Analysis and Prediction of Urban Traffic Congestion Based on Big Data

Zhenhua Wang[1], Yangsen Yu[2, *], Dangchen Ju[1]

[1]School of Information Engineering, China University of Geosciences, Beijing, China

[2]Geological Survey Institute, China University of Geosciences, Beijing, China

Email address:

wangzh@cugb.edu.cn (Zhenhua Wang), 61412129@qq.com (Yangsen Yu), judc@cugb.edu.cn (Dangchen Ju)

*Corresponding author

Abstract: With the rapid development of big data technology, its application has become more and more extensive. The application of big data technology in intelligent transportation systems is the best way to solve traffic congestion in big cities. The paper analyses in detail the main causes of traffic congestion in big cities and the classification and evaluation of traffic congestion. Utilizing the Internet of Things and modern communication technologies, large-scale traffic data and related data based on GPS are acquired, and data analysis is carried out to construct a traffic prediction vehicle prediction model. The forecasting model is used to predict the traffic flow in each direction of traffic intersections at a certain time, predict the possibility of congestion at a certain time at a certain intersection, the traffic flow and congestion probability of a certain section at a certain time, and the travel trajectory and travel habit forecast of pedestrians. At the same time, consider the impact of non-motorized vehicles and pedestrians on traffic congestion. Use forecasting results and real-time traffic information monitoring to solve traffic congestion problems. Combined with traffic control and optimization strategy control for traffic collaborative management, it provides valuable reference for decision-making in metropolitan traffic congestion solutions.

Keywords: Traffic Congestion, Big Data, Intelligent Transportation System, Road Capacity

1. Introduction

Traffic congestion has become increasingly serious, and traffic accidents have occurred frequently in many cities. These have become traffic management problems that need to be solved. In response to the increasing traffic demand and the pressure of transportation resources, the traditional way of traffic management seems to be inadequate [1]. Intelligent transportation system (ITS) has become an important way to improve traffic management. By use of modern information technology, based on Big Data (BD), the traditional road traffic management methods are deeply reformed to improve the efficiency of urban traffic network and ease urban traffic. Problems and reduce unnecessary losses and improve the efficiency of public transportation [2].

At present, most cities have completed or are building complete three-dimensional traffic monitoring platforms [3]. Vehicle GPS, pedestrian GPS, camera and other monitoring tools are used to comprehensively monitor vehicle travel speed, road cross-section flow, and intersection shunt. By use of these data, real-time evaluation of the operational status of urban roads is possible [4]. Huge and real-time monitoring data forms massive traffic information, which provides effective data support for traffic congestion prediction, and also requires traffic congestion prediction to efficiently and comprehensively cover the entire urban transportation network. Accurate prediction of traffic congestion is one of the core objectives of intelligent transportation systems [5]. Urban traffic conditions have certain self-similarity laws that can be used for prediction; however, road segment and environmental factors generate road segment diversity and road network dynamics caused by external factors, making the traffic congestion law highly complex and uncertain. The prediction model not only needs to adapt to the complex road network conditions, but also achieves high-precision prediction, and needs to achieve high-efficiency updates according to the changes in the road network environment [6].

In the field of traffic flow prediction, a variety of prediction models and methods have been developed, such as linear regression, time series, Kalman filtering, nonparametric models. Kalman filtering is a relatively advanced data processing method based on the filtering theory proposed by Kalman in the 1960s. After the Kalman filtering method was first proposed by Okutani and Stephanedes in 1984, a lot of improvements were made [7]. In 2004, Yang et al. proposed a recursive least squares method for short-term vehicle speed prediction based on maximum likelihood and Bayesian rule. The Kalman filter method was used to adapt to the fast-changing mode. In general, the Kalman filtering method is a matrix iterative parameter estimation method for linear regression analysis models, which has the advantages of flexible selection of predictive factors and high precision. However, due to the large number of matrix and vector operations, the algorithm is more complicated and difficult to use for real-time online prediction.

With the development of Big Data and artificial intelligence, the artificial neural network-based prediction model has been gradually introduced into the field of short-term traffic volume prediction, and has achieved satisfactory prediction results. Smith and Demetsky apply backpropagation neural networks to short-term traffic flow prediction [8]. The neural network models currently applied in the field of traffic prediction include backpropagation algorithms, recurrent neural networks, radial-based RBF neural networks, and multilayer feedback neural networks. This topic mainly uses big data technology combined with deep learning to analyze and predict urban traffic congestion [9].

