Research on the Large Data of Road Traffic and its key Technologies

Brooklyn Joseph Beckham

California American University, NewYork, 12201, USA

Abstract: Firstly, this paper summarizes the connotation, characteristics and research status of big data. Secondly, according to the situation of the people, vehicles, roads and environment in the road traffic, this paper analyzes the sources of the big data. Next, according to the road traffic information acquisition, transmission, processing and application, this paper analyzes the key problems of road traffic data. Finally, from the point of view of data processing and management, this paper analyzes the supporting technology of large data processing and management.

Keywords: Information technology; Large data; Large data processing and management; Road traffic; Traffic information

1. Introduction

With the rapid growth of the amount of data generated by the industry, the industry's requirements for data processing and analysis capabilities are also increasing. How to face the challenges of data and find new opportunities have been widely recognized. Since 2012, big data has become a hot research. There is no uniform definition of big data. Common definition is: big data refers to the amount of data involved in the large scale can not be processed through the current mainstream software. From the definition, big data is not a new product and technology, but only a phenomenon in the digital process. For the characteristics of large data, a more unified understanding is "3V", that is: large scale (volume), type (variety) and rapid change (velocity). In the field of road traffic, the traffic information system has accumulated a large amount of data. At the same time, the extensive application of information technology in the field of road traffic has accelerated the accumulation of data, and the emergence and application of the vehicle network has enriched the content of the data. The development of the road traffic in the data environment presents the characteristics of large data. International traffic data research features are: IntelliDrive program and University of Washington e-Science Transpor-Tation[1-2]. In China, China Intelligent Transportation Association believes that: Based on large data analysis to establish a new intelligent transportation information service system is a priority. In short, the domestic and foreign research on traffic data is still in the initial stage, while traffic data will become an important trend in the field of data accumulation.

2. Sources of Road Traffic Data

From the point of view of the subject classification, the road traffic data belongs to the category of traffic information engineering and control. In the field of road traffic, the subjects of traffic information engineering and control are human, vehicle, road and environment. Road traffic data is also a natural person, car, road and environment, the 4 elements of the data generated.

2.1. Data Drom Human

Refers to the people involved in road traffic, including drivers, traffic managers, passengers and pedestrians. According to the attributes of traffic data, the data of traffic participants are divided into static data and dynamic data.

In the process of road traffic operation and management, the static data is involved, and the dynamic data is semi-structured or unstructured. The amount of static and dynamic data will increase with the increase of the number of people involved in the traffic.

2.2. Data From Car

Road traffic is mainly refers to motor vehicles. The data derived from the vehicle include: static data generated during the vehicle life cycle and dynamic data generated during the vehicle running. The static data of vehicle is structured data, text type semi-structured data and a small amount of unstructured data.

Similarly, with the increase in the number of vehicles involved in traffic, the data from the car will be a lot of growth. Therefore, the data derived from the car also has the characteristics of large data 3V.

2.3. Data From Road

Road in the road is the highway and urban road. The data from the road are: the road attribute data (static data) and the use of road generated data (dynamic data).

Road static data refers to the morphological attributes of the road. Road dynamic data is the road performance and real-time road conditions. The road form attribute data includes structured data and unstructured or semi-structured data derived from the design unit. The dynamic data show the characteristics of real-time change and dynamic acquisition.

Therefore, the data derived from the road has the 3V features of large data.

3. Key Problems and Supporting Technology of Large Data Processing in Road Traffic

The research task of traffic information engineering and control is the collection, transmission, processing and comprehensive utilization of traffic information. From the application point of view, after the formation of the road traffic data environment, integration, storage, management, processing and mining data and the formation of road traffic management and use of knowledge has become a priority. Therefore, the storage, management, processing and application of road traffic data is the key problem in the study of road traffic data.

