Cities 82 (2018) 19–26

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Improving the approaches of traffic demand forecasting in the big data era

Yongmei Zhao[a,b,*], Hongmei Zhang[b], Li An[b], Quan Liu[b]

[a] School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, PR China
[b] Material Management and Unmanned Aerial Vehicle Engineering Institute, Air Force Engineering University, Xi'an, PR China

ARTICLE INFO

Keywords:
Big data
Travel demand management
Transport modeling
Traffic demand forecasting

ABSTRACT

Since the 2000s, an era of big data has emerged. Since then, urban planners have increasingly applied the theory and methods of big data in planning practice. Recent decades illustrate a rapid increase of the application of big data approaches in transportation, bringing new opportunities for innovation in transport modeling. This article analyzes the theories and methods of big data in traffic demand forecasting. In view of theory, the new models and algorithms are proposed in order to adapt to new big data and response to the limitations of traditional disaggregated approaches. In such approaches, three traffic demand-forecasting methods, the full sample-demand distribution model, the traffic integration model, the model organism protein expression database model, are discussed. Undoubtedly, the development of big data also presents new challenges to travel-demand forecasting methods regarding data acquisition, data processing, data analysis, and application of results. In particular, identifying how to improve approaches to traffic-demand forecasting in the big data era in the Third World will be a challenge to the researchers in the field.

1. Introduction

Transport has become a key issue in relation to sustainable urban development. It supports urban and regional development (Duranton & Turner, 2012), promotes social development (Jones & Lucas, 2012), and enhances human well-being (Storeygard, 2016). In the meantime, it causes problems for cities, such as traffic congestion, air pollution, greenhouse gas emissions, traffic accidents, and other issues. Therefore, travel demand management has been increasingly attracting the attentions of planners and politicians.

Travel demand forecasting is a process to forecast future travel demand based on history and status information of the transportation system and its external system, and to explore the development law and the future trend of the transportation system based on historical experience, objective data and logical judgment. The commonly used "four-stage" method is divided into traffic generation, traffic distribution, traffic mode, traffic flow, traffic status prediction and so on. Traffic prediction methods can be divided into different types according to mathematical methods and calculation processes, such as qualitative and quantitative, linear and nonlinear, dynamic and static, aggregated and disaggregated.

Since *Nature* published the special issue, "Big Data: Science in the Petabyte Era," in 2008, the urban transportation field has become increasingly concerned with big data methods and their applications. Big data refers to a set of data that cannot be crawled, stored, analyzed, or processed by traditional database software tools at a certain stage of development (Manyika et al., 2011). Big data has the following key attributes: (1) volume (i.e., large data volume); (2) variety (i.e., many data types); (3) velocity (i.e., fast data flow) (Bryant, Katz, & Lazowska, 2008); and (4) veracity (i.e., true and accurate data). Additionally, big data also has a value (Yang, 2017) attribute, i.e., it has a huge potential value. Nonetheless, its value density is low, and special tools are necessary to crawl the data. Compared with traditional data analysis methods, the big data method analysis has several unique characteristics. These include that: (1) the data source is extended from the sampled data to the whole data; (2) the single domain data extends to the cross-domain data; (3) the big data analysis surpasses the causal paradigm in the traditional data analysis, and pays more attention to exploring the relationship between data (Barwick, 2011).

In recent years, with the further development of smart city construction, big data has been widely used in urban and rural planning and urban and rural management. This provides a solid foundation for government decision-making, resulting in good value effect (Liu et al., 2015). Especially in the field of transportation, with changes in the living environment and lifestyle brought about by rapid urbanization development, the amount of information has increased dramatically—and the personalization, complexity, and uncertainty of resident's trips have been further strengthened. The travel demand forecasting based on traditional data acquisition methods, such as questionnaire survey and site inspection, has become inefficient and its

* Corresponding author at: Material Management and Unmanned Aerial Vehicle Engineering Institute, Air Force Engineering University, Xi'an 710051, PR China.
E-mail address: yong_zhao_2@163.com (Y. Zhao).

https://doi.org/10.1016/j.cities.2018.04.015
Received 21 December 2017; Received in revised form 28 April 2018; Accepted 30 April 2018
Available online 20 July 2018
0264-2751/

scientific nature has been questioned. Big data-based traffic information acquisition and travel demand forecasting methods have made up the shortcomings of the traditional travel investigation and travel demand forecasting methods. They are widely used in urban intelligent transportation system (ITS), traffic information service systems, and urban transportation planning, and have played a positive role in improving the urban transportation planning techniques.

The role of big data in the traffic management and trip characteristics analysis has been widely discussed by the scholars. However, the theoretical discussion on the application of big data methods in travel demand forecasting is currently limited. This paper discusses theories and methods based on big data-based travel demand forecasting. By comparing trip-based and activity-based forecasting methods, this paper analyzes the impact of big data on these two methods, and theoretically discusses the big data-based travel demand forecasting methods in detail from four aspects: mathematical methods, data acquisition, data processing and data analysis. Among them, the impact of big data on the traffic demand forecasting model was discussed through typical case studies.

2. Current urban travel demand forecasting method: problems and trends

2.1. Current travel demand forecasting methods and problems

Due to the randomness of travel and the openness of the transportation system, both human travel behaviors and the traffic flows of the road systems show uncertainties that bring about many difficulties in predicting traffic. A travel demand forecasting model must overcome the randomness of the transportation system itself. Scholars have tried to develop nearly 30 kinds of travel demand forecasting methods (Yang, Wang, & Guan, 2006) based on different theories or theoretical combinations. Most mathematical deductions are complex and require large data. Before big data methods became widely discussed, many forecasting methods remained solely in the theoretical discussion stage and were difficult to apply to actual work. Exploring more practical travel demand forecasting methods in the technical field has, thus, become one of the challenges in travel demand analysis and the forecasting field.