2. Causes of Urban Traffic Congestion

2.1. Population and Vehicles Increasing in Large Cities

In many big cities, population expansion and the surge of private cars and vehicles are one of the main causes of traffic congestion [10]. There are the following reasons: the distribution of population work and residence is not reasonable, the population density is too large in some areas; in urban public transportation, the investment of public transportation vehicles is also increasing; the increase of non-motor vehicles and pedestrians has a certain impact on traffic congestion [11].

2.2. Problems of Urban Basic Transportation Facilities

In some large cities, the width of many traffic roads is limited due to land shortages. Due to urban planning problems, the distance between some adjacent traffic intersections is relatively short [12, 13]. There are also traffic lights at the intersection of the bus stop and the number of buses stopping at the bus stop. The switching frequency of the traffic signal and the duration of the red and green lights in a certain direction are long. Another problem that cannot be ignored is that the entrances to schools and hospitals are relatively close to the traffic roads. These situations can all lead to traffic congestion [14].

In Figure 1 below, it is analysed that traffic congestion maybe happens due to unreasonable traffic facilities. Factors affecting the efficiency of the traffic intersection: the width of the AB unilateral road; the distance between O1 and O2; the direction of the turning of the vehicle at the O2 intersection; the switching frequency of the O2 traffic signal; The time of the red light (or green light) in a certain direction of O2; the conflict between the right turning vehicles and the straight non-motorized vehicles and pedestrians at the O2 intersection.
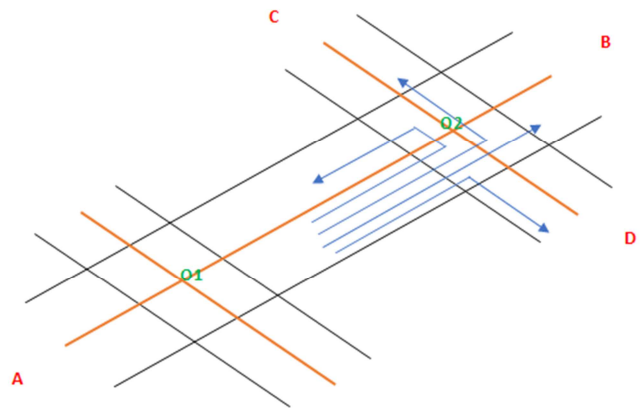

Figure 1. Factors affecting the passage of traffic at traffic intersections. In the figure: the road is between the black lines, the orange line is the middle of the road, and the blue line with the arrow is the direction of the vehicle. AB and CD represent roads. O1 and O2 are the centres of two traffic intersections.

2.3. Other Reasons Cause Traffic Congestion

Traffic accidents can also lead to traffic congestion. The degree of congestion is closely related to the number and severity of road traffic accidents and the speed of accident handling. The occurrence of major events and the necessary traffic control are also the causes of traffic congestion. In addition, traffic lights, such as accidental power outages, road construction, bad weather (such as: heavy rain, blizzard, typhoon, etc.). Sometimes, the traffic congestion in the city has some regularity. For example, the commuting time is more likely to cause congestion. Traffic congestion has a certain relationship with the season. In northern China, due to the cold weather, the number of vehicles and pedestrians traveling in winter is less than in summer, and the traffic conditions in winter are better [15].

In many large cities in China, sharing bicycles facilitates people's travel. The subway has alleviated the traffic congestion pressure of the city to a certain extent, but it also brings certain problems. The traffic congestion near the subway station is very serious [16].

3. Traffic Flow and Traffic Congestion Evaluation

Traffic flow is an important factor causing traffic congestion. Traffic flow refers to the number of passing vehicles per unit time in a particular direction of a road, or

the number of vehicles passing through a unit of time in a certain direction of a traffic intersection [17].

The intelligent transportation industry commonly uses three parameters to quantitatively describe the traffic flow:

(1) The first is traffic flow, also known as traffic volume, which indicates the number of vehicles passing through the designated section of the road in unit time. The unit is vehicles/hour (V stands for traffic volume).

(2) The second is the traffic flow speed, referred to as the flow rate, indicating the speed of the traffic flow, in meters or kilometres (S stands for traffic flow speed)

(3) The third is the traffic flow density, which indicates the degree of density of the traffic flow, that is, the number of vehicles included in the length of the road unit, the unit is km/km (D stands for traffic flow density).

