Big Data Platform & Typical APP Services for Urban Public Transportation

Lin Cao
University of Chinese Academy of Sciences
Beijing, China
caolin2016@ia.ac.cn

Xisong Dong
The Qingdao Academy of Intelligent Industries
Qingdao, China
xisong.dong@ia.ac.cn

Jiehan Yu
Guangzhou Communication Information Construction Investment and Operation Co. Ltd
Guangzhou, China
yujiehan@96900.com.cn

Gang Xiong (Correspondence Author)
Cloud Computing Center, Chinese Academy of Sciences
Dongguan, China
xionggang@casc.ac.cn

Fenghua Zhu
The Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation ,Chinese Academy of Sciences
Beijing, China
fenghua.zhu@ia.ac.cn

Zhen Shen
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences
Beijing, China
zhen.shen@ia.ac.cn

Bin Hu
Cloud Computing Center, Chinese Academy of Sciences
Dongguan, China
binhu@ia.ac.cn

***Abstract*—Along with more and more people and cars appear in cities, the stricter requirements, like traffic efficiency, safety, service and cleanness cannot be met easily by using the existing urban traffic plan and operation management methods, which are normally supported by the separated and partial Intelligent Transportation System (ITS). In this paper, big data platform for urban public transportation is proposed, and its architecture and subsystems are developed. Then, Guangzhou's typical APP services for urban public transportation are described in detail, including: real-time bus service, taxi inquiry and call service, mobile payment, and traffic information service, etc.**

Keywords—Big Data Platform, APP Services, Urban Public Transportation

I. INTRODUCTION

The explosive growth of population and the rapid development of urbanization lead to less and less land for urban transportation construction, which can have many negative effects, such as worsening traffic environment, traffic congestion, traffic accidents and so on. Therefore, it is important to develop a new kind of Intelligent Transportation System (ITS) to coordinate people, vehicles and roads more fluently, and then to improve traffic efficiency, to relieve traffic congestion, and to reduce traffic accidents, energy consumption and environment pollution.

With the introduction of Internet of Thing (IoT) and cloud computing, ITS based on big data is coming. Taking Beijing as an example, more than 60,000 taxis can produce hundreds of millions of GPS data daily, the data volume of license plate recognition and traffic monitoring video is even bigger. At present, traffic daily data volume jumps from TB level to PB level with a wide range of sources and forms, which has a huge impact on the operation and management of ITS. As a result, solving the traffic problems with big data becomes the inherent demand of ITS and the inevitable trend.

In fact, the "Big Data" concept was proposed by Americans in 1980s. In 2008, Nature published a special issue - "Big Data" [3]. In 2011, Science published "Dealing with Data" [4]. In 2012, The U.S. government officially released the big data re-search and development initiative [5] and ERCIM News published "big data" [6]. There is no uniform and authoritative definition of big data. According to the Gartner, "Big Data is high-volume, high-velocity, and/or high-variety in-formation assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" [7]. In Wikipedia, big data is a term for data sets that are so large or complex that traditional data processing tools are inadequate to deal [8]. Big data has the characteristics of real-time and high predictability, which can monitor the status of public transportation in real time and realize intelligent traffic management, especially with advanced detection means, such as GPS or Beidou [9]. Big data provides the possibility to solve the problem of urban public transportation.

In recent years, mobile Internet is integrated into some ITS applications. With the popularity of mobile terminals, they are used for information collection of road traffic status to analyze the travel rules and provide real-time traffic information, navigation information and other services. Recently, traffic Application (APP) can provide the personalized services to improve traffic management and alleviate travel problems. And real-time bus APP has a great ad-vantage on solving the existing problem, such as real-time service, reminder, and suggestion and so on.

The organization of the paper is as follows: Section II gives a comprehensive introduction to Big Data Platform for Urban Public Transportation. Section III presents typical APP services "XingXunTong (XXT)" for Guangzhou`s public

transportation. Then the conclusions are drawn, and future research directions are given in Section IV.