Since the 1950s, the "four-stage" model in Chicago transportation research, and its core technology based on the trip-based forecasting method, has dominated travel demand forecasting methods. Since the 1970s, the activity-based forecasting methods—proposed by Chapin (1971), Hägerstrand (1970), and Cullen and Godson (1975)—have attracted more attention in recent years.

2.1.1. Trip-based forecasting methods

Trip-based travel demand forecasting methods usually divide the travel forecasting unit in the form of "aggregate" (traffic zone), focusing on the population and the land use of the traffic zone. These methods usually take into account spatial coordination at the urban level, but ignore the actual travel needs and feelings of individual residents. Thus, with poor flexibility and a low degree of refinement for the prediction results, these methods can easily cause problems of unequal traffic distribution (Qin, Zhen, Xiong, & Zhu, 2013).

There are dozens of trip-based forecasting methods. In this paper, we compared and analyzed five of the most basic algorithms and models (see Table 1). These include, first, the Gray system theory (GST); in this system, some information is known, and some information is unknown. The second is Kalman filtering (KF), which has the ability to estimate the state of the dynamic system from a series of incomplete and noise-containing measurements. The third is Chaos theory (CT), which focuses on the orderly structure and regularity of seemingly random phenomena in dynamic systems. The fourth is the artificial neural network (ANN), a mathematical model that mimics the structure and function of biological neural networks. Finally, the fifth is

Table 1
Main trip-based travel demand forecasting methods.

Theory	GST	KF	CT	ANN	SVM
Theory generation	Deng Julong (1982)	Kalman (1960)	Lorenz (1963)	McCulloch and Pitts (1943)	Vapnik and Lerner (1963)
Start of international application		Okutani and Stephanedes (1984)	Disbro and Michael (1989)	Vythoulkas (1993)	
Start of domestic application	Liu (1989)	Hu (1985)		Gong and Li (1995)	Ding Ailing (2002)
Number of data samples	Few	No requirement on data volume	Large data volume	Huge data volume	Large data volume
Linear/nonlinear	Linear	Linear & nonlinear	Nonlinear	Nonlinear	Nonlinear
Dynamic/static	Dynamic & static	Dynamic & static	Dynamic	Dynamic	Dynamic
Spatial scale	Both available	Both available	Both available	Micro-dominant	Micro-dominant
Time scale	Both available	Short-term	Short-term	Short-term	Short-term
Preprocessing/ requirement	High data accuracy	The system mathematical model and noise statistics are known	Chaos shall be distinguished first	The network structure shall be designed, and the number of network layers and the number of nodes in a hidden layer shall be determined	Training samples, selection of kernel function
Advantages	Data loss is offset, and data randomness is reduced	Flexible predictor selection, good real-time, small data storage	The uncertainty of traffic operation is considered and the law behind the phenomenon of random traffic is revealed	Flexible self-organizing learning method, fully distributed storage structure	"Dimensional disaster" is avoided, fast convergence rate; global optimal solution
Limitations	Non-decreasing data; no adaptive change	Unable to truly describe the multi-structure of the transportation system	Cannot be used for long term forecasting	Slow convergence rate, easy to fall into the local minimum	Difficult to handle large training samples and multistructural systems
Optimization/ combination	Neural network (Chen & Wang, 2004), genetic algorithm (Zhou & Wand, 2002)	Wavelet (Li, Hou, & Han, 2007), CT (Shi, 2012), GST	Wavelet (Yang, Jia, He, & Kong, 2005), neural network, time series (Xue & Shi, 2008)	Wavelet (Zhang, 2007), genetic algorithm (Li, Hou, & Han, 2007; Li, Li, Hou, & Yang, 2007), CT (Li, Liu, & Xie, 2011), GST (Wen, Zhou, & Kang, 2010)	Wavelet (Yao, Shao, & Xiong, 2007), genetic algorithm (Wu, Yang, & Liu, 2009), CT (Kang, Li, Zhou, & Zhao, 2011)

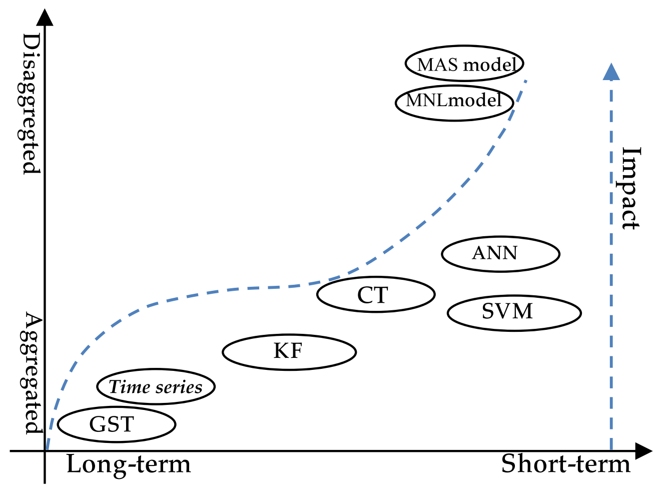

Fig. 1. Trend diagram of impact of big data on the travel demand forecasting models.

support vector machine (SVM)—a machine learning method.

In practical applications, each model shows advantages as well as limitations. As the regularity of traffic flow becomes more indistinct when time interval decreases, the linear models (gray system, etc.) are almost impossible to use for short-term predictions based on traditional survey data. Non-linear models (e.g., neural network, chaos model) often require a large volume of traffic monitoring data as training samples or decision samples. Subject to data acquisition, entry, transmission and storage, these methods require long-time data preprocessing, and show poor real-time nature. Additionally, due to the limitations of data acquisition, many trip-based forecasting models cannot accurately describe the self-similarity, multi-structure and multi-scale features of real networks (Xu & Fu, 2009). They also cannot meet the refinement requirements of travel forecasting.