3.1. Key Problems of Large Data Processing in Road Traffic

For the road traffic data storage, management and processing of the following issues need attention. [3]

New data representation. At present, the representation of the data may not be able to intuitively show the meaning of the data. If you need to use the data effectively and mine knowledge to improve intelligence, you need to find the most appropriate way to represent data. If you want to find the big data from this kind of inappropriate data in the model, the correlation and causal relationship will be due to the data representation of the reasons for the fixed conclusion. The manner in which data is represented is closely related to the original data entry. If the original data is identified, it will help to reduce the difficulty of subsequent data analysis, classification and classification. At the same time, in order to identify data, which will give the user additional work, and is not conducive to user use. Therefore, it is necessary to find an effective and simple way to represent data.

Data integration. Data integration can not play the value of big data. One of the most important challenges facing big data is whether individuals, businesses and service providers can easily integrate a variety of data. There are many kinds of natural language in human society, and the corresponding data in the information space can be expressed in many formats. In order to overcome the diversity of data formats in large data processing, we need to study and apply platform independent data representation. Big data has become the social, physical world and the information world link, which requires the use of a unified data sheet three space information systems.

3.2. Basic Software Supporting Technology for Large Data Processing of Road Traffic

Road traffic data has the characteristics of 3V, so the traditional data processing technology is not suitable for large data processing. MapReduce is a model for parallel processing and generating large data sets. Hadoop is a big data processing technology of MapReduce enterprise and academia.

NoSQL database. NoSQL database is a new technology which can not deal with unstructured data, difficult to expand laterally, and limit the scalability of RDBMS. This is part of the relational database management system (RDBMS) is not good at the supplement.

The main difference between the NoSQL database and the traditional RDBMS is shown in table 1[7].

Table 1. The Difference Between RDBMS and NoSQL Database

	RDBMS	NoSQL
Data type	Structured data	Mainly unstructured data
Database structure	Need to be defined in advance	No need to define
Data consistency	Maintain strict consistency	The state of being temporarily inconsistent
Expansibility	Upward expansion	Lateral spread
The server	On the premise of working on a server	Based on the distributed and cooperative work
Fault tolerance	In order to improve fault tolerance requires high cost	Low cost
Query language	SQL	Support a variety of non SQL language
Data quantity	Smaller scale data	Large scale data

4. Possible Future Applications of Road Traffic Data

The big data environment of road traffic has greatly enriched the data of people, vehicles, roads and environment in the field of road traffic. The big data environment for data sharing, a wide range of intelligent

applications to provide a data base, but also to further develop intelligent transportation systems. Therefore, this section analyzes the future possible applications of road traffic data from the subsystem of intelligent transportation systems.

Traffic information service system is designed to provide timely, accurate and relevant travel information. Road traffic data processing technology for timely and accurate traffic information services to provide technical support.

Road traffic data can be completed a wide range of road traffic information to make full use of and widely shared, therefore, real-time analysis, traffic road traffic under the big data environment short range road road traffic prediction based on possible.

5. Conclusion

This paper analyzes the research status of data traffic and road connotation and characteristics of big data at home and abroad, and analyzes the sources of traffic data, also analyzed the road traffic data processing and possible problems. This paper tries to sort out the basic

technical support for the research and development of intelligent application of road traffic data. The future will be selected from the application of the road traffic data may choose the typical application of in-depth research, and the establishment of large road traffic data system.

References

[1]   Li Fang. Research and Innovative Technology Administration, ITS Joint Program Office[J].AASHTO Connected Vehicle Infra structure Deployment Analysis，2011(1):23-33.

[2]   Qiao jun. Federal Highway Administration Office of Operations Research and Devcl-opmcnt[J].Data Capture and Management State of the Practice Assessment and innovations Scan Overview. 2011(2215):3 7-49.

[3]   Liu Peng. Cloud computing (Second Edition) [M]. Beijing: Publishing House of electronics industry, 2012.

[4]   Li Guojie. The scientific value of large data research [J]. Chinese society of computer communication, 2012 (9): 8-15.

[5]   Li Jianxin, Hu Chunming. The challenge and thinking of big data science and engineering [J]. Communication of China computer society, 2012 (9):22-28.