II. Big Data Platform for Urban Public Transportation

With the rapid development of Internet, big data pro-vides a new opportunity and all walks of life are always concerned about the development and application of big data, and public transportation also needs to solve the bottleneck problem through big data technology. The basic way to solve the urban traffic problem is to develop the public transportation, optimize the bus routes and improve the urban public transport management system, so as to plan the overall development of the city.

In the traditional mode of public management, public transportation enterprise is very difficult to generate re-al-time data, extract and analysis data, such as the bus's GPS data, vehicle speed, and real time traffic information. Therefore, data cannot be used to analyze the public decision-making management and support. However, by big data technology, the public transportation manager can master comprehensive data information and carry on the specialized analysis and processing. For example, when the bus company received the passengers' complaints about a road traffic congestion, they can track the line, analyze the site (transfer), operation time and so on to adjust departure interval and meet passenger travel demand.

The architecture of the big data platform is shown as Fig. 1, based on the existing ITS resources, mobile APPs are convenient application tool for the public to obtain comprehensive and personalized information services.

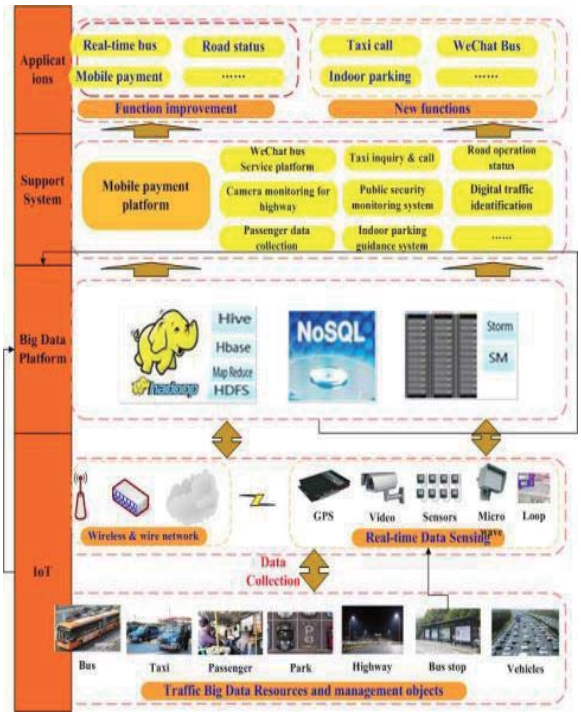

Fig. 1. Architecture of Big Data Platform for Urban Public Transportation

A. Traffic Big Data Platform

1) Integrated Traffic Information Platform

The platform is capable of data collection, storage, pro-cessing, and exchange. Based on SOA model, its system architecture is designed including presentation layer (release mode), service layer (service release) and data layer (basic information).

a) Presentation Layer

The layer (Fig.2) is used to receive traffic information service provided by service layer, and show the available services. The layer can be used by electronic maps based on spatial location, or diagrams, text and multimedia interfaces.

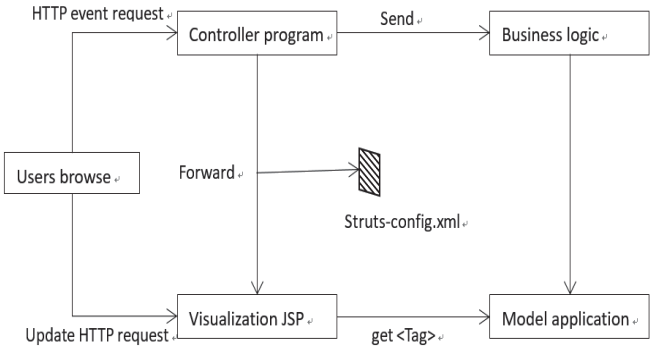

Fig. 2. Presentation layer

b) Service layer

Service layer (Fig.3) includes some functional modules such as road information, real-time bus, travel planning, which are defined as independent Web services. Data pro-cessing, such as analysis, combination, modeling, and out-put formatting, is executed by application server and Enter-prise Service Bus (ESB) to form function lists required by traffic guidance system and provide consistent service search and access interfaces. Furthermore, traffic infor-mation is packaged into Web service, so external users can register and look up service by Universal Description, Dis-covery and Integration (UDDI), and intelligent integration and management of services can be realized by ESB.