2.1.2. Activity-based forecasting methods

Taking individuals as the research subject, activity-based forecasting methods are a kind of disaggregated forecasting method that focuses on the reasons for human trips and the "activity-trip" model (Bhat, Guo, Srinivasan, & Sivakumar, 2003). Such methods also explore the causal relationship between activities and trips. The multi-agent system (MAS) model and multinomial logit (MNL) model are activity-based travel demand forecasting methods, and are common nowadays. Their principles and methods are clear and intuitive, and their theoretical basis is more easily to be accepted (Parker, Manson, Janssen, Hoffmann, & Deadman, 2003).

However, activity-based forecasting requires a large volume of individual sample data, which is difficult to achieve under traditional data acquisition and processing techniques. Additionally, due to the constraints of data processing capacity, the current activity-based forecasting methods are used primarily for the division of transportation mode, and are rarely used in traffic generation and traffic distribution prediction (Wang, Huang, & Gong, 2013).

2.2. Development trends of travel demand forecasting methods

During the last two to three decades, the data volume required by travel demand forecasting methods has shown a rapid upward trend, specifically from small sample stage, large sample stage to the current big data stage. The first stage took place in the late 80s. Limited by data acquisition methods at that time, there was relatively few traffic data, and the travel demand forecasting models were designed to achieve a better prediction effect through small samples. Generally, the demand for sample data was small then. Later, with substantial improvements to traffic monitoring technology in the 1990s, the volume of traffic data grew rapidly. The travel demand model that adapted to such a changing came into existence. However, since the beginning of the 21st century, with traditional data acquisition and processing methods, many research results have shown that the improvement of the prediction accuracy brought about by the increase in the number of samples is not significant, and that the real key is the accuracy of single sample information. With the emergence of big data methods, travel demand forecasting methods will usher in a new stage.

3. Big data and travel demand forecasting model

3.1. Big data-based travel demand forecasting

In recent years, under the influence of big data thinking and methods, many new trends have appeared in the field of travel demand forecasting. These trends fall mainly in the following eight aspects. First, the number of forecasting samples has increased significantly. Second, the information volume and accuracy of single samples have improved significantly. Third, there is more research on, as well as applications of, the disaggregated methods; additionally, individuals on the trips receive more attention. Fourth, aggregated methods and disaggregated methods are integrated, such as the integration of the Logit model and neural network model (Yin, Lu, Peng, & Kou, 2008). Fifth, a combination of models has become the development trend. Sixth, there are more micro-level travel forecasting methods and means. Seventh, random traffic data are avoided, and non-traffic data (instead of traffic data) is used for prediction. Finally, the cost limitations of travel demand forecasting are expected to be removed, and data acquisition should, therefore, become easier.

3.2. Impact of big data on travel demand forecasting model

Big data has made a great impact on trip-based and activity-based travel forecasting models, but the size of impact on the different types of travel forecasting models differs (Fig. 1).First, the impact on the short-term travel forecasting is greater than that on the long-term travel forecasting. The number of short-term samples exceeds that of long-term samples and, furthermore, the long-term forecast cannot reflect the real-time value of big data. Second, the impact on nonlinear travel forecasting methods is greater than that on linear travel forecasting methods. Linear travel forecasting methods require fewer data samples and non-linear forecasting methods usually require more data samples. Third, the impact on non-aggregated travel forecasting methods is much greater than that on aggregated methods. The traffic flow data widely used in the aggregated models is a part of the traffic data, however, more valuable are the comprehensive travel data of the individual, which is adapted to the characteristics of disaggregated models that takes individual as the research object.

3.3. Big data-based travel demand forecasting method: typical case

As mentioned above, compared with traditional travel demand forecasting models, big data-based travel demand models show eight major new trends. In this paper, we shall introduce three trends in detail: first, with the acquisition and application of big data, new big data-based travel demand forecasting mathematical methods appeared. Second, with the improvement of data acquisition and processing abilities, the traffic data acquisition, input and demand forecasting have become synchronized, and integration models on the basis of the "four-stage" models have attracted much attention. Third, with the improvement of the magnitude and accuracy of the data samples, the travel demand forecasting for non-motor vehicles (walking, cycling, etc.) has become more accurate.

3.3.1. Total sample demand distribution model

Simini, González, Maritan, and Barabási (2012) published a paper in the renowned international journal *Nature*. They improved the traditional gravity models and built radiation models based on total-population data using a big data approach. The model is mainly characterized by total sample analysis, namely, covering each individual and using more easily accessible population data to predict the demand distribution between regions. Because of the total sample analysis, the model has no parameters. Therefore, it is not necessary to select the survey samples, carry on the origin destination (OD) traffic investigation, or carry on the model parameter estimation and other processes as the gravity model.

Its core formula is as follows:

$$\langle T_{ij}\rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

wherein, T_{ij} is the commute flow between city i and traffic zone j, T_i is the total commute population from zone i, m_i is the population of departure place i, n_j is the population of workplace j, s_{ij} is the population within the circular arc that takes i as the center and the distance between i and j as the radius of the circle (excluding the populations in the departure place and the destination).

Wherein, $T_i = \mathrm{m_i}(N_c / N)$, N_c and N are the total commute population and the total population of the city, respectively.

The author used the data of New York County to carry on the empirical analysis, compared to the demand distribution prediction result of the radiation model with that of the gravity model, and found that the radiation model method can more accurately reflect the actual traffic distribution than the traditional gravity model method (Fig. 2).

