

Development Status and Future Trends of China's Big Data

Guanchen Wu

Carey Business School, Johns Hopkins University, MD, USA

gwu31@jh.edu

Abstract. China has developed rapidly in recent years. Chinese President Xi also pointed out: "Big data is a new stage of informatization development." Therefore, promoting and innovating the big data technology industry and building a digital economy with data as the key element will be the main development direction in the next few years. In this article, I intend to review the origin & development, current status, and trend of big data in China, then analyze the future trend of big data and give some thoughts and suggestions.

Keywords: Big Data Development, Data Governance System, Privacy and Data Security, Digital Transformation in China.

1. Origin and Development

The concept and trend of thought, "Big Data," initially originated in the field of computer science before progressively expanding to other science disciplines and commerce. Scholars within the field of science date the origin of Big Data back to 1998 [1]. John Mashey, a chief scientist at Silicon Graphics (SGI), one of America's high-performance computing companies, pointed out some of the key challenges posed by the exponential data volume growth in one of the international conference reports [1]. Mashey reported these challenges as understanding, obtaining, processing, and organizing data. Therefore, "Big Data" was coined to describe these challenges and spark a new wave of thinking and understanding in computing. In 2007, Jim Gray, a pioneer in the database field, also predicted that Big Data would revolutionize how humans interact with complex systems [2]. Based on Gray, Big Data would usher in a new scientific research paradigm focusing on data exploration. Gray's concept was later refined by scholars in the industry as a data-intensive scientific discovery, thus marking a new beginning in the study of Big Data [2]. Therefore, the value of Big Data lies in its ability to provide insights and tools for understanding complex systems by creating digital representations of the real world.

2. Status and Trend

Globally, there is a growing interest in researching and developing Big Data technology to drive economic growth, enhance social governance, and improve government services [3]. In other words, applying Big Data technology to various facets of information analysis and presentation is a powerful tool for promoting innovation and efficiency in various sectors. In the following sections, I will explore the current status and trends of Big Data from both an application and governance perspective.

2.1. Levels of Big Data Application

The realm of big data applications has seen numerous successes. Nevertheless, its global understanding regarding impact and depth remains in its infancy. Although the future of big data remains in its ability to analyze and predict outcomes, it must guide decision-making processes. Big data applications can be categorized into three levels:

Level 1: descriptive analysis application summarizes and extract pertinent information from big data to help individuals analyze past events and understand their developmental history. For instance, DOMO Company in the US aggregates data from various enterprise information systems [4]. It presents it to business personnel and managers in visual formats such as statistical charts, aiding in informed decision-making.

Level 2: predictive analysis application, which focuses on analyzing relationships and patterns within big data to forecast future trends. For instance, David Rothschild, a researcher at the New York Research Institute in Microsoft, predicted 21 out of 24 Oscars in 2014 and 2015 by creating a prediction model based on data from the gambling market, the Hollywood Stock Exchange, and social media posts [5].

Level 3: instructive analysis application, analyze the concerns of several decisions based on the insights gained from the first two levels, ultimately guiding and optimizing decision-making processes. For instance, driverless cars utilize high-precision map data and real-time sensing data from sensors such as radar and cameras to analyze and predict the outcomes of various driving behaviors, guiding the automatic driving process.

2.2. Challenges in Big Data Governance

One major challenge in applying big data remains the lack of a comprehensive governance system. There are existing conflicts between the need for privacy protection, data security, and data sharing and utilization. These existing conflicts act as barriers to the advancement of big data technologies. To address this issue, all stakeholders must recognize the importance of establishing a robust governance framework, ongoing research, and practical efforts are being made. The need for a well-defined data governance system remains a major obstacle to its development as big data continues to gain prominence as a strategic resource [6]. The status of data assets has yet to be universally agreed upon, with debates surrounding concepts such as "The Metaverse" gaining traction. The Metaverse is defined as a network of 3D virtual worlds focused on social connections [7]. Furthermore, it is usually depicted in futurism and science fiction as a universal virtual world accessed through virtual and augmented reality headsets.