The relationship between the parameters is:

$$V = S * D \qquad (1)$$

In Equation 1, it can be seen that when there are few vehicles on the road, the driver can choose a higher speed. At this time, the traffic flow speed is larger, but because the traffic flow density is small, the traffic flow is also relatively small. As the number of vehicles on the road increases, the traffic density increases, the speed of the vehicle decreases due to the constraints of the vehicles before and after, and the flow rate decreases, but the traffic flow increases until the product of flow rate and density reaches a maximum under certain conditions. The value, that is, when the traffic flow is maximum. The flow rate at this time is called the optimum speed, and the density is called the optimum density. If the vehicle on the road increases again, the density continues to increase [18].

Traffic congestion evaluation indicators are mainly based on road speed, road traffic density, traffic volume and travel time [19].

4. Research on Traffic Congests of Big Data

Among the factors affecting traffic congestion, the collected traffic data is divided into static data and dynamic operational data. Static data refers to long-term fixed data, such as the width of traffic roads, the setting of traffic intersection signals, the lane setting of roads, the distance of adjacent intersections, and so on [20]. Dynamic operational data refers to data that changes in time in traffic information data, such as: location information of vehicles, number of pedestrians at traffic intersections, number of motor vehicles at traffic intersections, and so on. In the case of legal permission and protection of personal privacy, the location information is obtained by the GPS positioning system, the location information of the vehicle or pedestrian is obtained, and the movement trajectory is predicted.

Traffic information is collected by remote terminal equipment, and transmit it to the Big Data centre, establish a dedicated storage system, analyse, classify, organize, and make decisions and implement control. First of all, the electronic map library is the basis of the intelligent traffic control system. It must be first established and updated in real time to provide guarantee for the accuracy of GPS navigation. Road information, traffic flow information, vehicle status information, and parking space information must ensure real-time performance to ensure traffic. Scheduling accuracy and security [21].

Big Data uses road information and human travel information for analysis. Big Data technology greatly reduces the requirements for the structure of data [22]. It uses real-time processing of information such as road information, behavioural habits, and preference information of pedestrians to outline various characteristics of each individual to discover a large amount of traffic flow information. Implicit patterns and rules. Big Data will be dispersed in different departments of traffic data, such as personal information, public transport network information, logistics information, weather information and other information related to the traffic related departments, so that the information of all departments is open and interoperable, to achieve multi-level, cross-sectorial Information resource exchange and sharing. The collaborative computing of heterogeneous data is to enhance the knowledge discovery ability, not only to solve the problem of information fusion, but also to solve the cross-domain association problem of multi-source data [23].

4.1. System Structure Based on Big Data

System architecture discusses the use of the bottom layer of the Internet of Things perception layer to obtain data, the network layer to collect data, Big Data technology data pre-processing, cloud computing platform to provide computing services, large data analysis, and finally to the Internet of Things upper application layer. By use of large data analysis, using historical data and real-time data to predict traffic congestion, remote control traffic lights. According to the corresponding strategy, traffic signal switching frequency or delay the green light time in the direction of congestion are carried out for single intersection congestion. For multi-intersection congestion, multi-intersection signal coordination is used to smooth traffic congestion [24]. The hierarchical structure of the system is analysed in detail (see Figure 2).
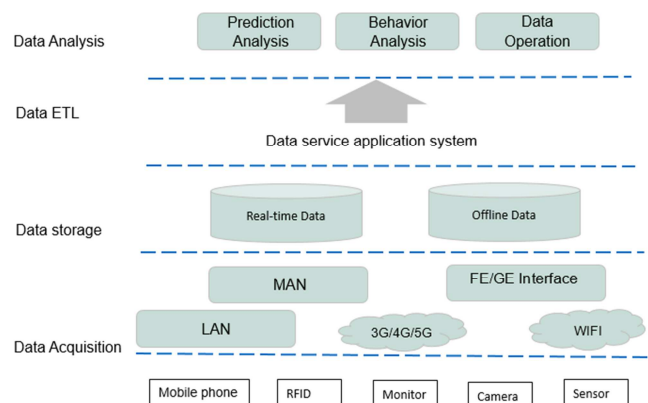
Figure 2. *ITS architecture based on Big Data.*

4.2. Traffic Data Acquisition

By use of a series of data acquisition devices, such as mobile phones, car network, GPS receiver, video surveillance, cameras, etc., a series of information can be obtained such as the time, location, and number of the traffic vehicle. The purpose of obtaining this information is to analyse the vehicle's trajectory and predict the amount of traffic and the likelihood of congestion in a place [25]. The information in Table 1 below is required and how to obtain the information. Car GPS is developing rapidly, generating a large amount of spatio-temporal data, which can be used to predict congestion road segments.

