

## China's Next Big Challenge: Big Data

Yue Zhu<sup>1, a</sup>

<sup>1</sup>Graduate School of Chinese Academy of Social Sciences, Beijing City, 102488, China

<sup>a</sup>aeonis@qq.com

**Keywords:** Big Data, Data mining, Computer science, China

**Abstract:** The main purpose of this paper is to argue that big data can be seen as the next big challenge to China. At first, this paper tries to describe what is the importance and value of big data. In section II and III, it expounds the benefits and risks of big data in China. Big data will have big positive impact on Chinese government. And it is a new impetus to the development of economic transformation in China. But there are still many risks, like data security, violation of data privacy, Chinese Internet censorship like the GFW. Finally, this paper concludes that big data are the next big challenge to China and introduces how to promote the development of big data.

### 1. Introduction

The epoch of big data has begun. Big data is “a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them” [1]. Hilbert described big data by the following 5 “V” characteristics: Volume, Variety, Velocity, Variability, Veracity [2].

Nowadays, big data, like energy and food, have become one of national basic strategic resources. It will promote a revolution in thinking. Big data are an important factor of production, management, governance and capital. The use of big data to promote economic development and improve the quality and capacity of government services is becoming a consensus of western world. Several developed countries, like the USA, have formulated and implemented strategic plans for promoting the development and application of big data industry. Big data is also becoming a more and more significant impact on Chinese economy, society and governance. In the near future, big data will become a key basis of national competitions. China should seize the opportunity.

### 2. The Importance and Value of Big Data

#### 2.1 Revolution in Thinking

“Change the instruments, and you will change the entire social theory that goes with them” [3]. Big data can be seen as a creative tool of instrumental rationality. The use of big data will change the way we analysis the world, the meaning of learning and the definition of knowledge. Thus, big data will advocate a revolution in thinking. The collection, storage and analysis of vast information was unlikely to be implemented in the past. But the application of big data can sharply reduce the cost of getting information and provide more accurate, and detailed information efficiently. Mayer-Schönberger and Cukier mentioned, “big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and government, and more” [4]. Big data help make people's access to information so much more efficient and effective than ever before. Furthermore, big data have the power to shift creates a radical shift in how we learn about information networks and human communities. By using it, people have built many new systems of knowledge we never see before, like google, wikipedia and quora, which have advocated a data-driven empirical thinking model. This new model is greatly expanding the range of human knowledge, reducing the cost of learning, also changing the objects of knowledge.

## **2.2 Predicting Human Life**

To some extent, big data can be seen as the new prophet in the Information Age. By using big data, we can deal with extracting information from data, and use it to predict trends and behavior patterns. The core of big data predictive analytics relies on a technology called “data mining”, which could capture relationships between explanatory variables and the predicted variables from past occurrences, and exploiting them to predict the unknown outcome [5]. A research report gives several illustrations to point that we can predict businesses and where the market is going by using Big Data [6]. NPR also given an example that Netflix executives were able to predict that House of Cards would be just what Netflix viewers would want to watch by using big data technology [7]. There are already a dozen successful technology companies who sell the “predictions”. Sense Networks, a company headquartered in New York City, uses real-time and historical personal location data for predictive analytics [8]. Palantir Technologies, the “unicorn” company with clients like the NSA, FBI and CIA, is specializing in big data analysis and prediction which once helped Marines in Afghanistan to predict insurgent attacks, also helped track Mexican drug cartels [9]. With the development and application of big data, the predictive analytics will be a necessary step for decision optimization and have a great impact on capital markets, crisis response, defense, disease response, cyber security, disaster preparedness, etc.

## **2.3 Ruled by Big Data**

Big data is changing forever how organizations decide what to do. We often made decisions by subject feelings in the past. But now we can make decision by big data. Big data-driven decision means you have the talent and leadership strength to get the right information to the right places at the right time and take action. A report by MIT Technology Review unveiled why we are moving into a new epoch ruled by data: “On Nasdaq, trading bots exchange a billion shares a day. Online, advertisers bid on hundreds of thousands of keywords a minute, in deals greased by heuristic solutions and optimization models rather than two-martini lunches. The number of variables and the speed and volume of transactions are just too much for human decision makers” [10]. To compete in such a so large and so complex globally-integrated situation, today’s organizations, including governments and companies, need speed and sophistication in their decision-making. This demands the capability of effective use of information and analytics. Governments and companies can use big data to transform themselves to take advantage of the vast array of available information to improve decision-making and performance mathematically. More specifically, big data can help governments and companies making much better management plans and the optimal decision which are very vital in providing more precisely tailored products or services to citizens and customers.