In Mark Zuckerberg's vision, Metaverse users are not simply connected to the internet but live within it by engaging in social interactions with the real world. However, challenges related to the authenticity, circulation, and data control within the Metaverse continue to pose challenges. For instance, data barriers hinder the sharing and accessibility of information. However, the slow progress in developing relevant laws and regulations exposes big data applications to security and privacy risks. These factors collectively impede the extraction and utilization of the valuable insights contained within data resources. The issue of balancing privacy, security, and data sharing is a significant challenge in today's digital landscape [7]. The need for open and shared data is increasingly pressing, as advancements in artificial intelligence rely heavily on analyzing vast amounts of high-quality data. However, individual organizations often need help to amass sufficient data. The true power of big data can be linked to integrating and analyzing data from multiple sources. It is also associated with how it allows for a comprehensive understanding of various perspectives [8]. Unfortunately, data within a single system or organization usually offers only a limited view of the overall picture. Thus, promoting data sharing and facilitating cross-domain data exchange is essential to gain a holistic understanding [8]. Promoting data sharing also helps to establish a more complete and accurate information set, enabling more informed decision-making and driving innovation in various fields.

China has protected personal information by developing and passing various legal documents to safeguard online personal data. These legal documents include The Decision of the Standing Committee of the National People's Congress on Strengthening the Protection of Network Information [9]. Other legal documents include The Provisions on the Protection of Users' Personal Information on

Telecommunications and the Internet, The Decision of the Standing Committee of the National People's Congress on Safeguarding Internet Security, and The Law on the Protection of Consumer Rights and Interests. In 2019, China's Cyberspace Administration introduced a Draft for Comments called Data Security Management [9]. Data Security management focused on gathering feedback from the public. It outlined standards and regulations for the collection, processing, use, and security supervision of personal data [9]. It was anticipated to play a crucial role in promoting the lawful use of data and safeguarding data security and personal privacy. Nevertheless, to ensure steadiness and systematization, there is a need for specialized legislation on data security and personal information protection. It is important to achieve a balance between development and security, as well as efficiency and risk, in order to avoid any negative impact on the extraction and utilization of big data value while maintaining safety. This remains a key challenge in data governance worldwide.

3. Trend of Big Data Development in China

China has a vast capacity for generating data, resulting in abundant big data resources despite being densely populated and industrially advanced. The ongoing efforts to build a digital China have played an important role in collecting and utilizing data resources across various industries. This has led to a rapid increase in data accumulation across different sectors. Projections indicated that China will contribute 21% of the global data volume by 2023 [10]. This solidifies its position as a leading data resource country and a global data hub.

The field of Internet big data in China is experiencing significant growth. Several Internet companies have established cutting-edge big data storage and processing platforms, making remarkable advancements in application areas such as mobile payment, network credit reporting, and e-commerce internationally. However, a significant gap must still be in integrating big data with the real economy [3]. This has resulted in a need for more breadth and depth in industrial big data applications. There is room for improvement in bridging the gap between data technology and real-world applications. Bridging this gap will harness the potential of big data to drive economic growth and innovation.

One of the challenges China's big data faces is the need for a governance system. Laws and regulations in China cannot match the rapid growth of data management needs. Therefore, China needs data management regulations that depend on data management and security laws [11]. However, the existing gap in the governance systems affects big data's ability to manage the increasing volume of data. Another challenge is a low level of openness in data sharing. This challenge limits big data's ability to break down barriers between different departments and systems. It also affects data circulations and the development of big data resources. China's lagging basic theory and core technology in the field of big data has led to a hollowing out of information technology in the country. It is crucial to address this issue and make advancements in basic theory, core devices, algorithms, software, and other key areas to avoid falling behind in the era of big data.

Murphy and colleagues believe there is a need for deeper integration of big data with the real economy in China [11]. Challenges include inadequate infrastructure, difficulties in data acquisition, lack of effective guidance and support for digital transformation, and a need for more independent and controllable data-sharing platforms. Addressing these issues will be crucial for integrating big data with the real economy and unlocking its full potential.

4. Solutions

4.1. Vigorously develop industry big data application

China's big data applications in the Internet sector are currently highly commercialized and well-established. However, the breadth and depth of industrial applications still need to be improved, and there is an urgent need to form and develop the ecosystem. Big data applications integrated with the real economy hold important development value and potential [3].

Using manufacturing as an example, a McKinsey research report indicates that manufacturing enterprises can minimize production costs by 10%-15% through big data technology [12]. Big data

technology on manufacturing is essential in extending beyond cost reduction. Various industries and sectors in China actively promote digital transformation, network restructuring, and intelligent upgrades to advance industrial big data applications. This is a crucial approach and foundation for promoting the construction of a digital China.