Table 1. *Vehicle information on traffic roads.*

Items	Data Sources
Buses	Time, Location, Car number
Taxis	Time, Location, Car number
Private cars	Time, Location, Car number
Logistics vehicles	Time, Location, Car number
Other vehicles	Time, Location, Car number

Logistics-related vehicles include oil tank trucks, cement tank trucks, garbage collectors, trucks with detachable decks, car carriers, and water sprinklers etc.

Pedestrians and non-motorized vehicles are an important factor affecting traffic congestion in some large cities. Pedestrians and non-motorized vehicles are an important factor affecting traffic congestion in some large cities [26]. Especially at some traffic intersections, pedestrians or non-motorized vehicles who do not obey the traffic rules affect normal driving, which is likely to cause traffic congestion. Due to improper setting of traffic lights at some traffic intersections, turning vehicles will also collide with pedestrians and non-motorized vehicles, causing traffic accidents and causing traffic congestion. The information in Table 2 is about the data of the Pedestrians and non-motorized vehicles. In China, motorcycles are driven on non-motorized roads. The subway has slowed down the pressure on the ground traffic, but the subway station is a place that is likely to cause traffic congestion. Non-motor vehicles are divided into bicycles, tricycles, electric bicycles, disabled motor wheelchairs and animal vehicles [27].

Table 2. *Pedestrians and non-motor vehicles information on traffic roads.*

Items	Data Sources	Description
Pedestrians	Time, Location information	Bus card, Mobile phone GPS, etc.
Non-motor vehicles	Time, Location information	Mobile phone GPS
Motorcycles	Time, Location information	Mobile phone GPS
Other auxiliary public transport	Number of passengers transported by subway	Bus card, Mobile phone GPS, Ticket sales, etc.

The condition of roads is also a factor affecting traffic. Traffic will be good with the more lanes, the wider lane, and the longer distance of adjacent intersection and the higher of the lane. The contents in Table 3 detail this aspect.

Table 3. *Road conditions related to traffic congestion.*

Items	Data Sources	Description
Number of lanes	Quantity	Lanes of the road
Lane width	width	The wider the better
Adjacent intersection distance	length	The long the better
Speed limit	Quantity	Usually 60 or 40

There are many other complex issues that affect the state of urban road traffic. It not only includes the differential impacts of different regions of the city, including differences in population density in different regions of the city, differences in road infrastructure construction specifications, etc., but also due to environmental factors such as weather, for example heavy rain, blizzard and other abnormal weather [28, 29]. There are other factors that affect traffic congestion: the occurrence of major events, traffic flow control, and maintenance of transportation facilities. The information in Table 4 below is descript in detail.

Table 4. *Other factors affects traffic congestion.*

Items	Data Sources	Description
Weather	Time	Heavy rain, strong winds, heavy snow.
Important events	Time, place, driving route	Sports competitions, foreign leaders visiting, etc.
Traffic control	Traffic accident limit	Other special events
Traffic maintenance	Time, place	Road repair, bridge repair.

4.3. Data Analysis and Processing

Based on the analysis of dynamic traffic flow and road network congestion status, combined with the temporal and spatial characteristics of traffic data and traffic domain constraints, the potential similarities, correlations and correlations of the data are analysed in depth [30]. The traffic data is clustered, predictively analysed, correlated with analysis, and anomaly detected, so as to discover the knowledge hidden by different feature dimensions and different data granularities, and use the dimensionality reduction technology to process the data.

4.3.1. Short-Term Intersection Traffic Flow Forecast

Data mining technology is used to predict traffic flow. Traffic flow data is a form of time series that is segmented according to changes in time series data characteristics. The most commonly used time series segmentation method is a piecewise linear description, that is, the sequence is

segmented and described piece by piece using a linear model. K-Means clustering algorithm is used as the basis of sequence segmentation. The traffic flow is predicted by the combined model algorithm of sequence segmentation and BP neural network. The traffic flow pattern is divided according to the two-dimensional clustering algorithm of flow and time [31, 32].

4.3.2. Traffic Congestion Forecast

At present, many research results of traffic congestion prediction are mainly concentrated in the subject area, mainly based on time series correlation prediction analysis, neural network prediction, Bayesian network prediction, and multi-classifier combination prediction [33].

