ORIGINAL ARTICLE

Research on the Integration of Urban Traffic and Big Data

Wangshu He[*]

Dano Information Consulting Limited. E-mail: wangshu@163.com

Abstract: The powerful data processing ability of big data technology can allocate traffic resources more efficiently, and can deal with various sudden traffic problems flexibly. It is an unprecedented opportunity and challenge for urban transportation and smart cities to effectively collect and utilize traffic big data to meet the application requirements of high timeliness traffic administrative supervision, traffic enterprise management and traffic citizen service. This article expounds the concept and characteristics of big data, discusses the application research of big data in urban transportation at home and abroad in recent years, summarizes its application research scope and trend, points out that intelligent transportation is the focus of the application of big data in urban transportation, and finally looks forward to the future research direction.

Keywords: Urban Traffic; Big Data; Intelligent Transportation

Introduction

With the development of mobile phone network, global positioning system (GPS)/ Beidou vehicle navigation, car networking and traffic Internet of Things, the information of people, cars and roads of traffic elements can be collected in real time, and the sources of urban traffic big data are increasingly abundant. By using big data technology, cross-regional and cross-platform data resources sharing can be realized, and its combination efficiency and information integration advantages can be brought into full play to create a comprehensive and perfect traffic information service system. Meanwhile, traffic resources can be allocated more efficiently with powerful data processing capability, and various sudden traffic problems can be dealt with flexibly. At the same time, the research on the integration and analysis technology of urban traffic big data is of strategic significance to the development and exhibition of smart cities in China. Traffic big data has many features, such as va-riety, heterogeneity, large span of time and space scale, dynamic variability, high randomness, locality and limited life cycle, etc. It is an unprecedented development opportunity and challenge for all large and medium-sized cities to effectively collect large data for communication and meet the intelligent demand of urban traffic, such as high time efficiency and knowledge traction.

1. Overview of big data

There is no unified and authoritative definition of big data at present. Its definition is written in Baidu Encyclopedia: a huge collection of data that cannot be captured, managed and processed with conventional software tools within an affordable time frame. Viktor Mayer-Schonberger believes that big data is a data set composed of a large number of data with many types and complex structure. It is an intellectual resource and knowledge service ability formed by data integration, sharing and cross-reuse instead of analyzing some data

by random analysis. It is a data processing and application mode based on cloud computing. Tu Zipei holds that the "greatness" of "big data" lies not only in the amount of information, but also in the increasing number of data that human beings can "analyze and use". Through the exchange, integration and analysis of data, new knowledge can be discovered and new value created, thus bringing "great knowledge", "great technology", "great development" and "great profit". According to the definitions elaborated by the above experts, although the descriptions are different, one thing is the same, that is, big data is a data set with many kinds of data and types, but it can quickly obtain information from it. "4V" is usually used to describe the characteristics of big data:

(1) The data Volume is large, and the storage unit of general large data sets is TB level, while the data of big data are mostly PB level, even reaching EB level or even ZB level;

(2) The data processing speed is fast. Big data is different from traditional data. It uses different mining technologies, and data processing does not need to consume warehousing, so the processing speed is very fast;

(3) Data types (Variety). Big data includes not only traditional structured data, but also semi-structured data and unstructured data, such as weblogs, pictures, videos, audio, geographic location information, etc. The diversity of data increases the complexity of big data, and at the same time puts forward higher requirements for data processing capabilities;

(4) The Value density is low. Take video as an example. In an hour of video, during the uninterrupted monitoring process, the data that may be useful is only one or two seconds. The characteristics of big data reveal the difficulty of its management, and also contain considerable value.

2. Development status and trends of big data

In recent years, the rapid growth of data has become a severe challenge and precious opportunity faced by many industries, and the Information Society is entering the era of large data. Large data refers to a data set whose data volume is so large that it cannot be perceived, retrieved, managed, handled and served within a reasonable time by the current mainstream software tools. From around 2009, "big data" began to become a popular word in the Internet information technology industry. The connotation and value of data not only bring great development opportunities for enterprises, but also have a very positive impact on improving the management and service level of government departments and better serving the general public and enterprises. With the rapid popularization of technologies and equipment such as Internet of Things, Internet of Vehicles, and sensors, what is worrying is not the lack of transportation data, but the information overload caused by exponential growth of traffic data. In the face of massive multi-source traffic data, it is necessary to use big data technology to collect, analyze and apply data efficiently and at low cost. At the same time, there are a series of challenges in the development of big data traffic, such as data security, privacy leakage, lack of effectiveness of intelligent algorithms and application models.

2.1 Data analysis platform based on cloud computing

Cloud computing provides flexible and relatively cheap storage space and computing resources for big data, which enables small and medium-sized enterprises to complete big data analysis through cloud computing like Amazon. Cloud computing IT resources are huge and widely distributed, which is a powerful and even the only way for enterprises with many heterogeneous systems to process data timely and accurately. Big data moving towards cloud computing depends on the improvement of data communication bandwidth and the construction of cloud resource pool. It is necessary to ensure that the original data can be migrated to the cloud computing environment and the resource pool can be flexibly expanded as needed.