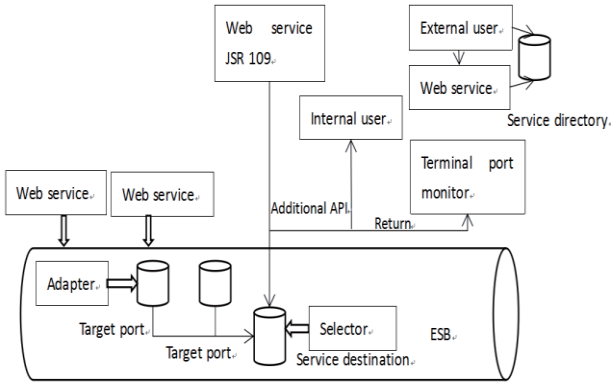

Fig. 3. Service layer

c) Data Layer

In order to facilitate the data calls and modifications, the traffic information is stored in database, the service list is stored in the service database, and actualizes standard data call interface. The data layer analyzes and makes better use of these big data in the platform, like distributed real-time processing and value-added application such as analyzing user habits.

The traffic information platform integrates almost all the basic data of the traffic system and the database should prevent the data from being accessed by users without permission to access. Therefore, it is necessary to use data encryption technology and define different roles for the database system and each role has different access rights to control access rights for different users. When the database is accessed through network, it needs to increase network security mechanism to ensure security of network. Using data encryption and authentication authority in the network transmission can prevent data being monitored in the network.

2) Application Technology Framework

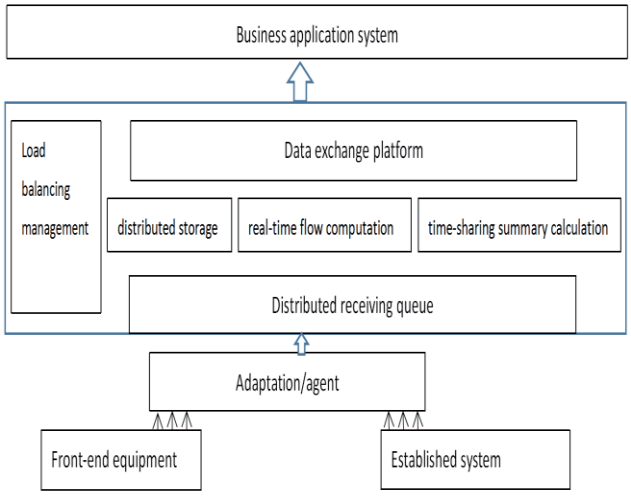

Fig. 4. Traffic data processing architecture

Front-end equipment data includes operational data of people, vehicles and roads. Established system data includes passenger flow data, video data and so on. The data of front-end equipment and established system can be converted into uniform standards through adaptation/agent treatment. Distributed receiving queue provides loose coupling method for distributed application in synchronous or asynchronous modes. There are three processing options for those data entering the distributed receiving queue, those data without calculation adopts distributed storage; those data with on-line calculation adopts real-time flow computation; those data which needs to be calculated based on time can be added to the distributed storage after summary calculations. All data can be exchanged on data exchange platform. Load balancing management can dynamically allocate storage and computing resources.