The traffic flow feature vector is constructed by summarizing the basic data such as the traffic flow parameters, the environmental state, and the time period, and the four predicted states are determined (then the levels are determined according to different driving speeds, such as smooth: $V \geq 30$; Crowded: $10 \leq V \leq 30$; Congestion: $3 \leq V \leq 10$; Blockage: $V \leq 3$, V stands for speed and is measured in km / h). A self-encoding network method using deep learning learns from the unlabelled data set to obtain hidden layer parameters that can characterize deep features of the data and generate new feature sets. Softmax regression is used to learn the new feature set with label to generate predictive classifier, and the model predicts the polymorphism of traffic congestion [34].

4.3.3. Pedestrian and Non-motorized Vehicle Trajectory Prediction Based on GPS Data

At many traffic intersections, the collision of pedestrians and non-motorized vehicles with motor vehicles has led to traffic congestion. Based on GPS data, the moving trajectory of a pedestrian or a non-motor vehicle at certain time is predicted by using relevant clustering algorithm. In this way, the number of pedestrians and non-motorized vehicles can be predicted at a certain traffic intersection at a specific time [35].

5. Conclusions

By use of Big Data technology to get the traffic information of the entire city, and provide practical data and solutions for traffic guidance and urban planning, it can solve the following problems: traffic congestion prediction analysis and processing; traffic flow forecast; scientific planning of transport infrastructure. By the technology of the Internet of Things to collect data, obtain real-time data and historical data, Using Big Data technology, the data is cleaned and pre-processed, and the appropriate algorithm is selected to establish a traffic prediction model. With the widespread use of driverless technology in large cities, intelligent transportation systems based on Big Data can provide accurate and highly reliable traffic information for the Internet of Vehicles.

Funding Statement

Abbreviations

The following abbreviations are used in this manuscript:
BG: Big Data
ITS: Intelligent Transportation System

References

[1] Pedro Wences; Alicia Martinez; Hugo Estrada; et al. Decision-Making Intelligent System for Passenger of Urban Transports [J]. Ubiquitous Computing and Ambient Intelligence, 2017, Vol. 10586: 128-139.

[2] Guan, Shuqi; Yun, Jun; Zhang, Qinkang; et al. Intelligent transportation system contributions to the operating efficiency of Urban traffic [J]. Journal of Intelligent & Fuzzy Systems, 2016, Vol. 31(4): 2213-2220.

[3] Li Zhijie, Li Yuanxiang, Wang Feng, et al. Overview of online learning algorithms for big data analysis [J]. Computer Research and Development. 2015(08).

[4] Hu Peifeng. Research on short-term forecasting method of traffic flow [D]. Beijing Jiaotong University 2007.

[5] Qin Ming, Zhang Wenqiang, Zhong Xianfei, et al. Space-time optimization of intersection based on Kalman filter traffic prediction [J]. Science and Technology Plaza. 2015(01).

[6] Yang Qingfang, Zhang Wei, Gao Peng. Short-term Forecasting Method of Traffic Volume Based on Improved Dynamic Recurrent Neural Network [J]. Journal of Jilin University (Engineering Science). 2012(04).

[7] Dong Rui, Jia Yuanhua, Qi Guchang. Traffic Flow Prediction Based on Wavelet Denoising and Chaotic Time Series [J]. Science Technology and Engineering. 2010(31).

[8] Jiao Guangting, Xu Jianbiao, Ma Yonghong. Research on Traffic Volume Forecasting Method Based on QPSO-RBF [J]. Transportation and Computer. 2008(04).

[9] Tan Manchun, Li Yingjun, Xu Jianbiao. Combined Traffic Flow Prediction Based on ARIMA and SVM Based on Wavelet Denoising [J]. Journal on Highway and Transport Engineering. 2009(07).

[10] Catalin Gosman; Tudor Cornea; Ciprian Dobre; et al. Controlling and filtering users data in Intelligent Transportation System [J]. Future Generation Computer Systems, 2018, Vol. 78: 807-816.

[11] Chin, Cheng Siong; Zhong, Xionghu; Hamdan, Mohammad; et al. Intelligent Autonomous Transport Systems Design and Simulation [J]. JOURNAL OF ADVANCED TRANSPORTATION, 2018, 1-2.

[12] LIN Chi, XU Bo, XUE Yuwei, YU Cheng. Experimental Platform for Intelligent Data Protection of Big Data in Intelligent Transportation [J]. Laboratory Research and Exploration, 2017, Vol. 36 (7): 39-42, 49.

