Urban smart transportation based on big data

Mai Liu
Computer Science and Technology, Tianjin University of Technology (TUT), 300384, Tianjin, China

Abstract. Traffic congestion has become more and more serious in the last decade. Because of the backward transportation system, it is impossible to analyze the large amount of feedback data. At the present stage, the development trend of big data is gradually emerging, and traffic problems are taken as the starting point. Therefore, it is of great significance to study the methods of big data to improve traffic congestion. In this paper, the development status of intelligent transportation based on big data is mainly described. In combination with Hadoop technology, positioning data acquired by on-board terminals is taken as an example for analysis, and the driving radius and driving distance map are experimented. It is of great significance to study the application of big data platform based on Hadoop technology in the transportation industry. A large number of data resources and low-cost big data technology can be effectively utilized to promote the development of transportation, thus promoting social development and progress.

1. Introduction
In recent years, the development of information and communication technologies (ICTs) has made it technically and economically feasible to collect huge amounts of spatiotemporal data about human movement. Typical spatiotemporal big data include mobile phone data, smart card data, taxi trajectories, and social media data (Chen et al. 2016; de Bruijn et al. 2018; Kwan, 2016). These spatiotemporal big data record the everyday movement of a large number of individuals with unprecedented spatiotemporal resolutions. Such big data allow us to better understand the activity-travel behaviors of individuals from different social groups and their interactions with the transportation environment and transport policy, therefore providing new opportunities for developing advanced models and algorithms to make various transportation systems smarter and safer (Chen et al. 2018; Kwan2012, 2018; Ma, He, and Zhang 2018).

With the continuous improvement of China's economy, the number of vehicles in urban transport networks has increased significantly. This change has brought tremendous pressure on our urban transport. In order to improve the situation, urban intelligent transportation system with intelligent features has been gradually used in traffic management. In order to optimize the system's functions and improve the efficiency of system data processing, big data technology can be introduced into the urban intelligent transportation system.

At the same time, the massive data include all kinds of traffic monitoring. Such as roads and other video monitoring data, traffic flow detection of city road and highway, meteorological data, urban public transport and vehicle satellite positioning data, etc., these types of traffic data are numerous and huge. Through market research and analysis, there are some systematic products for traffic management both at home and abroad, but there are still some problems, such as single system function, lack of integration and backward technology, and that is mainly reflected in the construction of application system distributed mass data, lack of effective integration of traffic data, low utilization...
rate, the data value cannot be brought into full play, and limited, traffic information dissemination is difficult to timely access to traffic warning etc. With the development of information technology, traffic departments urgently need a more advanced intelligent data analysis method, in order to carry on the efficient, real-time analysis.

2. **Big Data Technology**

2.1 *The value of big data technology*

The value of Big data technology mainly included the following: First, through the integrated analysis of massive data, it can solve the problems and analysis the reasons timely to reduce the cost of failure; second, to provide reliable data base for actual analysis by using the data mining technology to analysis the correlation between the huge amounts of data (such as, product marketing, criminal case investigation).

2.2 *The characteristics of big data technology*

In fact, the characteristics of big data technology include the following: (1) High-speed processing characteristic. Big data technology data processing speed fits the 1 second law. This characteristic effectively solved the pressure of traditional processing methods to deal with massive data processing tasks. (2) Volume characteristic, traditional data volume level is TB level, while the big data technology has developed to the PB level directly. (3) Low-value-density, this feature is mainly for the production of useful data in terms of probability. Take the process of screening video by public security agencies as an example. In video content that lasted for hours, the useful data may only last for a short time of 1-5 seconds.

3. **The City Intelligent Transportation System Technology Structure Analysis**

The technology and structure of urban intelligent transport system are as follows:

3.1 *The Key Technologies in Urban Intelligent Transportation System*

From an overall point of view, the key technologies in urban intelligent transportation system mainly include the following:

1) **GPS Technology**

In the city intelligent communication system, the function of GPS technology is mainly to provide users with precise positioning function to meet the real-time navigation needs of users. From the application effect point of view, the GPS technology provided a good data base for relieving the pressure of urban traffic.

2) **GIS Technology**

In actual use, GIS technology can be based on the actual needs of users, with the help of GPS technology to convert the relevant geographic information into an intuitive form and feed it back to users, so that users can make appropriate decisions and avoid prolonged user communication due to traffic congestion.

3) **Communication Technology**

To some extent, the process of using the city intelligent transportation system can be regarded as the process of data transmission between system and user through system channel.