3) Logical Function Framework

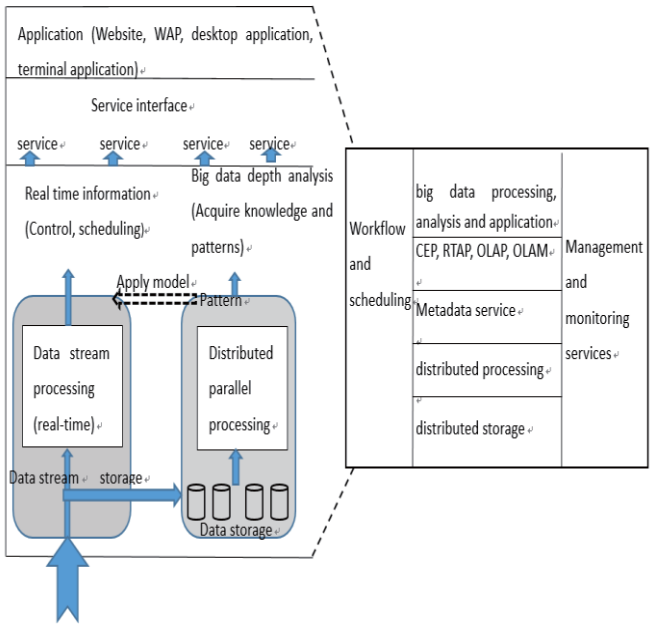

Fig. 5. Design logic function framework

From the bottom to the top, the logical function structure includes distributed storage layer, distributed processing layer, Metadata service layer, processing and analysis layer (including CEP, RTAP, OLAP, OLAM), traffic big data processing, analysis and application layer. Besides, it is necessary for distributed system to execute these tasks such as resource scheduling and management, and middleware monitoring. The left half of the logical framework shows the component structure diagram of the traffic big data analysis and processing platform, which can be divided into real-time data stream processing and big data depth analysis subsystem logically. The platform can provide traffic information services and release various modes of traffic information, such as Web traffic information services, radio and television stations, touch screen query terminals and so on. The applications and services can be connected through unified service interface which provides the upper application with consistent call interfaces that isolate applications and service to achieve the flexibility of the function expansion of platform.

In order to ensure the correctness of the logic function of the platform, the database system should provide transaction processing ability, if one step fails, the database system must return and change operation to ensure the consistency and integrity of traffic data.

Besides, through the switch (Fig.6), data enters the distributed storage system or the traditional relational database, and customizing the intermediate table conversion system is required between them to convert the suitable data formats and exchange. The business system queries the data of the distributed storage system with API call and the relational database with SQL call.

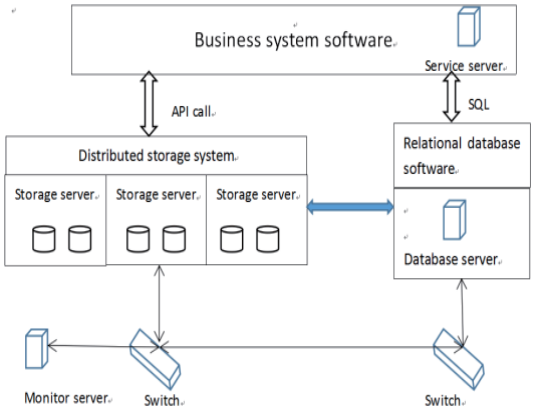

Fig. 6. The combination of big data system and traditional system

B. Subsystems of Big Data Platform for Urban Public Transportation

1) Big Data Acquisition Subsystem

The subsystem mainly realizes the automatic collection and access from the existing traffic database. It can standardize the multi-sources and heterogeneous data, and then store the data together into the platform. The data access needs to follow unified data exchange specification, data control strategies and security mechanisms. The ways to access data include FTP, Web Service, and direct database and so on. The data acquisition subsystem of urban public transportation uses cloud computing, high-definition monitor, mobile communication technology, car networking to achieve a full range of traffic data collection. The front-end equipment can access information of vehicle, running state information, environment information around and determine the pavement status and detection on the environment. Because of collection of multiple intelligent traffic system, data update frequency of the subsystem is faster.

2) Big Data Preprocessing Subsystem

The subsystem is to ensure the access data accurate and effective, then the stored data is inspected. The methods include data cleaning, standardized processing, data fusion analysis. The accuracy of observed data is the basis for making full use of observed data. Most observed data has a lot of noise data, the zero point data, which makes the observed data poor rationality, identicalness and practicability. Therefore, it needs to check and correct data by software and scientific model.