## **3. How China can Benefit from Big Data**

In the context of the rapid development of global information technology, big data have become an important national strategic resource which has big positive impact on Chinese government, and lead the development of economic transformation in China.

### **3.1 The Impact of Big Data on Chinese Government**

Firstly, big data will reshape the government management and social governance progress. By using big data, Chinese government will greatly enhance its analysis ability which provides a new effective method to solute complex social problems. Big data can also help Chinese law enforcement become more transparent, effective, and efficient.

Secondly, Governance based on big data will promote governmental data sharing and integration, establish more scientific decision-making. Enhancing the capabilities of using big data will effectively release the potential value of data resources, reinforce the protection of network space data, safeguard the national security, and strengthen the national competitiveness of China.

Finally, the development and application of big data will conduce to accelerating the construction of socialist market economic system, promoting to build a rule of law Government, innovation

government, clean government and service-oriented government. It's almost certain that big data will play a key role in the progress of the modernization of Chinese governance capacity.

### **3.2 The Impact of Big Data on Chinese Economy**

Big Data has become a new impetus to the development of economic transformation in China. Big data will spur new waves of economic growth and create more jobs in the background of Chinese economic recession. Big data has become an important driving force in the Internet and other emerging areas, which is promoting the development of economic transformation in China.

In one hand, big data will change the traditional mode of production and economic operation mechanism. ISACA argued that “Enterprises that master the emerging discipline of big data management can reap significant rewards and differentiate themselves from their competitors” [11]. It can deeply affect the organization mode of social division of labor, rapidly promote the innovation of the mode of production organization, and significantly improve the efficiency of economic operation. Moreover, big data will stimulate business model innovation, and enhance the core value of the enterprise. More-targeted marketing that injects customer feedback into product design will help companies do faster innovation through a shorter research and development cycle.

In another hand, big data industry is becoming a new economic growth point. A research by McKinsey Global Institute estimated that “services enabled by personal-location data can allow consumers to capture \$600 billion in economic surplus” [12]. Given that the total number of China's Internet users, including mobile Internet users, is the No.1 in the world, it is more likely that China has the biggest market of big data industry. In “THE DIGITAL UNIVERSE IN 2020: big data, Bigger Digital Shadows, and Biggest Growth in the Far East”, IDC estimated that “by 2020, a third of the data in the digital universe (more than 13,000 exabytes) will have big data value” [13]. China alone will generate 21% of the bit stream entering the digital universe [14]. China is on the cusp of a new wave of productivity growth enabled by big data.

## **4. The Risks of Big Data**

Several risks including data security, data privacy and the Internet censorship in China are threatening the development and application of big data in China.

### **4.1 Fragile Data Security Protection System**

Recent years, data security violation and cyber-attacks are getting bigger and more rampant. China Internet Security Report 2015 shows that in 2015, the number of malware samples are 356 million, a growth of 9.9% compared with 2014. The number of malware attacks are 85 billion 540 million times, a substantial increase of 49.4% compared to 2014 [15]. A report by Internet Society Of China showed that in 2015 the economic losses caused by cyber-attacks is 91 billion 500 million RMB Yuan, equal to 13 billion 200 million dollars [16].

The bigger your data, the bigger the target it presents to criminals with the tools to steal and sell it. The data files used for big data analysis can often contain individual private data. One of the recent hot topics in Chinese web is the leaks of Naked selfies of female students. The 10-gigabyte file, Thousands of photographs and videos of 161 naked female students used as collateral for loans on a Chinese online lending service have been leaked onto the web [17].

### **4.2 Privacy Violations**

With increasing storage and integration of personal information, privacy advocates are increasingly concerned. The White House report “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights” points that: “the algorithmic systems that turn data into information are not infallible--they rely on the imperfect inputs, logic, probability, and people who design them” [18]. Actually there are numerous big data analytics being used to mine personal details, in order to learn customer preferences and buying behavior. What's more terrible is that Today, privacy has become a commodity that can be bought and sold. Large commercial organizations who collect vast

amount of data on millions often leak or sell those private data. In August 2016, Xu Yuyu, a college-bound student, died after suffering a sudden cardiac arrest on Sunday after her tuition fees were swindled in a telephone scam. The suspects conceded that they easily brought large amount of personal information including Xu's by using online platforms, like Sina Weibo and Taobao. As Joseph W. Jerome argued that "however, the big challenge presented by big data is that the value may not be clear, the motives let alone the identity of the data collector may be hidden, and individual expectations may be confused" [19]. Today, under the present policy and legal framework, privacy violations are very difficult to solve. An ever increasing number of privacy violations is bringing us more risks, which may become the one of the major threats to Chinese big data industry.

