurate media advertising and marketing services; NTT DoCoMo has greatly strengthened CRM system and knowledge base, accurately positioned target customers, and improved the success of business management by making refined forms and collecting user details. [5]

Big data in the information network is transformed into usable information. On the other hand, establish a business model to generate external revenue, including direct sales of data for revenue, as well as cooperation projects with third-party companies to create profits for operators, including at & T, Spain Telecom, dynamic insights, Verizon, Deutsche Telecom and Vodafone [6]. At & T sells data related to users to governments and enterprises for profit; Telefonica has established a dynamic insight department; dynamic insights carries out big data business, provides data analysis and packaging services for customers, and cooperates with GfK, a market research institution, to launch the first product in the UK and Brazil called smart steps; Verizon Precision marketing department established Division, which provides services such as precision marketing insight, precision marketing, mobile commerce, etc., including big data analysis of its user base with a third-party organization, and then providing valuable information to the government or enterprises for additional value. The profitability of data business accounts for a very high proportion of its entire business[7].

In the field of outsourcing, big data technology has become the "next big event" in the information technology industry. At present, some outsourcing giants have begun to enter the big data market, trying to carve up this big cake. India's National Association of software and service enterprises forecasts that the size of India's big data industry will reach US $1.2 billion in three years, six times the current size, and twice the average growth rate of the global big data industry.

In the information security industry, international enterprises such as fire eye and Spelunk are developing rapidly in big data security, and their technology in big data security is also worth learning from domestic enterprises. Websense Company, which specializes in DLP products, their analysis technology based on data flow is very conducive to the analysis and mining of big data [8].

2 Big data technology progress

2.1 Big data life cycle

The bottom layer is infrastructure, including computing resources, memory, storage and network interconnection, which is embodied in computing nodes, clusters, cabinets and data centers. On top of this is data storage and management, including file system, database and resource management system similar to yarn. Data analysis and visualization are based on the computational processing layer. Analysis includes simple query analysis, flow analysis and more complex analysis (such as machine learning, graph calculation, etc.). Query analysis is mostly based on table structure and relation function, flow analysis is based on data, event flow and simple statistical analysis, while complex analysis is based on more complex data structure and method, such as graph, matrix, iterative calculation and linear algebra. Visualization of general meaning is the display of analysis results. However, through interactive visualization, we can also ask exploratory questions, so that the analysis can get new clues and form iterative analysis and visualization. [9]The visualization analysis of real-time interaction based on large-scale data and the introduction of automation in this process are the current research hotspot. There are two areas that have opened up the above-mentioned layers vertically and need to be viewed as a whole and coordinately. The first is programming and management tools. The direction is that the machine can realize automatic optimization through learning, as far as possible without programming and complex configuration. Another area is data security, which runs through the entire technology stack. In addition to these two fields, there are also some technical directions that span multiple layers. For example, "memory computing" actually covers the entire technical stack.

2.2 Big data technology ecology

Hadoop is a big data distributed system infrastructure developed by the Apache foundation. Users can easily develop and run distributed programs dealing with large-scale data on Hadoop without knowing the details of the distributed bottom layer, making full use of the power of cluster for high-speed computing and storage. Hadoop is a data management system. As the core of data analysis, it collects structured and unstructured data, which are distributed in every layer of traditional enterprise data stack. Hadoop is also a large-scale parallel processing framework, with supercomputing power, which is positioned to promote the implementation of enterprise applications. Hadoop is also an open source community, which mainly provides tools and software for solving big data problems. Although Hadoop provides many functions, it should still be classified as a Hadoop ecosystem composed of multiple components, including data storage, data integration, data processing and other special tools for data analysis.

2.3 Big data acquisition and preprocessing

For multiple heterogeneous data sets, further integration processing or integration processing is required. After data collection, sorting, cleaning and transformation from different data sets, a new data set is generated to provide a unified data view for subsequent query and analysis processing. There have been many researches on heterogeneous database integration technology in management information system, entity identification technology and Deep Web integration technology in Web information system, and data fusion technology in sensor network, and great progress has been made. A variety of data cleaning and quality control tools have been launched, for example, data flux of SAS company in the United States, data flux of IBM company in the United States Data stage, Informatica power center of Informatica company in the United States.

2.4 Big data storage and management

Traditional data storage and management are mainly structured data, so RDBMS can unify the whole world to meet all kinds of application needs. Big data is usually semi-structured and unstructured data, supplemented by structured data, and various big data applications are usually for different types of data content retrieval, cross comparison, deep mining and comprehensive analysis.

