2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI) | 978-1-7281-9753-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAACI50733.2020.00017

Application Scenarios and Practice Essence of Data Science Based on Big Data Analysis

Yanfang Zha
Suzhou Industrial Rark Institute of Services Outsourcing, Suzhou, Jiangsu, China
zhayanfang@163.com

***Abstract*—The synchronous development of urbanization, industrialization and informatization is an important characteristic of China's urbanization. At present, the construction of the major of data science and big data technology in China has become a new hot topic. The major of data science and big data technology has been added by the state in recent years against the background of "big data" era. The aim is to train high-level big data talents with big data thinking, big data thinking and analytical application technology. In view of the increasing importance of data science and big data specialty, there are different personnel training programs for data science and big data specialty. The major of data science and big data technology is a combination of software and hardware, based on computing technology and characterized by data science and big data technology. This paper analyzes the current situation and application scenarios of data science, and introduces the challenges and practical methods that big data brings to data analysis tools in data science.**

Keywords—Data science, Big Data Technology, Personnel training

I. Introduction

Since the reform and opening up, China's urbanization, industrialization and informatization have been developing rapidly. The development of the three are closely linked and promote each other. With the rapid development of information technology and economy, data is becoming a national basic strategic resource at an alarming rate. Big data has penetrated into all fields and industries, and has an important impact on global production, consumption, economic operation system, social lifestyle and national governance capacity [1]. Data has penetrated into various industries and business functional areas and has become an important factor of production. People's mining and application of massive data indicates the arrival of a new wave of productivity growth and consumer surplus [2]. With the development of information acquisition technology, communication networks and computer networks for fast and safe transmission of information have begun to take shape. More and more kinds of information are acquired in real time, showing the characteristics of "large amount", "high speed", "diversity" and "low value density" of big data [3]. The major of data science and big data technology is a wide-caliber major combining software and hardware, based on computing technology and featuring data science and big data technology. It is a core idea with big data analysis. Under the background of the evolution of the major global forces, the imbalanced development of traditional industries in the economic field, and the development of the big data industry, building a national big data system is a practical and feasible route to national security, social governance, and economic development [4].

Data science is an emerging cross-discipline that spans many fields such as information science, network science, and economics, and is still in the early stages of development. Judging from the latest development trend of artificial intelligence, the latest development of artificial intelligence reflects the new trend of the fusion development of artificial intelligence algorithms and big data. The knowledge bottleneck of traditional algorithms can be solved by learning knowledge from big data [5]. The main disciplines of data science and big data technology are computer science and technology and statistics. The goal is to systematically train students to master computer science knowledge and statistical principles, making them a decision control method that can use big data to predict and analyze specific application scenarios. Of high-level engineering talents [6]. Data science and big data technology majors have become the most popular new majors in China. However, how to construct the major according to local conditions is a common problem faced by all universities [7]. Data is rapidly expanding and growing, it determines the future development of the enterprise. Cultivating big data talents facing industries and fields is an urgent task entrusted to universities by the times. To do a good job in data science and big data technology majors, it is necessary for colleges and teachers to clearly understand the theoretical basis, methodology and educational laws and cognitive laws of the big data major [8]. Over time, people will become more and more aware of the importance of data to the enterprise. This paper analyzes the development status and application scenarios of the discipline of data science, and on this basis introduces the challenges and practical methods that big data brings to data analysis tools in data science.

II. Knowledge System of Data Science and Big Data Technology Specialty

As a very important aspect of data science, big data not only provides great opportunities for the development of science and education, but also brings great challenges to frontier science projects. With the proliferation of Internet users, mobile phones have been digitized and broadband, and data is growing at an exponential rate. The urban management process is the process of formulating urban planning and urban public policies, organizing the implementation of planning and policies, monitoring the implementation process, correcting the irrationalities of planning and policies, and evaluating the effects of planning and policies. Entering the era of big data, the amount of data

978-1-7281-9753-1/20/$31.00 ©2020 IEEE
DOI 10.1109/ICAACI50733.2020.00017

is increasing rapidly and the format of data is also diversified. For example, the data of banks and supermarkets will be in text format, and the data of digital phones will be in voice format [9]. Big data operation and maintenance is crucial to the support of data platform. The sources of big data mainly include data information generated by electronic equipment and Internet in the process of production and life and information of various perceived objects collected along with the construction of Internet of Things, digital city and smart city. Digital city, digital city management, intelligent city and other information technology-dependent urban construction and management programs have shown great advantages in improving urban management, and have been recognized by all levels of urban government. Information technology plays an increasingly obvious role in urban management.