[13] The Experience and Enlightenment of American Intelligent Traffic Management Experience [J]. Road traffic science and technology, 2016, (6): 42-50.

[14] Wen Huimin, Quan Yuxiang, Sun Jianping. The Development Direction of Urban Intelligent Transportation System in the Age of Big Data [J]. Urban Transport, 2017, Vol. 15 (5): 20-25.

[15] Agachai Sumalee; Hung Wai Ho. Smarter and more connected: Future intelligent transportation system [J]. IATSS Research, 2018.

[16] Zhao Na, Yuan Jiabin, Xu Wei. Summary of Intelligent Transportation System [J]. computer science, 2014, (11): 7-11，45.

[17] LU Huapu, SUN Zhiyuan, QU Wencong. A Review of Big Data and Its Application in Urban Intelligent Transportation Systems [J]. Journal of Transportation Systems Engineering and Engineering, 2015, Vol. 15 (5): 45-52.

[18] Shaojun Zhang; Tianlin Niu; Ye Wu; et al. Fine-grained vehicle emission management using intelligent transportation system data [J]. Environmental Pollution, 2018.

[19] Bao Yuchen, Xie Xiaoyan, Gao Lei. Data technology of intelligent transportation in the era of big data [J]. Chinese information, 2018, (5): 5.

[20] Feng Kai. Development status and trend of urban intelligent transportation system [J]. Global market information guide, 2017, (1): 111-112.

[21] Jinchao Wu; Bokui Chen; Kai Zhang; et al. Ant pheromone route guidance strategy in intelligent transportation systems [J]. Physica A: Statistical Mechanics and its Applications, 2018, Vol. 503: 591-603.

[22] Wei Yanfang, Chen Peng. Discussion on the theory of urban intelligent transportation [J]. urban construction under the Internet + background (2017), (31): 201-202.

[23] Zhang Hong, Wang Xiaoming, Cao Jie, et al. Intelligent Transportation System Architecture Based on Big Data [J]. Journal of Lanzhou University of Technology, 2015, Vol. 41 (2): 112-115.

[24] Chilà Giovanna; Musolino Giuseppe; Polimeni Antonio; et al. Transport models and intelligent transportation system to support urban evacuation planning process [J]. IET Intelligent Transport Systems, 2016, Vol. 10(4): 279-286.

[25] Yang Yong. Henan Intelligent Traffic Management System Based on Big Data [D]. Beijing University of Technology, 2016.

[26] Zhang Peng. Analysis of Intelligent Traffic Signal Control System Based on Traffic Big Data [J]. Science and Informatization, 2018, (6): 17-18.

[27] Sergey Sheptunov; Yuriy Solomentsev; Nataliya Suhanova; et al. Improvement of the urban transport management system with elements of artificial intelligence. [J]. Bulletin of Bryansk State Technical University, 2016,: 93-98.

[28] Wang Peng. Research on urban intelligent traffic controller based on embedded system [D]. University of Science and Technology Liaoning, 2016.

[29] Hacène Fouchal; Emilien Bourdy; Geoffrey Wilhelm; et al. A validation tool for cooperative intelligent transport systems [J]. Journal of Computational Science, 2017, Vol. 22: 283-288.

[30] Yen-Wen Lin; Yuan-Kai Hsiao; Zih-Shiuan Yeh. A New Mobility Management Scheme for Intelligent Transportation Systems [J]. Wireless Personal Communications, 2017, Vol. 96(2): 3081-3112.

[31] Alfred Daniel; Anand Paul; Awais Ahmad; et al. Cooperative Intelligence of Vehicles for Intelligent Transportation Systems (ITS) [J]. Wireless Personal Communications, 2016, Vol. 87(2): 461-484.

[32] J Jariyasunant; A Carrel; V Ekambaram; et al. The Quantified Traveler: Changing transport behavior with personalized travel data feedback [J]. The Ninth China (International) urban intelligent transportation Forum, 2012.

[33] Lu Hua Pu, Sun Zhiyuan, Qu Wen cog. Large data and its application in urban intelligent transportation system [J]. transportation system engineering and information, 2015, Fifteenth volumes (5): 45-52.

[34] Wen Huimin, Yu Xiang, Sun Jianping. The development direction of Urban Intelligent Transportation System in the age of big data [J]. urban traffic, 2017, Fifteenth volume (5): 20-25.

[35] Li Jun. Research on dynamic path planning in urban intelligent transportation [D]. Hangzhou Dianzi University, 2016.