### **4.3 The GFW**

The GFW(the Great Firewall of China), is becoming a significant obstacle to Chinese innovation. The GFW has blocked at least 18,000 websites are blocked from within mainland China like Google, Facebook, twitter, youtube and Wikipedia's Chinese site, according to a Harvard study [20]. From an economic point of view, the GFW can be seen as a new type of trade barrier. Local Chinese IT companies such as Baidu, Tencent and Alibaba, some of the world's largest Internet enterprises, benefited from the GFW which has blocked international rivals from the market, encouraging domestic competition [21]. To some extent, some successful IT companies are imitators of their international rivals.

Chinese government should know that the political targets of censorship could never be achieved because the GFW can be easily circumvented by many methods like using proxy servers or VPN. However, the GFW curbs the capability of innovation and creativity of Chinese IT industry. the GFW just blocks those western creative companies. The key words for today's innovation are open and collaboration. Big data is all about integration. The GFW transforms WWW to CWW (China Wide Web).

In addition to the above, other risks, such as deficiency of top-level design and overall planning, delay of legislation and policy making, incapability of innovation, government data sharing inefficiency, need to be solved or limited.

## **5. Conclusion: Big Data are the Next Big Challenge to China**

Obviously, big data are the next big challenge to China. Big data will be a new opportunity to reshape the competitive advantages of China. Although there are many obstacles we need to confront. The development and application of big data is the internal needs and inevitable choice to ensure steady growth, promoting reform, structural adjustment, improving people's livelihood and promoting the modernization of governance capability.

Gratifying, the State Council of China issued guidelines to promote the development of big data in Sep. 5, 2015. The guidelines aim to "forge a new model for social governance in the coming five to 10 years, highlighting accurate management and multi-dimensional cooperation"[22]. This can be seen a sign of the government's increasing emphasis on big data industry.

However, the guidelines are far from being desired. China should do more legislation and policy-making to build data sharing platform, boost industrial innovation, foster new business patterns, and speed up the priority researches like next generation technology of data storage and protection, annalistic tools of unstructured massive data like video and picture, machine reading and hearing technology and etc.

Last but not the most important one is that if China desires adhere to innovation driven development, China should loosen its censorship of Internet for stimulating more competitions to promote more innovations and creativities.

## References

- [1] [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)
- [2] Hilbert, M. Big Data for Development: A Review of Promises and Challenges [J]. *Development Policy Review*, 2016, 34(1): 135–174.
- [3] Latour, B. Tarde's Idea of Quantification [J]. In M. Candeia (Ed.), *The Social After Gabriel Tarde: Debates and Assessments* [M], London: Routledge, 2009: 145-162.
- [4] Mayer-Schonberger, V., Cukier K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (1st ed.) [M]. New York, New York: Houghton Mifflin Harcourt. 2013: 6
- [5] [https://en.wikipedia.org/wiki/Data\\_mining](https://en.wikipedia.org/wiki/Data_mining)
- [6] <https://www.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html>
- [7] <http://www.npr.org/2013/03/07/173176488/the-big-data-revolution-how-number-crunchers-can-predict-our-lives>
- [8] <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
- [9] <http://www.forbes.com/sites/andygreenberg/2013/08/14/agent-of-intelligence-how-a-deviant-philosopher-built-palantir-a-cia-funded-data-mining-juggernaut/#51898ba53da8>
- [10] <https://www.technologyreview.com/s/523646/the-power-to-decide/>
- [11] <http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Big-Data-Impacts-and-Benefits.aspx>
- [12] <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
- [13] <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- [14] <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- [15] <http://zt.360.cn/1101061855.php?dtid=1101062370&did=1101654296> (In Chinese)
- [16] <http://www.isc.org.cn/zxxz/xhdt/listinfo-33759.html> (In Chinese)
- [17] <http://finance.qq.com/a/20161130/036750.htm> (In Chinese)
- [18] [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf)
- [19] Jerome, J. Buying and Selling Privacy, Big Data's Different Burdens and Benefits [J]. *The Stanford Law Review*, 2013(66): 50.
- [20] <https://cyber.harvard.edu/filtering/china/>
- [21] <http://www.bbc.com/news/world-asia-china-29119121>
- [22] [http://www.china.org.cn/business/2015-09/05/content\\_36505087.htm](http://www.china.org.cn/business/2015-09/05/content_36505087.htm) (In Chinese)