3.2 *The Analysis of Intelligent Transportation System Structure*

In the design of urban intelligent transportation system structure, except for the reference to relevant standards, the practical value of urban intelligent transportation system should be improved in combination with the actual use demand of Chinese cities [1]. The feasible urban intelligent transportation system structural elements design patterns are as follows:

(1) The central system elements. In the urban intelligent transportation system, this element has various functions such as charging management, traffic management travel management and so on.
The remote system elements. The main function of this element is to provide real data support for users in remote areas. Its functions include remote travel support function and personal information access function. (3) The elements of the road system. In actual use, this part is mainly responsible for inspecting various commercial vehicles, assessing the road conditions, and carrying out parking management. (4) The vehicle system elements. The function of this part mainly aimed at the data acquisition and information management of different vehicles such as buses, emergency rescue vehicles and commercial vehicles.

4. The Application of Big Data Technology in Urban Intelligent Transportation System

4.1 The Practical Application Demand of Urban Intelligent Transportation System Based on Big Data

When applying big data technology in urban intelligent transportation system, it should be applied reasonably according to the actual application requirements. Through the analysis we can see that the actual application needs of urban intelligent transportation system mainly include the following:

1) Inclusive Needs and MEET Traffic Requests Demand

The inclusive demand of urban intelligent transportation system is basically the same as the basic data processing tasks of big data technology. In addition, in order to ensure the satisfaction of users, diversified traffic requests proposed by users should be satisfied in the process of constructing urban intelligent transportation system using big data technology.

2) Modular Needs

With the increasing number of commercial vehicles and the continuous development of urban transportation network, users have also made some changes in the functional requirements of the urban intelligent transportation system. Therefore, the structure of intelligent transportation system should have good modular characteristics in order to use different modules to play different functions.

4.2 Big Data Technology Application Function

The functions of urban intelligent transportation system based on big data technology mainly include the following:

1) Massive Data Acquisition Function

Refer to the current situation of traffic in China's major cities: In order to relieve the traffic pressure and improve the quality of urban traffic management, the number of information gathering devices such as video monitoring devices and positioning devices in urban traffic network has increased significantly.

Take the data of bayonet electric police in a city of Northeast China as an example, the number of lanes in this city is 1,300 in total, and the peak value of daily passing data for each lane can reach 12,000. In the background of massive data, PB has become the basic level of magnitude for which data are collected by urban intelligent transportation systems. Therefore, it is necessary to collect real-time traffic data information from different lanes and different devices through the application of big data technology to lay a foundation for subsequent data analysis and data integration.

2) Mass Data Computing Capabilities

For massive data generated in urban intelligent transportation system, traditional data processing technology is difficult handle massive data, providing users with accurate traffic navigation and other functions. In contrast, the application of big data technology can make use of cloud computing cluster, through the distributed way to complete the massive data high-speed calculation.

In the urban intelligent transportation system, the massive data calculation process of big data technology is: the distributed computing system based on the M/S architecture, the mass data is classified, decomposed and scheduled by the dispatching service manager Master. At the same time, the slave of the basic unit directly completes the calculation task from the Master, and while accurately calculating the massive data, feeds back accurate data calculation results to the Master and
provides the urban intelligent transportation system with reasonable traffic planning or guidance and decision-making.

3) Massive Data Retrieval Function

The massive data retrieval mechanism of urban intelligent transportation system based on Big data technology is as follows: referring to the characteristics of the industry data query and the actual traffic data usage requirements of the users, and customizing the search engine of the intelligent transportation system, and using the Big data technology to optimize the query speed of the system, Making the city intelligent transportation system can support users record 10 billion high-speed query[2]. In addition, in order to ensure the timeliness of the massive data retrieval functions, the cluster mechanism is also used to enhance the scalability and fault tolerance of the system search function and meet the personalized data retrieval needs of the users so as to enhance the user's satisfaction with the urban intelligent transportation system.

4) Massive Data Value-Added Applications

In addition to meeting the functional needs of commercial vehicle users such as traffic navigation and traffic congestion avoidance, in urban intelligent transportation systems, it is also possible to use big data technology to provide value-added services to enhance the practical value of urban intelligent transportation systems.

Taking the actual situation of the public security organs in analyzing cases and chasing suspects as an example, after introducing Big data technology into the urban intelligent transportation system, the intelligent transportation system can input relevant information such as vehicle license plate number, model, etc., and the intelligent transportation system Internal case clustering model, vehicle trajectory model and each lane traffic model, data mining integration, and then quickly for the public security agencies to provide accurate vehicle trajectory collision results, intelligent tracking results and other key information to enhance the actual efficiency of public security. In addition, in the urban intelligent transportation system, big data technology can also be used to gain value-added functions such as face comparison and intelligent decks, and to provide reliable technical support for the development of other industries while improving the intelligence of traffic management.