Additionally, Yan, Wang, Gao, and Lai (2017) established a generic model that can accurately predict various individual and collective flow patterns in different countries and cities at different spatial scales, such as scaling behavior and trajectory patterns. This model can explain the universal basic mechanism of various human mobility behaviors.

3.3.2. Transportation integration model

The traditional travel demand forecasting is based on the existing traffic survey. As sample selection, data acquisition, questionnaire entry, and analysis during the investigation all lag behind, there is a great margin of error in travel demand forecasting. Moreover, the current travel forecasting still focuses on the balance between the network distribution of traffic flow (congestion level) and the choice of traffic modes. In fact, the level of congestion will also affect the trip generation (whether to take a trip?) and trip distribution (where to go?). The big data methods provide planners with the opportunity to collect traffic statuses more accurately. With the support of big data, transportation integration models that can reflect real time information will become widely used.

For example, trip distribution-mode selection-traffic assignment integration model:

$$\mathrm{MinimiseZ} = \eta \sum_{\alpha} \int_0^{\mathrm{V}_\alpha} \mathrm{c}_\alpha(\upsilon)\mathrm{d}\upsilon + \sum_{\mathrm{ij}} \mathrm{c}_{\mathrm{ij}}^{\mathrm{b}}\mathrm{T}_{\mathrm{ij}}^{\mathrm{b}} + \frac{1}{\lambda}\sum_{\mathrm{ijm}} \mathrm{T}_{\mathrm{ij}}^{\mathrm{m}} \log \mathrm{T}_{\mathrm{ij}}^{\mathrm{m}}$$

Restrictions:

$$\sum_{\mathrm{ijr}} \eta \mathrm{T}_{\mathrm{ij}}^{\mathrm{c}} \delta_{\mathrm{ijr}}^{\alpha} = \mathrm{V}_\alpha$$

$$\sum_{\mathrm{in}} \mathrm{T}_{\mathrm{ij}}^{\mathrm{k}} = \mathrm{D_j}$$

wherein, D_j is the traffic generating amount in region j; $T_{ij}{}^{k}$ is the traffic volume of traffic mode k (or b, c) from i to j; $c_{ij}{}^{b}$ is the transportation cost (time) of traffic mode b from i to j; c_α is the path cost (time + cost) of path a at traffic flow v; V_α is the traffic flow of section a; $\delta_{ijr}{}^{\alpha}$ is 1 (if traffic path r between i and j uses road section a) or 0 (if road section a is not used); η is the vehicle ride rate.

In the empirical application, Lu, Sun, and Qu (2015) used big data and genetic algorithms to identify and analyze real-time traffic flow. Wang, González, Hidalgo, and Barabási (2009) used mobile information to detect and predict the number of travelers' destinations to be visited.

3.3.3. Big data-based walking and cycling travel demand forecasting model

Due to limitations of aggregated methods and traffic zone scales, traditional travel demand forecasting methods are often not accurate for the demand prediction of non-motor vehicles such as walking and cycling. In China's cities, the travel survey does not count walking within 400 m, nor does it calculate moving within the zone and the factory. Therefore, local traffic microcirculation data are insufficient, which may bias the pedestrian traffic forecasting in the city. With the implementation of urban strategies such as green traffic, slow traffic and livable city, it is objectively necessary to improve the accuracy of traffic forecasts on walking and cycling.

Big data methods provide an opportunity for this. For example, in the Portland pedestrian travel demand forecasting, the Baltimore model was improved and the model of MoPeD 2.0 (model organism protein expression database) was proposed. The model used a spatial analysis unit in a smaller scale (80-m square grid) and the new walking environment measurement method supported by big data method, and formed an independent pedestrian planning tool to enhance the ability to predict the needs of walking (Patrick, Christopher, & Robert, 2014) in metropolitan planning organizations (MPOs) travel model. Gupta, Vovsha, Kumar, and Subhani (2014) used a big data method to construct a cycling route selection model (Subhani, Stephens, Kumar, & Vovsha, 2013) in the study of Ottawa-Gatineau (a pedestrian-friendly community).

4. Types and methods of data acquisition

In the big data era, traffic data expanded from the original single structural static data set to the multi-source (Shao, Zhao, & Wu, 2006), multi-status, multi-structural data set in combination of both static and dynamic data. The traffic data acquisition no longer includes split time steps such as investigator designing a questionnaire, sample selection, questionnaire issuance, and data entry. It now achieves integration and synchronization of collection and entry of travel information and social economic attribute information of traffic users. Traffic data acquisition has changed from its original sample path observation and data entry to whole network data collection. It includes not only transportation data such as traffic flow, fleet length, vehicle type, vehicle traveling direction, travel time, instantaneous speed, travel speed, etc., but also includes attribute data of individual person or car. The traffic data acquisition process is as follows (Fig. 3).

The traffic data acquisition targeting people mainly includes bus IC card data and mobile phone GPS data, which includes information such as public traffic origin-destination (OD) data, bus corridor flow and passenger traffic of each site. At present, the cooperation between the traffic information center and a mobile phone service provider has become the main mode of big data application. The Beijing Traffic Information Center uses the mobile phone GPS data provided by China Mobile to collect the traffic travel information and analyze the characteristics of the residents' travel distribution.

The data acquisition targeting motor vehicles is similar to that targeting people, which collects vehicles' real-time information and obtains the traffic conditions of each road through the GPS equipped on the electronic license plates of vehicles. The traditional traffic data acquisition methods, based on floating vehicles, have been applied at home and abroad. For example, in 2005, Beijing used the data of the taxis to reflect the operation of the road network of the whole city. By 2014, the number of taxis has increased to 30,000. However, in the future, with the popularity of electronic license plates, the acquisition, processing, and analysis of the total sample driving states will become available with the support of big data.