3) Urban Road Operation Analysis Subsystem

Based on the historical data of bus arrival and the latest road conditions or other factors, the arrival time (or distance) of the nearest bus can be forecasted by the prediction algorithm, which possesses the self-learning function and can continuously improve the accuracy of forecast stations. And the function improves the convenience of public transport. Based on public transportation data, we extract, integrate and carry out careful analysis to obtain useful information and knowledge. Using relevant theories and techniques of data integration, decision support and expert model can analyze traffic flow or congestion trend, etc. In addition, data mining techniques are used for traffic flow prediction, traffic analysis, road traffic safety, at the same time, artificial intelligence and neural network are used to provide effective technical support for traffic data collection, data management, and intelligent analysis.

4) Big Data Storage and Management Subsystem

The complete and orderly storage of massive data is the core part of the platform and establishing database is the primary task of the information sharing platform. The in-formation platform should not only solve the reasonable storage of mass data, but also realize the sharing of information. Regardless of the various applications and queries, the retention of historical information and the management of real-time information should be carried out through the database. At the same time, the data can be extracted and managed, and stored in the corresponding module or sub database after data conversion. The subsystem can provide efficient data access interface for the upper application, access to PB or even EB magnitude of operational data, complete data access within acceptable response time and ensure the correctness and availability of data. For bottom device, it can manage the storage resources and make full use of the physical characteristics of the devices.

5) Real Time Data Stream Processing Subsystem

The data stream records the space data and attribute da-ta such as speed or video, image data. The subsystem is the database of the real-time traffic information service, which can provide support for traffic information management such as bus scheduling, travel planning and decision making, and provide more convenient service for users. The subsystem includes some key technologies in which high speed data conversion makes event data stream random access format to distributed parallel data acquisition format and deals with traffic data real-time to get the latest results. Flexible resource allocation scheme makes different types of data processing components and distributed key value storage connected to easily construct a new service and does not affect the existing system operation.

6) Unified Management and Scheduling of Resource Subsystem

The subsystem allows different computing frameworks adopted by different applications to deploy to public clusters and allows unified management and scheduling of re-sources (memory, CPU, I/O, etc.) to improve resource utilization and reduce operation costs. Mesos which is cluster resource manager intends to be used to build resource sharing platform and makes each computing framework responsible for task scheduling and execution control by de-fining the minimum interface for resource sharing among multiple computing frameworks.

7) Depth Analysis Subsystem

On the basis of the comprehensive traffic information, data with multiple sources and inconsistent formats can be extracted and integrated, and then achieve depth analysis and processing by advanced data processing technology to obtain patterns, models, rules and knowledge of decision. To meet the needs of big data for public transportation, traditional data mining, machine learning algorithms are needed to be transformed.

Traffic data platform system is a multi-user and multi-task system. It needs to support the concurrent access of users and interact with each traffic subsystem at the same time. It needs to obtain real-time and dynamic traffic data from various data sources, process data according to specified requirements, and send the data to each user, and make a response on real-time dynamic data request of multiple users. Therefore, database module should be able to handle concurrent processing requirements.

III. Application Case

Based on traffic big data platform, Guangzhou, China provides a comprehensive traffic information service mobile APP "XingXunTong (XXT)", which mainly provides such functions like real-time bus service, taxi inquiries and taxi call service, traffic information service, mobile payment service, parking, travel planning, subway, etc., and integrates more than 700 bus routes, real-time traffic information and arrival of more than 10,000 buses (coverage above 90%). As a result, with XXT, everyone can conveniently grasp the real-time traffic conditions of Guangzhou at any time and place.

A. Real-time Bus Services

It provides bus lines, transfer information, site location and other information query services.

- Line diagram shows the real-time vehicle position of the current line, electronic bus stops show real-time arrival information of the current site. The real-time bus inquiry of XXT also realizes the mutual correlation among the sub modules, which is convenient for users to change the query pages.
- The function of waiting reminder or arrival reminder: users choose departure/arrival site, specific route, reminder condition, and it can play a reminder when the next car meets the conditions.
- The users can view the crowded degree of all buses in real-time. Video detection devices for bus passenger flow and intelligent scheduling terminals are connected through the serial port, and the number of on-off passengers will be transmitted in real time to the vehicle intelligent scheduling terminal by video flow counter, and to the background management platform through GPRS, so that the number of on-off passengers, stranded passengers and other in-formation can be obtained.
- The accuracy of bus arrival can be improved by using station identification. The intelligent bus system in Guangzhou mainly uses the vehicle terminal of installing satellite positioning system (GPS or Beidou) to collect location data. And, the precise positioning of the vehicle can be obtained with the digital identification of the traffic factors such as bus, bus station/parking.