5) Massive Data Storage Function

The application of big data can provide high quality data storage function for urban intelligent transportation system in the face of massive data information from traffic data collection devices in various lanes. Compared with the traditional relational database, big data technology used the Hbase distributed storage structure to process different formats of data information flexibly (such as unstructured data, structured data, etc.), and provide convenient traffic data storage and access functions for users according to the rapid horizontal expansion of data,. In actual use, if the user needs to use some stored data information, the required historical traffic data information can be quickly derived from the database[3]. In essence, the massive data storage function based on big data technology has the characteristics of high performance and high fault tolerance.

In addition, compared with the traditional data storage capabilities, large data-based mass data storage features the application of advantages mainly reflected in: This feature can store a large number of different types of data and information simultaneously, and to protect the security of data to meet the user's data Query, retrieve functional requirements.

4.3 Big Data Technology Application form

Specifically, the application forms of Big data technology in urban intelligent transportation system mainly include the following:

1) Data Sharing form

In essence, the data sharing function of urban intelligent transportation system is based on cloud computing of big data technology. In the actual work of traffic management, after all levels of management departments introduced the urban intelligent transportation system, the big data technology collected, integrated and analyzed the massive data through the integrated form. According to the analysis results of traffic data information, a reasonable allocation plan of transportation public
resources is established. With the operation of the distribution scheme, the management departments at all levels can obtain the status of all areas and lanes, and formulate feasible traffic management schemes according to current traffic changes.

2) Traffic System Perception Form

Uses the urban intelligence transportation system to promote the traffic management quality undoubtedly is an extremely complex function. Among them, the perceived ability directly affects the practicality of the intelligent transportation system. In order to enhance the perceived ability of ITS, sensors can be added to all lanes and some of the vehicles in the urban transport network respectively. After obtaining the traffic data, the sensors can be transmitted to ITS database. With the help of big data technology analysis technology and data mining Technology, access to user needs information[4]. For example, when a user uses a city intelligent transportation system to check whether a route is congested, the system's sensing ability and data analysis and mining functions are simultaneously linked together to promptly provide the user instant traffic conditions for the user to make reasonable traffic decisions.

3) Formulation of Transport Options

The key to improving the quality of traffic management is to fully integrate the big data technology with the urban intelligent transportation system so that the urban intelligent transportation system has the function of formulating the transportation plan in real time. The intelligent transportation system obtains the real-time traffic information of relevant lines through the sensor, provides several different versions of travel plans according to the user's request, and the user can decide which scheme to use for traveling.

From the application effect point of view, the application of big data technology effectively enhances the data processing speed of the urban intelligent transportation system, effectively promotes the improvement of data usage efficiency, and provides technical support for the urban traffic safety.

5. Experimental design of transportation industry application research based on Hadoop

5.1 Core content of Hadoop Platform

1) MapReduce

MapReduce originated from Google's MapReduce paper, which was published as early as December 2004. MapReduce has an excellent horizontal scaling cluster architecture. MapReduce has the following characteristics: excellent scalability, excellent fault tolerance, suitable for offline processing of massive data.

During the operation of MapReduce, the files already stored in HDFS are first read by the tasks run by Mapper. Call your own way, process the data, and then output it. Secondly, the Reduce task will receive the data output by the Mapper task and use it as its own input data. In the same way, it will call its own methods to process the data and finally output it to the HDFS file. In layman's terms, a MapReduce job typically divides input data files into many equally large, independent data blocks according to fixed criteria, which are processed in full parallel by the Map task. MapReduce distributed computing framework sorts the output of Map first [10], and then inputs the results to Reduce, which processes the data again.