If common data storage technology is applied to the storage of big data, it will cause a large amount of resource consumption. Therefore, it is necessary to combine the characteristics of big data and adopt new methods for big data storage. Ensure fast and stable storage of big data information. Figure 1 reshapes the structure of Internet financial supervision for big data.

Figure 1 Big data reshapes the structure of Internet financial supervision

At present, compared with other industries, big data applications are shaped by Internet enterprises, which hold a large amount of data assets. These Internet enterprises are eager to transform big data information into commercial value and improve user experience. Besides the traditional relational data, the data form also includes original, unstructured and semi-structured data from web pages, e-mails, social media forums, search indexes, internet log files, etc. By means of big data analysis, find the most suitable product features to meet the needs of users, thus guiding product design and development, and continuously tracking and analyzing users' online ordering and usage problems after the business goes online, thus providing data support for optimizing business strategies. Faced with such a large amount of data and various forms of data, data analysis tools are required to organically combine the methods of processing structured data with the methods of new unstructured data. Many applications of big data have the characteristics of real-time. In this case, the most important index of big data application is no longer the accuracy of the algorithm. The algorithm needs to strike a balance between accuracy and real-time. The data collection technology in the big data era is to continue to develop new intelligent identification technology based on the mature and practical traditional technology, and form a data collection technology system in which human and technical intelligence complement each other.

III. Thinking on Development of Data Science and Discipline Construction of Data Science

A. Selection of Practical Teaching Mode

The upsurge of big data has given birth to a new discipline, namely data science. Data science is in the early stage of development and is a continuously developing discipline. The main research objects of data science include data acquisition, data storage and calculation capabilities and the ability to extract data value, which are also the three major elements of big data application. The core of data science involves using automated methods to analyze massive data and extract knowledge from it. The scope of data collection is expanding, the quantity is increasing and the speed is getting faster, which puts forward new requirements for data transmission technology from the perspectives of transmission range and transmission speed. After continuous upgrading and evolution, computer big data has entered a new technology end and field, and even a special computer new technology has been available. The means and methods used by Internet education to help students learn are becoming more and more abundant. A common Internet learning model can be described as a "catalytic reaction", as shown in Figure 2.

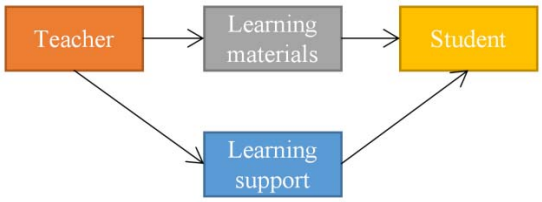

Figure 2 Catalytic reaction mode of Internet learning

In practical application, data science includes data collection, data cleaning, data analysis, data visualization and the whole iterative process of data application, which ultimately helps organizations make correct development decisions. The practitioners of data science are called data scientists. Data scientists are compound talents with broad vision. They not only have a solid foundation of data science, such as mathematics, statistics, computer science, but also have a wide range of business knowledge and experience. In almost all areas of knowledge discovery, data science provides a powerful new way to explore discovery, which provides a new source of insight for companies that have a large amount of data but do not know how to extract value from it. Data science involves the whole life cycle of data, including data collection, analysis management and visualization as well as its wide application in all walks of life. Structural data is a traditional form of data, which usually has structure before data. Semi structured data usually has data first and then structure, while unstructured data cannot form an obvious data structure.

B. Construction of Practical Teaching System

Big data collection mainly obtains massive structured, semi-structured and unstructured data through Internet of Things, social networks, mobile Internet and other channels, which can be accessed and interoperated simultaneously by tens of thousands of users. The curriculum system of the major of data science and big data technology includes general education curriculum module, basic subject curriculum module, professional core curriculum module and centralized practice curriculum module. In the big data

scenario, due to the explosive growth of data volume and the increased demand for storage capacity, hundreds of large servers cannot meet the data storage needs of an enterprise [11]. The source of big data is the basic data and operation data of each component of the city, which provides the basic guarantee for the overall analysis of the city system. As a new interdisciplinary subject in the era of big data, the main knowledge structure of data science and engineering comes from computer science, applied mathematics, information system and information management. Because the big data system needs to meet the needs of a large number of users at the same time, and to provide efficient services for the processing and analysis of big data sets. In the development of big data information system, we must meet the actual needs of the premise. See Table I for system delay data of different network parameters.

TABLE I SYSTEM DELAY DATA OF DIFFERENT NETWORK PARAMETERS

Network parameter	Minimum value	Maximum value	Median	Average value
0.6	5.94	6.89	6.42	6.59
0.8	3.38	7.26	5.51	4.82
0.1	1.87	6.82	4.76	5.18

Each service circuit set must pass as few communication nodes as possible in order to improve the compactness of the topology. Figure 3 shows the situation of original data and predicted data.