In the era of big data, the unstructured data in the transportation field is also worthy of attention. Useful information from data sources,

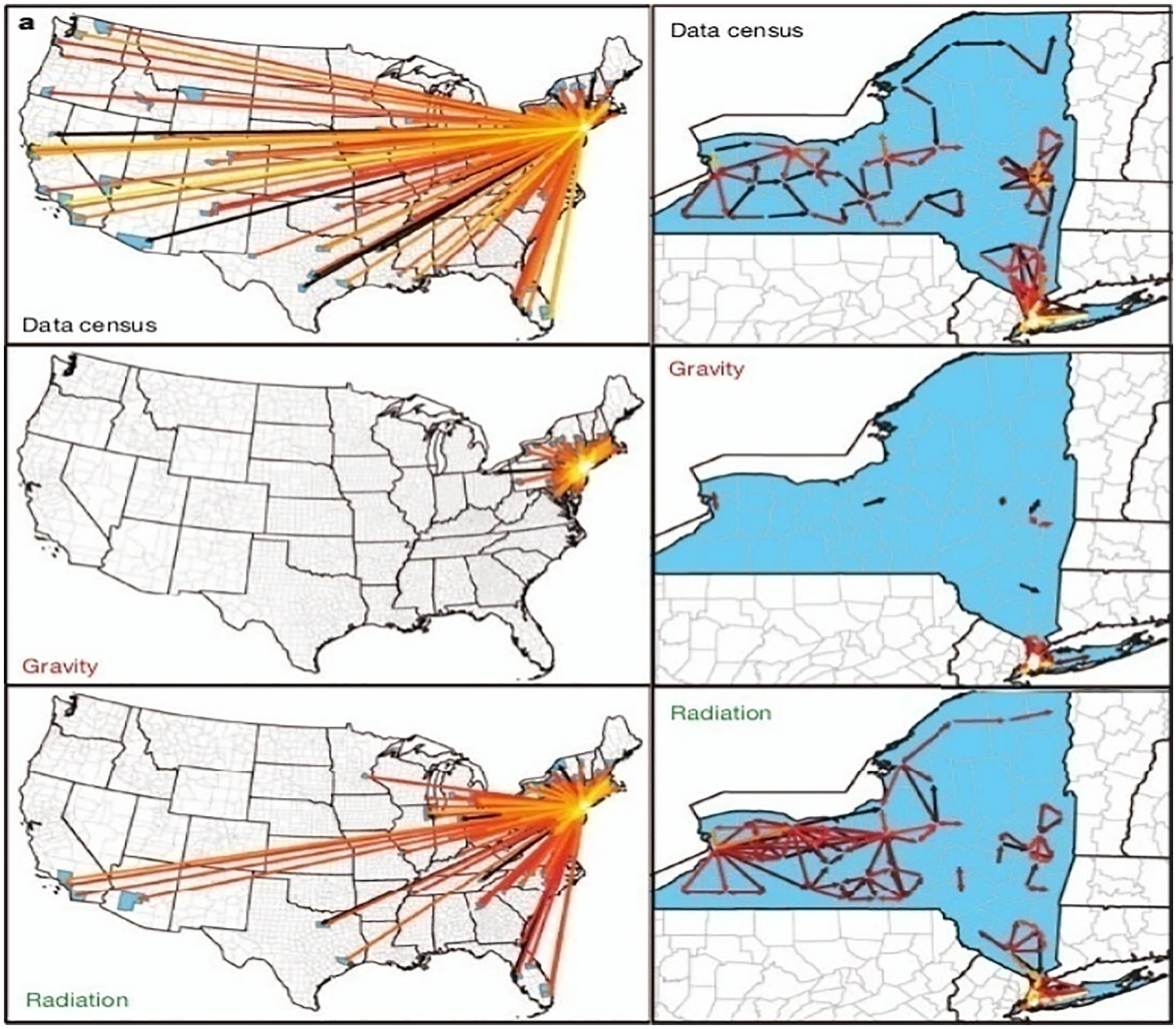

Fig. 2. Comparison of prediction results between gravity model and radiation model.
Notes: (1) Source: Simini et al. (2012). (2) The upper part of the left figure shows the actual trip distribution of residents outside the county in the census. The right figure shows the actual trip distribution of the residents between traffic zones within the county area. The middle figure shows the simulation result of the gravity model. The lower figure shows the simulation result of the radiation model. From the figures, it is evident that the simulation result of the radiation model is closer to the actual trip distribution.

such as Web click flows, documents, social networks, Internet of things, phone call logs, videos, photos (Zhao, 2013), RFID data and so on are captured to study people's travel behaviors. Especially from social networks, such as microblog, WeChat and other platforms, the trip statuses reflect accurately real-time and dynamic nature, and break the fixed limitations of traffic data samples collected in traditional questionnaires. Mark and Nick (2011) extracted Twitter data for four months from 9223 users of Leeds in the UK. Then, in combination with GIS technology, they used intelligent models to determine the basic behaviors of urban life, education, work, entertainment and shopping, and the travel patterns closely linked to these behaviors. Additionally, the spatial GIS data such as vehicles, drivers' data and population, land use, remote sensing, road network, road planning, and transportation facilities network have been collected as a prediction basis. Finally, the traffic information database and the basic information database related to transportation field are formed.