2) HDFS

When the amount of data is too large to be stored in one operating system, it is allocated to more operating system disks, but this is inconvenient, so there is an urgent need for a versatile system to manage the files on several machines. This is why the distributed file management system HDFS came into being. HDFS is very powerful and can be used to store big data. We can analyze how it works by cutting files into equal chunks and storing them on different computers. Data segmentation is fault-tolerant and can also be used for load balancing. Multiple copies of HDFS data are automatically saved. The high reliability of HDFS is due to the following characteristics: excellent scalability, high fault tolerance, suitable for archiving sea volume data. HDFS is developed in the Java language. Because the Java language is so portable, HDFS can be deployed on many types of machines. The key
performance of HDFS is copy storage. Improved copy storage is not available on most distributed file systems and is an important feature of HDFS. In HDFS, files can only be written once, and it is strictly required that there is only one writer. An HDFS is composed of a Namenode and a number of Datanodes. The Namenode is a central server that provides performance support for managing file system namespaces and client access to files. A Datanode is typically one for anode and is responsible for storage on the node where it is located. HDFS displays the name space of the file system, and users can store data in HDFS as a file. In this paper, the data stored in HDFS space is in the form of a file. And during the operation, there can only be one write at a time. The data is first uploaded to the HDFS space and then calculated. HDFS will have a browser interface that allows you to view the contents of files in HDFS directly from the browser.

3) HBase

HBase is Hadoop's database, ready to provide real-time access to large data reads and writes. Is one of the open source implementations of Google's Big Table. HBase exists for the purpose of storing and processing large amounts of data, and more specifically for large databases with thousands of rows and columns that can be configured with ordinary hardware.

HBase is open source and has many versions. HBase is a columns-oriented storage model. It can use a local file system or it can take advantage of HDFS storage. However, in order to improve the reliability of the data, the robustness of the system, and to highlight HBase's ability to handle large data, it is better to use HDFS to store folders. On the other hand, HBase stores loose data, specifically, HBase stores data that is somewhere between mapped and relational data. The data stored in HBase is logically analyzed as a large table, and its data columns can be dynamically increased as needed.

HBase gets its name from Hadoop Database, or Hadoop Database. HBase originated from Google's Big Table paper and was published in November 2006. HBase has the following characteristics: excellent reliability, excellent performance, column oriented, and excellent scalability.

The Server architecture of HBase follows a simple master-slave Server architecture pattern, which is composed of HRegion Server group and HBase Master Server. The HBase architecture is shown in FIG 1 below.

5.2 Research and experiment ideas of Hadoop Big Data Platform in the transportation industry

At present, the most widely used method to improve intelligent transportation is GPS positioning technology. GPS positioning technology can quickly and accurately know the location of the locator, so as to feed back to the traffic management platform for analysis. It is very difficult to analyze the massive data in the increasingly complicated big cities. On the one hand, data storage, for traditional
stand-alone file systems, it is impossible to store big data, so we use HDFS distributed storage system to store our sample data. We used pseudo distribution to build Hadoop platform [18] to analyze one-hour vehicle-mounted terminal positioning data of transportation enterprises. On the Hadoop platform, we chose MapReduce for data processing because its model is simple and easy to use. Write the calculated data results to HBase, then read the data from HBase for chart presentation, and make specific analysis based on the situation. The three main tools we used, all of which have been highlighted in the previous section, are tools for the Hadoop ecosystem. Therefore, this paper chooses to combine Hadoop with positioning data of on-board terminals to analyze vehicle driving radius and driving distance so as to improve traffic management construction, which is a good scheme.

1) Preliminary data processing

The sample data obtained from the transportation enterprise is both huge and irregular, so it needs to use Hadoop big data platform for preliminary processing. We used MapReduce, a distributed computing framework in Hadoop big data platform, to sort the data according to the car ID and time and divide the data into different files for preliminary data processing. These data are then used as the input data to calculate the driving radius, and the statistical distribution map of the driving distance is obtained experimentally.

2) Data extraction

The Hadoop platform is installed and configured on Linux, while the Linux system runs on the virtual machine platform. VMware is the virtual machine platform selected in this paper. VMware is more network-friendly and requires access to the virtual machine’s browser to perform some operations. Hadoop platform experiment, this paper uses pseudo-distributed to build the platform. You are reading files in HDFS. HDFS has been highlighted in chapter 3.

According to the idea introduced above, we carried out the experiment of Hadoop platform processing data to obtain the distribution diagram of driving radius and driving distance. MapReduce has already done some preliminary processing of the data, and we start the concrete operation with the processed data as input data.