B. Taxi Inquiry and Taxi Call Services

The service can realize the integration of the vehicle, driver, industry regulation and social call software, playing an important role in promoting the service quality of the city. Taxi call service function is shown as Fig. 7, and its flow chart is shown as Fig. 8.

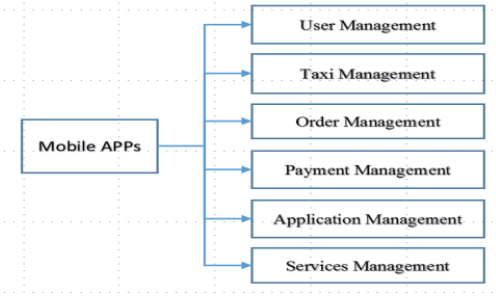

Fig. 7. Taxi call service function

- User management module mainly realizes user registration, login, information maintenance and other functions.
- Taxi module. Passenger can call taxi at any time and any place by using XXT, then the background system will automatically assign taxi for a call in accordance with scheduling principle. The module mainly includes automatically getting passenger position, selecting taxi type, choosing vehicle information, confirming vehicle information and protocol, call, etc.
- Order management module. Taxi call orders have 6 kinds of states, respectively, submitting, scheduling, completing, cancel, complaints and freezing.
- Payment management module mainly achieves the related payment function of the electronic payment channels and data validation function of other payment channels, including electronic payment, cash payment, and can inquire the detailed costs of the call service.
- Application management module mainly realizes the related application management, and it increases the interaction between passengers and taxi information platform, mainly including trajectory, mile-age statistics, travel bills, general information management and traffic service information.

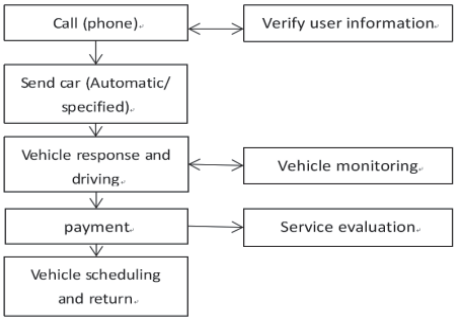

Fig. 8. Flow chart for call service

C. Traffic Information Services

Real-time monitoring of the city's major traffic intersection (Fig. 9) can be obtained on the real-time video image of the

mobile phone software and the whole traffic condition of the city and real-time congestion status of main roads can be queried, and it has good enough accuracy rate. Sub functions mainly include road condition inquiries, traffic diagram, traffic ranking, traffic information and micro-blog sharing and so on. The advantages of the service are as follows.

Fig. 9. Real-time monitoring video

- Covering all features of the Google map road query, and getting more comprehensive and accurate query information, and more intuitive results.
- Achieving query function of traffic diagram, and showing road map of the trunk roads, which does not have similar function in the existing software.
- Road congestion ranking, not only shows the ranking of congested road sections, the starting point and the end point of each section, but also can be synchronized to the road map to view real-time traffic information.
- Traffic information shows traffic alert in the day.

The acquisition of traffic video information usually uses high definition video camera to monitor motor vehicles, non-motorized vehicles of the road sections in real time and record the relevant image data with advanced technology of photo electricity, computer, image processing, pattern recognition, remote data access and so on, and it can help to avoid congested roads, making traffic guidance from passive to active.

D. Mobile Payment

Combined with third-party payment such as Yang-ChengTong, NFC, mobile payment, UnionPay flash pay, XXT can realize service query, online consumption and electronic payment.