5. Data processing

5.1. Data storage and transmission

Big data has brought great challenges to fields of data storage and transmission. In south China's Guangzhou City, the traffic comprehensive processing service platform has more than 1.2 billion pieces of newly added urban traffic data records every day. The amount of data generated each day reaches 150 to 300GB. Distributed storage and object storage architectures have to be more efficient in order to store huge amounts of data and support fast and concurrent access to data at a lower cost. For specific data forms, scholars have proposed a method of compressing data. For example, Ma, Liu, and Sun (2011) compressed and coded the mass GPS data of urban traffic, with a compression ratio of no less than 60%, removing redundant information and improving the real-time nature of mass information services.

Video-based traffic data are often larger. The transmission modes based on 3G, 4G, WIFI, and other mobile Internet options are more advanced. This can dynamically adjust video frame rates based on the characteristics of mobile Internet such as instable signal, narrow bandwidth, and low speed, to achieve rapid data transmission and meet the real-time requirements by the users to the maximum extent (Wang, Xiong, Zuo, & Meng, 2012).

More than 85% of the data collected at present are unstructured and semi-structured data (Li, 2012). It is not suitable to use a traditional relational database—and a non-relational database is required. At present, non-relational databases mainly have four data storage types:

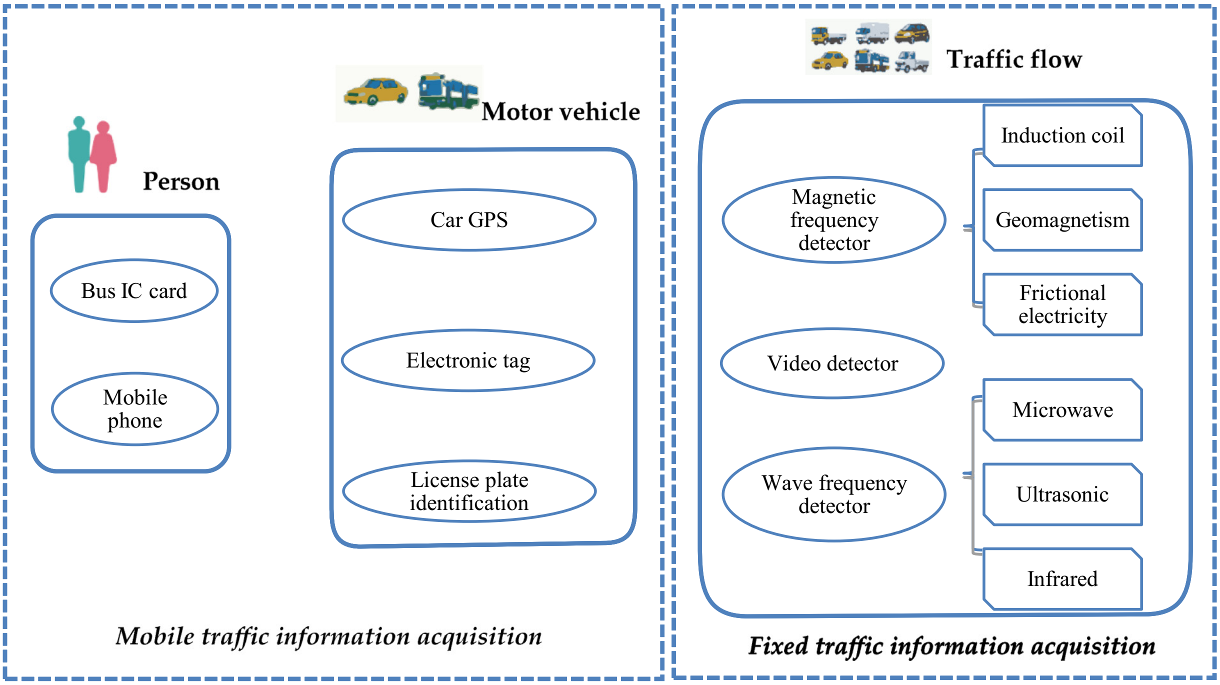

Fig. 3. Traffic information acquisition modes.

key-value, column-oriented, document store and graph database. Each type of storage will solve the corresponding problem that the relational database cannot solve. In practical applications, these kinds of situations will combine to achieve the corresponding functions. For example, Amazon DynamoDB integrates NoSQL's Document stores and Key-value stores. Although relational database is still dominate, due to the "explosive" growth of pb-level unstructured data, the potential of non-relational databases is enormous.

5.2. Data preprocessing

The actual data collected often has an inconsistent or incomplete structure with the data needed for big data analysis processes. The traffic big data preprocessing process can initially organize and comb the collected data through data extraction, conversion and loading, thus improving the quality and efficiency of big data analysis. Instead, the instrument and equipment will carry on such processing as integration, identification, and screening, and then transmit the useful information to the platform, which can also ease the pressure of transmitting and storing big data. Abella, Ortiz-De-Urbina-Criado, and De-Pablos-Heredero (2017) propose a new model which selects the most valuable data to the user by evaluating the re availability of data, reusing value, and the impact of generating services on the economy and society.

5.3. Data analysis

The analysis of big traffic data are divided into real-time analysis and non-real-time analysis, according to different real time requirements (Fig. 4). In many cases, simple data are more valuable for real-time decision-making. Real-time big traffic data analysis requires simplification of processing steps and application of stylized information analysis methods to obtain publicly accepted traffic decision information such as traffic congestion index. However, non-real-time big traffic data analysis focuses more on research, such as analysis of residents' time and space behaviors based on individual traffic data. The accuracy of the travel demand forecasting model can also be checked by big traffic data. Big data also provides a basis for traffic data mining: The specific mathematical models and algorithms are used to analyze big traffic data to explain the relationship, patterns and trends hidden behind the data, and to provide new knowledge for decision-making.