In the face of this kind of application demand, the traditional database is difficult to sustain either in technology or function. [10] Therefore, in recent years, oldsql, NoSQL and newsql coexist. In general, according to the different types of data, big data storage and management adopt different technical routes, which can be roughly divided into three categories. The first category mainly faces large-scale structured data. The second category mainly deals with semi-structured and unstructured data. The third category faces the big data mixed with structured and unstructured data.

3. Development trend of big data technology

With the continuous development and research of big data technology, the technical development of each link presents new development trends and challenges. In December 2015, the big data expert committee of China Computer Society (CCF) released the report on China's big data technology and Industry Development, and looked forward to the development trend of China's big data, mainly including the following six aspects.

3.1 Visualization promotes the popularization of big data

In recent years, the concept of big data has been deeply rooted in people's minds, and the big data that the public directly sees is more embodied in a visual way. Visualization is to help users better understand and analyze data objects, find and insight into their internal laws by transforming complex data into interactive graphics. In fact, visualization has greatly narrowed the distance between big data and ordinary people. Even ordinary people who don't know about it technology and non-technical professional regular decision makers can better understand the effect and value of big data and its analysis, so that they can give full play to the value of big data from both the national economy and the people's livelihood. It is suggested that in the research, development and application of big data, the corresponding proportion should be kept for visualization and visual analysis.

3.2 Multidisciplinary integration and the rise of data science

Big data technology is the integration of multi-disciplinary and multi-technical fields, involving mathematics and statistics, computer technology, management, etc., and the application of big data intersects with many fields. This kind of interdisciplinary integration calls for and gives birth to a special basic discipline - Data discipline. The consolidation of basic disciplines will make the interdisciplinary integration more perfect. In the field of big data, many related disciplines, on the surface, have different research directions, but in fact, from the perspective of data, they are interlinked. With the gradual deepening of digital society, more and more disciplines tend to be consistent in the data level, so similar ideas can be used for unified research. People engaged in big data research include not only scientists in the field of computer, but also scientists in mathematics and other fields. It is hoped that the industry will take a broader and more inclusive attitude towards the boundary of big data, including the so-called "small data", and even generalize the boundary of the field to the whole data field and data industry corresponding to the "data science". [11]It is suggested that we should support the basic research of "data science" and try our best to introduce the results of basic research into the field of technology research and application.[12]

3.3 Big data security and privacy are worrying

The security and privacy problems brought by big data mainly include the following three aspects: first, the threats to big data, which are often called security problems, will inevitably become the target of attack when big data technology, system and application gather a lot of value; second, the problems and side effects brought by excessive abuse of big data With, a typical one is personal privacy disclosure, which also includes trade secret disclosure and state secret disclosure brought by big data analysis ability[13]; third, mental and conscious security issues. The threat to big data, the side effects of big data, and the extreme mind of big data will hinder and destroy the development of big data. It is suggested that in the research and development of big data, a basic proportion should be kept for the corresponding security research, and the driving force for the

substantial progress in security may be the negative research on the attack and abuse of big data.[14]

References

[1] Alsheikh M.A., Niyato D., Lin S., Tan H.P., Han Z. (2016). Mobile Big Data Analytics Using Deep Learning and Apache Spark. arXiv preprint arXiv:1602.07031.

[2] Bekkerman, R., Bilenko, M., & Langford, J. (Eds.). (2011). Scaling up machine learning: parallel and distributed approaches. In. Cambridge: University Press.

[3] Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews, 44, 58–75.

[4] Chang, V. (2015a). Towards a Big Data system disaster recovery in a Private Cloud. Ad Hoc Networks, 35, 65–82.

[5] Chung, I. H., Sainath, T. N., Ramabhadran, B., Pichen, M., Gunnels, J., Austel, V., et al. (2014). Parallel deep neural network training for big data on blue gene/q. High performance computing, networking, storage and analysis, SC14: international conference for (pp. 745–753). IEEE.

[6] Doctor, F., Hagras, H., & Callaghan, V. (2005). A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. IEE Transactions on Systems, Man and Cybernatics—Part A Systems and Humans, 35(January (2005)), 55–56.

[7] Doctor, F., Iqbal, R., & Naguib, R. N. (2014). A fuzzy ambient intelligent agents approach for monitoring disease progression of dementia patients. Journal of Ambient Intelligence and Humanized Computing, 5(1), 147–158.

[8] Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., et al. (2013). The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Medicine, 10(1), e1001363.

[9] Hawkins J., George D., (2006). Hierarchical Temporal Memory: Concepts, Theory, and Terminology.

[10] Hawkins, J. (2004). On intelligence. New York: Times Books.

[11] Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). Neural networks and learning machines (vol. 3) Upper Saddle River: Pearson Education.

[12] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.

[13] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82–97.

[14] Hirsch, D. D. (2013). Glass house effect: big data, the new oil, and the power of analogy. The Maine Law Review, 66, 373.