As shown in FIG 2 below, the statistical distribution diagram of the maximum driving radius is finally obtained. The following figure shows the maximum driving radius of 5000 data:

FIG 2 Statistical distribution of mileage of 5000 cars
5.3 Analyze the graph of maximum driving radius

The number of interval distributions can be seen more intuitively by combining the two figures 3 and 4. The different intervals are colored in different colors, and we can roughly infer from the figure above that the number of cars distributed between 0 and 10 is the highest. As can be seen from the figure, since the signal frequency of GPS received by the vehicle-mounted positioning terminal is the same, the premise time for calculating the radius is the same. So the maximum driving radius of different vehicles varies greatly in the same time. That is to say, during working hours, some roads are open and some are congested. A large proportion of vehicles are in congested roads, so we can feed back this information to the transportation enterprises, so that the transportation enterprises can timely adjust the transportation plan, reasonably allocate the transportation road, save time, reduce costs, and improve the work efficiency in transportation. We also can get the data back to the intelligent transportation platform, which can be combined with the urban road traffic platform analysis make the hot tried to show, timely adjust and control of traffic, promote the development of transportation industry, overall lead to social progress thereby, because social phenomena, can also reduce congestion to the residents in the waste of travel time.

The interval is separated and a small amount of data is presented for clearer analysis and more intuitive research. As can be seen from the picture, within one hour, most cars cover a distance of 0-20 kilometers. But the general or each interval has the distribution, that is to say, the transportation business has run long distance, have run short distance. Further analysis is based on the number of miles traveled, we can know how much an employee works every day, whether the employee is working hard, cutting corners or using the company car for personal work, which can be seen from the mileage data. Then, according to the mileage of the vehicle, as well as the employee's fuel fee reimbursement, you can know whether the employee intentionally engaged in private affairs. On the other hand, we can also know the operation of the enterprise, which aspect is better. In general, this enterprise short - distance transport projects more. Transportation enterprises are also a kind of transportation industry, by small to big, through this method we can analyze the status quo and development of the entire transportation industry, this experimental study is very practical significance.

6. conclusions

The above analysis shows that the application of big data technology significantly enriched the practical value of urban intelligent transportation system. In the practical application processes, with the help of cloud computing and clustering mechanism, big data can generate a good amount of data acquisition, mass data retrieval and other functions.

With the continuous updating and development of urban transport network in our country, The link between big data and urban intelligent transportation system will become more and more closely, which will provide users with diversified and high reliability service function support.

Bibliography

[1] Can hui Fu. Exploration on responsibility Education for Children of New Jinjiang people [J]. Speed Reading, 2016, (4) : 68-69.
[2] Bo cheng Le. The prediction method and application of Pedestrian traffic flow based on Hadoop Big Data Platform [D]. Guangzhou: Guangdong University of Technology, 2016:3-4.
[3] Da wei Zhang. The Prosperity and loneliness of urban intelligent transportation [J]. China Traffic Informatization, 2011, (9) : 28-29.
[4] Zheng-xi Li. The industrialization trend of Urban intelligent transportation system in China [J]. Automation Expo, 2015, (7) : 60-62.
[5] Yu wang. The engine of big data intelligent transportation development [J]. China Science and Technology Wealth, 2015, (10) : 40-41.
[6] Ming yao Zhou. Development of cloud computing technology and its application in security industry [J]. Public Security in China, 2014, (20) : 60-63.
[7] Zhong xin Liu, Hai ming Zhang, Jian hui Li. Research and Implementation of online Performance Test Platform for distributed file System [J]. Scientific Research Information Technology and Application, 2015, (2) : 11-18.

[8] Ping Zhu, Jian tao Zhu, Jian gang Gao. And other key technologies of performance computing and storage [J]. Computer Research and Development, 2011, (48) : 354-364.

[9] Yong feng Li, Min qi Zhou, Hua liang Hu. Overview of Cluster Resource Unified Management and Scheduling Technology [J]. Journal of East China Normal University, 2014, (5) : 17-30.

[10] Hong jiang Chen, Research and Optimization of fault tolerance Mechanism under MapReduce [D]. Harbin: Harbin Institute of Technology, 2014:22-33.

[11] Guo qing Jiang. Research on data Security of intelligent Mobile Terminals based on cloud Computing [J]. Software Engineer, 2015, (11) : 22-23.

[12] A Wibisono, W Jatmiko, HA Wisesa. I Traffic Big Data prediction and visualization using Fast Incremental Model Rains-Drift Detection [J]. Knowledge-based Systems, 2015, (93) : 33-46.

[13] Ren tong Peng. Hadoop's core technology research or overview [J]. Science plaza, 2012, (5) : 39-41.

[14] Fang Ji. Nortel running Zhongshan Electric Power [J]. Electronic Commerce, 2002, (12) : 45-46.

[15] Tao He. Design and Research of storage Mode for parallel and efficient processing of massive spatial data [D]. Chengdu: University of Electronic Science and Technology, 2014:3-12.