Since Eric Schmidt, Google's CEO, put forward the concept of cloud computing for the first time in 2006, it has derived a series of calculation methods such as fog calculation, deuterium calculation, and edge calculation. Darwish and Bakar (2018) believe that the big data analysis phase needs to be distributed between cloud computing and fog-computing layers. The fog-computing node can consider the user's location awareness and mobility requirements at the service terminal and is an effective solution for instant big data analysis. For non-immediate traffic big data analysis, cloud computing that uses more computing power and greater storage capacity performs better than fog computing.

6. Conclusion

This paper analyzes and summarizes two traffic demand forecasting methods (trip-based and activity-based) and focuses on three major new trends presented by the traffic demand model based on big data. The first is a traffic prediction model with big data characteristics, the second is the integrated model based on the "four-phase" model, and the third is the more accurate traffic demand prediction model for non-motorized (walking, bicycle, etc.) vehicles. From the system perspective, some methods and mechanisms of big data for data acquisition and data processing were discussed. At the data acquisition stage, the types and methods of big data acquisition were combed. At the data processing stage, the big data pre-processing technology, unstructured data storage, and some typical big data calculation methods were analyzed.

Big data brings new thinking to the field of travel demand forecasting methods and expands the development directions of the travel demand forecasting methods. The study of big data-based transportation could enhance urban transportation planning techniques. Urban planners and policy-makers should pay more attentions to big data and related approaches. Undoubtedly, the development of big data also presents new challenges to the travel demand forecasting methods from the aspects of data acquisition, data processing, data analysis, and application of results. First, data are closely related to mathematical methods. The big data-based traffic models need to consider new mathematical measurement methods and build new models and core

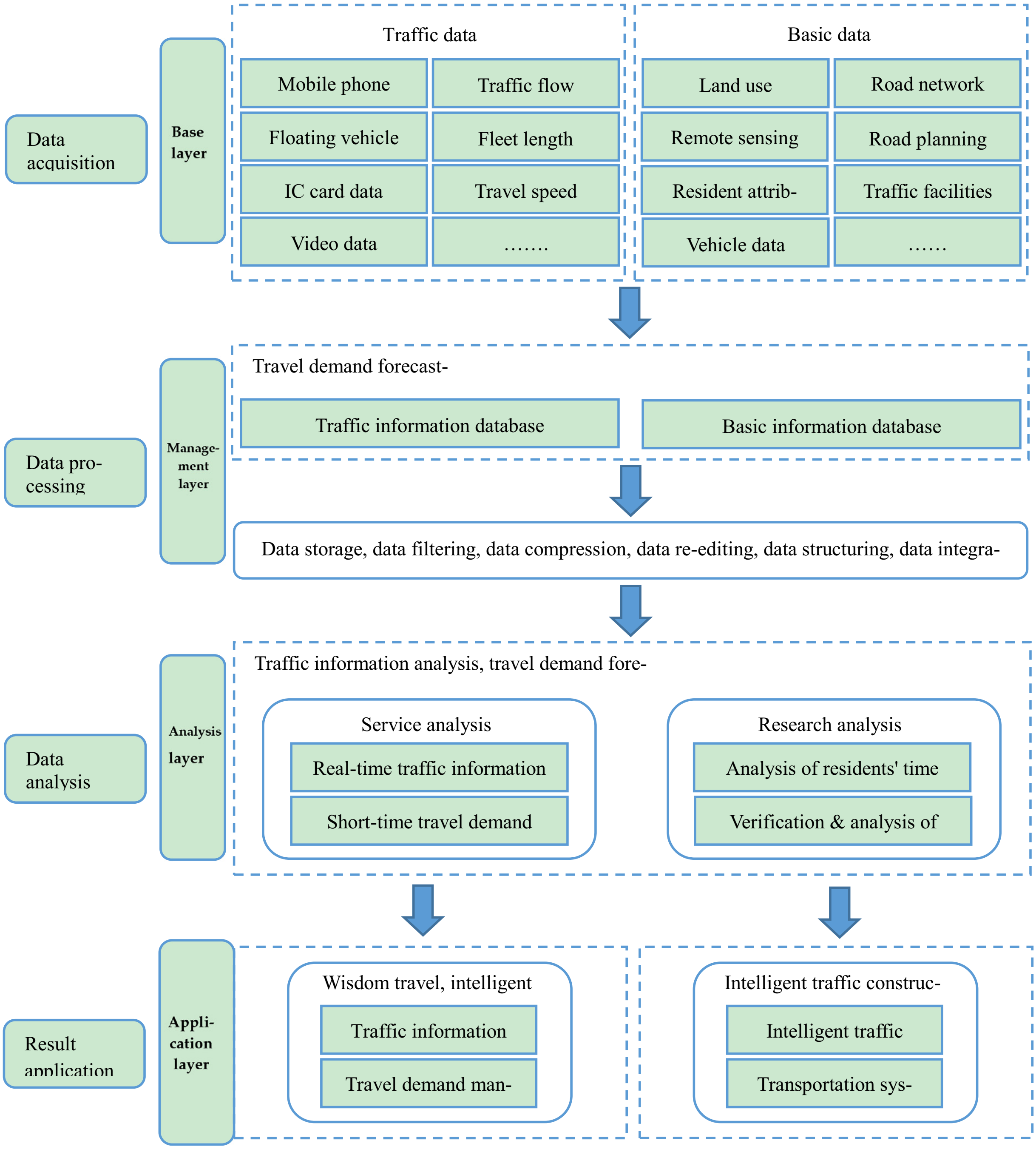

Fig. 4. Acquisition, processing and analysis of traffic big data.

algorithms to adapt to the requirements of big data. Second, it is not yet clear whether big data is a method or a theory. However, big data cannot completely replace traditional travel demand theories. A future challenge is to identify how to explain theoretically big data-based travel demand forecasting methods. Third, in terms of specific applications, it will face a series of practical problems. These may include integration of data resources, formation of a comprehensive database, the current traffic-related data acquisition systems are not interconnected, the integration and utilization of multivariate data are low; it is necessary to enhance the compatibility of various data acquisition methods, and to strengthen the processing capacity of non-structural data. Fourth, big data and related approaches develop slowly in developing countries. Due to a shortage of techniques and financial support in these countries, the application in the traffic models may lag behind in big data generation and data acquisition itself. Improving approaches for traffic demand forecasting in the big data era in the Third World will be a challenge to the researchers in the field.