IV. Conclusion and Discussion

With the rapid development of social economy and the continuous improvement of people's living standard, traffic congestion has become a severe problem in many cities. The existing urban traffic plan and operation management method cannot meet the stricter requirements. In the paper, urban public transportation by using big data is discussed, big data platform is developed, and application case of Guangzhou's APP is introduced. Clearly, the use of mobile APP services in public transport will be more and more popular, which can provide real-time service for passengers to reduce the time waiting for the bus. In the future, with more data integrated into the platform, more and better APP services will be available.

Acknowledgment

Research supported by the National Natural Science Foundation of China under Grants 61233001, 71232006, 61533019, 61773381, 61773382, and 91520301; Chinese Guangdong's S&T project 2015B010103001, 2016B090910001, Dongguan's Innovation Talents Project (Gang Xiong) .

References

[1] H. Lu, Z. Sun, W. Qu, Big Data and Its Applications in Urban Intelligent Transportation System [J]. Journal of Trans-portation Systems Engineering and Information Technology, 2015, 15 (5):45-52.

[2] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, L. Yang, Big Data for Social Transporta-tion. IEEE Transactions on Intelligent Transportation Systems, 2015, 17(3): 620-630.

[3] F. Frankel, R. Reid, Big data: Distilling meaning from data [J]. Nature, 2008, 455(7209): 30-30.

[4] W. Los, J. Wood, Dealing with data: Upgrading infrastructure [J]. Science, 2011, 331(6024): 1515-1516.

[5] D. Li, Y. Yao, Z. Shao, Big data in smart city [J]. Geomet-rics and Information Science of Wuhan University, 2014, 39(6): 631-640.

[6] Big data [J]. ERCIM News, 2012, 89.

[7] Big data. http://www.gartner.com/it-glossary/big-data.

[8] Big data. https://en.wikipedia.org/wiki/Big_data.

[9] Y. Yu, L. Lei, Z. Jing, W. Wang. Brief probe into the status and trend of compass navigation's application in domestic intel-ligent transportation systems [J]. Journal of Engineering Studies, 2014, 6(1): 86-91.

[10] Q. Shi, A. Mohamed, Big Data applications in real-time traffic operation and safety monitoring and improvement on urban ex-pressways. Transportation Research Part C: Emerging Technol-ogies, 2015, 58: 380-394

[11] L. Floridi, Big data and their epistemological challenge. Philoso-phy &Technology, 2012: 1-3.

[12] A. Anthony, Model for determining optimum bus-stop spac-ing in urban areas. Journal of Transportation Engineering, 2001, 127(3): 195-199.

[13] X. Wu; X. Zhu; G. Wu; W. Ding. Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(1): 97-107.

[14] G. Nowacki, Development and Standardization of Intelligent Transport Systems [J]. Transnav the International Journal on Marine Navigation & Safety of Sea Transportation, 2012, 6(1): 224-226.

[15] G. Xiong, F. Zhu, X. Liu, X. Dong, W. Huang, S. Chen, K. Zhao, Cyber-physical-social System in Intelligent Transportation. IEEE/CAA Journal of Automatica Sinica (JAS), July 1, 2015, 2(3): 320-333.

[16] Y. Lv, Y. Duan, W. Kang, Z. Li, F. Wang, Traffic Flow Prediction with Big Data: A Deep Learning Approach, IEEE Transac-tions on Intelligent Transportation Systems, 2015, 16(2): 865-873.

[17] Y. Liu, Y. Peng, B. Wang, S. Yao, Z. Liu, Review on Cyber-physical Systems. IEEE/CAA Journal of Automatica Sinica , 2017, 4(1), 27-40.

[18] L. Li, Y. Lin, N. Zheng, F. Wang, Parallel learning: a perspective and a framework. IEEE/CAA Journal of Automatica Sinica, 2017, 4(3), 389-395.

[19] G. Xiong, F. Zhu, X. Dong, H. Fan, B. Hu, Q. Kong, W. Kang, T. Teng, A Kind of Novel ITS based on Space-Air-Ground Big-data, IEEE Intelligent Transportation Systems Magazine, 2016, 8(1): 10-22.