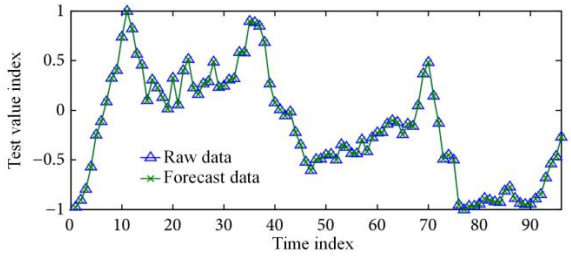

Figure 3 Raw data and forecast data

Traditional single-table data description cannot meet the needs of the big data era, so a distributed real-time data management system composed of multi-dimensional tables and oriented to column storage is needed to organize and manage data. The important characteristic of data science and engineering is its strong comprehensiveness, which requires students to learn through a comprehensive study and construct a basic system framework with professional knowledge and ability [12]. Data analysis is the most core part of the big data processing process. Analyzing and calculating the content of big data will discover the value of big data. The construction of practical teaching system requires various professional knowledge and practical experience, and at the same time follows the corresponding teaching rules. Data mining in big data scenarios requires that structured, semi-structured and unstructured data can be analyzed together. When the amount of data increases, only distributed service nodes are required to be added.

IV. CONCLUSION

In the era of big data, if data can be organized and used more effectively, people will have more opportunities to play the great role of science and technology in promoting social development. Although many enterprises may not realize the hidden dangers brought by the explosive growth of data, with the passage of time, more and more people realize the importance of data to enterprises. Big data technology integrates everything into digital codes, connects everything and transforms everything, providing an important foundation for the development of the Internet of Things and artificial intelligence. Colleges and universities should take the needs of the industry as the guide, the interdisciplinary basis as the support, adapt to the new era of young students to understand the laws of objective things, and design a scientific and forward-looking curriculum system to meet the needs of practical application. The construction of practical teaching system needs professional knowledge and practical experience in many aspects, and follows the corresponding teaching rules. Data mining in big data scenario requires the analysis of structured, semi-structured and unstructured data together. The future of big data needs to be shared by all countries. While exploring the broad development prospects of big data, we also need to pay close attention to data security and privacy protection.

REFERENCES

[1] Zhou Aoying, Qian Weining, Wang Changbo. Data Science and Engineering: Emerging Interdisciplinary Disciplines in the Era of Big Data. Big Data, no. 02, no. 90-99, 2015.

[2] Yang Jing, Wang Xiaoyue, Bai Rujiang, et al. Status and development trend of data science analysis tools in the context of big data. Information Theory and Practice, vol. 038, no.003, no.134-137, 144, 2015 .

[3] Zhu Yangyong, Xiong Yun. Training of data scientists in the era of big data. Big Data, vol. 000, no. 003, no. 106-112, 2016.

[4] Guo Yike. Big data; data science; data-driven scientific research. Bulletin of the Chinese Academy of Sciences, no. 6, no.599-607,2016.

[5] Liu Lei. From Data Science to the Fourth Paradigm: The Scientific Origin of Big Data Research. Advertising Concept: Theoretical Edition, vol. 000, no. 002, no. 44-52, 2016.

[6] Li Hui, Zhang Biao. Thinking on the cultivation of data science and big data technology professionals in agricultural universities. Higher Engineering Education Research, no. 5, no. 16-22, 2019.

[7] Yang Qianqian, Wang Long. An Analysis of the Talent Cultivation Model of Data Science and Big Data Technology Specialty in Independent College Undergraduate Colleges. Bohai Rim Economic Observation, vol. 000, no. 005, no. 116-117, 2019.

[8] Yang Yin, Huang Yunqing, Liu Shaoyue. Research on the training model of data science and big data technology professionals in local colleges and universities. Education Modernization, vol. 6 (04, no. 23-25, 2019.

[9] Wang Zhong, Shi Peibei. Research on Data Science and Big Data Technology Specialty Construction. Journal of Anqing Normal University (Natural Science Edition), vol. 25, no. 01, no. 123-126, 2019.

[10] Cui Chen, Yao Li. Research on the Course System Construction of Data Science and Big Data Technology in Public Security Colleges. Computer Age, vol. 312, no. 06, no. 92-94 + 97, 2018.

[11] Liu Minrui. Data Science and Engineering:Emerging Interdisciplinary Disciplines in the Era of Big Data. Smart City, vol. 004, no. 019, no. 123-124, 2018.

[12] Geng Xiurong. Guiding ideology and basic measures of data science and big data technology specialty construction. Journal of Mudanjiang Education College, vol. 193, no. 10, no. 22-23, 2018.