2.2 Expanding data analysis and the enterprise-level data warehouse becoming the mainstream

When people taste the sweetness from big data analysis, the data analysis set will gradually expand. At present, the amount of data analyzed by most enterprises is generally in TB. According to the current data development speed, the data volume will soon enter the PB era. Especially, the number of analysis data sets in the range of 100~500 TB and 500+TB will multiply. With the expansion of data analysis set, the previous depart-

ment-level data sets will not meet the needs of big data analysis, and they will become a subset of enterprise-level database (EDW). Therefore, data analysis within an enterprise will transition from departmental level to enterprise level, and from departmental demand to enterprise demand, which will surely gain more benefits than departmental perspective. With the opening of government and industry data, more external data will enter the enterprise data warehouse, which makes the data warehouse larger and the value of data greater.

2.3 Hadoop: less dependent on MapReduce

Hadoop is a software framework that can handle a large amount of data in a distributed way, and can handle P B-level data. Hadoop has the characteristics of high reliability, high scalability, high efficiency and high error tolerance. Its new edition not only serves MapReduce, but also replaces MapReduce with a SQL query engine or other methods like Impala of Cloudera. HBase NoSQL database is a good example of Hadoop leaving MapReduce constraint. Hadoop platform will play an increasingly important role in big data processing in the future.

3. Opportunities

3.1 Problems and challenges of modern transportation

Since the reform and opening up, China's industrialization and urbanization have advanced by leaps and bounds, which has greatly improved the overall traffic situation of the society, but also brought new pain points and challenges to the development of the transportation industry? According to incomplete statistics, by the end of 2017, the number of motor vehicles in China has reached 310 million, and about 33% of the urban congestion index has continued to rise; In civil aviation transportation, the punctuality rate of flights is less than 70%. Therefore, how to provide convenient, comfortable and efficient transportation experience has become an important livelihood issue of common concern to the general public and the government.

3.2 Big data processing technology

Big data preprocessing technology is to further access the data of the platform according to specific business rules.

Processing, including checking the validity of accessed data, cleaning big data, etc. Big data standardization processing technology takes out the cleaned data from the database, and converts the data format of the external system into the standard format defined by the platform according to the business rules.

3.3 Big data fusion processing technology

Big data fusion processing technology refers to the use of multi-source traffic information fusion method, combined with feature fusion technology (recognition/classification, neural network, Bayesian network, etc.), target maneuvering information processing technology (adaptive noise model, etc.) and multi-target tracking information fusion technology to improve the robustness and reliability of information systems. Multi-source traffic big data information fusion is divided into three levels: the basic level is data-level fusion, which only completes data preprocessing and simple association; The second level is feature-level fusion, which predicts traffic parameters according to the characteristics of existing data; The third level is state level fusion, which judges the traffic state according to the current traffic flow information. The basic process of traffic flow information fusion includes multi-source information extraction, information preprocessing, fusion processing, target parameter acquisition and state estimation.

3.4 Intelligent application system of traffic big data

The intelligent application system of traffic big data is an intelligent application system based on traffic big data center and traffic cloud computing support platform. It adopts the mode of "central data storage and processing" and "local service application", captures real-time data from massive traffic data, analyzes and mines historical data, predicts the future situation based on historical data, and provides decision-making suggestions for intelligent transportation. Traffic big data intelligent application system can provide intelligent traffic information services for government, enterprises and the public respectively. The system can provide traffic administrative supervision support for government departments, mainly providing fine geographic information services, traffic management services, emergency re-

sponse services, roadside parking space supervision services, public transportation supervision services, etc.; Traffic information services based on mobile phone applications can be built for the public, and data of public daily travel behavior can also be collected through traffic information services, mainly including fine geographic information services, accurate real-time road condition services, accurate traffic information services, real-time vehicle information services, traffic guidance information services, and parking guidance information services; Provide traffic information value-added services for enterprises, mainly including fine geographic information services, bus company vehicle scheduling and auxiliary decision-making, commercial data analysis, etc. Different users can share industry data, computing resources and personalized intelligence analysis results, which has outstanding advantages in data collection and sharing, large-scale data real-time processing and analysis, and enterprise emergency handling, thus greatly saving system resources and costs and improving work efficiency. Technologies used in the system mainly include multi-source heterogeneous traffic information fusion technology based on decision tree-support vector machine (DT M -S V M), traffic information basic data service design based on SOA, ZigBee wireless sensor network technology, traffic information application service design based on mobile Internet, travel time prediction based on machine learning, pedestrian traffic information service technology based on location based service (LBS), etc.

4. Conclusion

Remote sensing spatial data, traffic visual frequency data, all kinds of perception data, public opinion data and other information are scattered and used separately in the urban transportation system. Generally speaking, the amount of data is huge. However, due to the single application and insufficient deepening of data in various aspects, the value of data cannot be fully tapped. On the basis of these data, building a big data platform for urban intelligent transportation can ensure timely updating of data in terms of scale, accuracy and time equality. Besides, it can also realize a more detailed and timely dynamic monitoring service data system.

References

1. Zhao G. The characteristics and prospects of the integrated development of big data and transportation (in Chinese). Macroeconomic Management 2018; (8): 60–67.
2. Luo S, Pan F, Wang X, *et al*. Research on application of big data in urban transportation. Modern Transportation Technology 2016; 13(5): 76–80.
3. Xiong G, Dong X, Zhu F, *et al*. Big data technologies and intelligent application system for urban transportation. Big Data Research 2015; 1(4): 81–96.