4.2. Establish a systematic and comprehensive big data governance system

Establishing a systematic and comprehensive big data governance system is crucial in the current landscape. Data sharing and openness serve as the foundation for constructing big data resources. Balancing open data sharing, privacy protection, and data security is imperative. Prioritizing application and security is key when developing a data governance system.

5. Conclusion

The concept of Big Data has continued to evolve since its origin in the late 1990s. Big data has developed from computer science challenges to a transformative force. It can now support several sectors, such as governance, commerce, and science. It has proved essential in improving economic growth, decision-making, and societal outcomes, which are key to influencing global interest. China is one of the countries that has made substantial steps in developing data technologies and governance frameworks. Despite the expansion of Big Data applications, increased challenges continue to face Big Data, especially based on its governance, privacy, data security, and integration into the real economy. Furthermore, there needs to be a comprehensive governance system to unlock Big Data's potential. This approach will positively influence big data because it enhances data sharing, thus creating conflicts between data utilization and privacy protection.

Good governance structures ensure data security is maintained to realize the true value of Big Data while promoting open data sharing. By addressing the challenges and gaps in the technological field, Big Data within industries will foster the integration of the real economy, innovation, and economic development. China's experience shows the importance of advancing technological capabilities and legal frameworks. This experience shows how big data shapes the dynamics of an effective Big Data ecosystem. This implies that future research and collaborations are essential in overcoming Big Data challenges based on the ever-changing global landscape.

References

- [1] Diebold, F. X. (2019). On the Origin(s) and Development of the Term “Big Data.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2152421>
- [2] Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). Big Data. In *SpringerBriefs in Computer Science*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-06245-7>
- [3] Murphy, B., Creemers, R., & Kania, E. (2021, October 14). Xi Jinping: 'Strive to become the world's Primary Center for Science and high ground for innovation.' DigiChina. <https://digichina.stanford.edu/work/xi-jinping-strive-to-become-the-worlds-primary-center-for-science-and-high-ground-for-innovation/>
- [4] DOMO. (2024). *Secil Centralizes and Leverages Disparate Data with Domo*. Domo.com. <https://www.domo.com/news/press/secil-centralizes-and-leverages-disparate-data-with-domo>
- [5] Wiens, S. (2017, December 17). Here are David Rothschild's data-powered predictions for 2014, from the Super Bowl to Hillary Clinton - The Official Microsoft Blog. The Official Microsoft Blog. <https://blogs.microsoft.com/blog/2013/12/17/here-are-david-rothschilds-data-powered-predictions-for-2014-from-the-super-bowl-to-hillary-clinton/>
- [6] Tang, B., Kang, L., Zhang, L., Guo, F., & He, H. (2021, January 5). Collaborative filtering recommendation using nonnegative matrix factorization in GPU-Accelerated Spark Platform. *Scientific Programming*. <https://www.hindawi.com/journals/sp/2021/8841133/>
- [7] Tai, K. (2019, May 31). *Translation: China's New Draft "Data Security Management Measures"* (L. Laskai, R. Creemers, M. Shi, K. Neville, & P. Triolo, Eds.). New America. <https://www.>

- newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinas-new-draft-data-security-management-measures/
- [8] Dionisio, J. D. N., III, W. G. B., & Gilbert, R. (2018). 3D Virtual worlds and the metaverse. *ACM Computing Surveys*, 45(3), 1–38. <https://doi.org/10.1145/2480741.2480751>
 - [9] Greenleaf, G. (2013). China's NPC Standing Committee Privacy Decision: A Small Step, Not a Great Leap Forward. *Privacy Laws & Business International Report*, (121).
 - [10] Sun Chi. (2024). *China's data industry is galloping toward a wider digital future*. Chinadaily.com.cn. <https://global.chinadaily.com.cn/a/202408/26/WS66cbe00fa31060630b924e7c.html>
 - [11] Mao, D. (2021, August). Analyzing Digital Transformation in a Selected Chinese Company. University of Ljubljana; <http://www.cek.ef.uni-lj.si/magister/mao4447-B.pdf>
 - [12] Espel, P. (2020, April 19). How industrial companies can cut their indirect costs—fast. *McKinsey & Company*. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-industrial-companies-can-cut-their-indirect-costs-fast>