Acknowledgements

The research is funded by 2017GY-049.

References

Abella, A., Ortiz-De-Urbina-Criado, M., & De-Pablos-Heredero, C. (2017). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities, 64*, 47–53.

Barwick, H. (2011). IIIS: The 'four Vs' of big data. *Implementing Information Infrastructure Symposium*http://www.computerworld.com.au/article/396198/iiis_four_vs_big_data/.

Bhat, C. R., Guo, J., Srinivasan, S., & Sivakumar, A. (2003). *Guidebook on activity-based travel demand modeling for planners.* Austin, TX: Univ. Texas at Austin (4080-P3).
Bryant, R., Katz, R. H., & Lazowska, E. D. (2008). Big-data computing: Creating revolutionary breakthroughs in commerce, science and society. http://www.cra.org/ccc/files/docs/init/Big_Data.pdf.
Chapin, F. S. (1971). Free time activities and quality of urban life. *Journal of the American Institute of Planners, 37*, 411–417.
Chen, S., & Wang, W. (2004). Grey neural network forecasting for traffic flow. *Journal of Southeast University (Natural Science Edition), 34*(4), 541–544.
Cullen, I. G., & Godson, V. (1975). Urban networks, the structure of activity patterns. *Progress in Plannin, 4*, 1–96.
Darwish, T. S., & Bakar, K. A. (2018). Fog based intelligent transportation big data analytics in the Internet of vehicles environment: Motivations, architecture, challenges and critical issues. *IEEE Access, 6*, 15679–15701.
Deng Julong (1982). Grey control system. *The Journal of Huazhong University of Science and Technology, 3*, 9–18.
Ding Ailing (2002). Prediction of traffic flow time series based on statistical learning theory. *Computer and Communications, 20*(2), 27–30.
Disbro, J. E., & Michael, F. (1989). Traffic flow theory and chaotic behavior. *No. Special report 91*. New York (State): Dept. of Transportation.
Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies, 79*(4), 1407–1440.
Gong, Z., & Li, Z. (1995). Applying artificial neural network technology to estimate urban road traffic volume. *System Engineering Theory and Practice, 15*(5), 32–36.
Gupta, S., Vovsha, P., Kumar, R., & Subhani, A. (2014). Incorporating cycling in Ottawa-Gatineau travel forecasting model. *TRB Conference on Innovations in Travel Modeling, Transportation Research Board*.
Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science, 24*, 7–24.
Hu, L. (1985). Application of Calman filtering method in public traffic flow forecasting. *System Engineering, 3*, 003.
Jones, P., & Lucas, K. (2012). The social consequences of transport decision-making: Clarifying concepts, synthesising knowledge and assessing implications. *Journal of Transport Geography, 21*, 4–16.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering, 82*, 34–45.
Kang, H., Li, M., Zhou, P., & Zhao, Z. (2011). Prediction of traffic volume of SVM based on chaos efficient genetic algorithm. *Journal of Wuhan University of Technology (Transportation Science and Engineering Edition), 35*(4), 649–653.
Li, G. (2012). *The scientific value of big data research*.
Li, J., Hou, X., & Han, Z. (2007). Application of Kalman filter and wavelet in traffic prediction. *Journal of Electronics and Information Technology, 29*(3), 725–728.
Li, J., Li, Q., Hou, H., & Yang, L. (2007). Traffic flow prediction based on the wavelet neural network with genetic algorithm. *Journal of Shandong University (Engineering Edition), 37*(2), 109–112.
Li, S., Liu, L., & Xie, Y. (2011). Chaotic prediction for short-term traffic flow of optimized BP neuralnetwork based on genetic algorithm. *Control and Decision, 26*(10), 1581–1585.
Liu, Y. (1989). Application of grey model in transportation forecasting. *Shan Dong Jiao Tong Ke Ji, 2*, 60–68.
Liu, X., Song, Y., Wu, K., Wang, J., Li, D., & Long, Y. (2015). Understanding urban China with open data. *Cities, 47*, 53–61.
Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20*(2), 130–141.
Lu, H. P., Sun, Z. Y., & Qu, W. C. (2015). Big data-driven based real-time traffic flow state identification and prediction. *Discrete Dynamics in Nature and Society,* 2015.
Ma, Q., Liu, W., & Sun, D. (2011). A high-speed compression scheme for vast quantities of GPS data. *Journal of Sichuan University (Engineering Science Edition), 43*(Supplement 1), 123–128.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
Mark, B., & Nick, M. (2011). Microscopic simulations of complex metropolitan dynamics. http://eprints.ncrm.ac.uk/2051/1/complex_city_paper[1].pdf.
McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.
Okutani, I., & Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. *Transport. Res. Part B, 18B*, 1–11.
Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337.
Patrick, A., Christopher, M., & Robert, J. (2014). Introducing MoPeD 2.0: A model of pedestrian demand, integrated with trip-based travel demand forecasting models. *The 5th Transportation Research Board Conference on Innovations in Travel Modeling.* Baltimore: ITM.
Qin, X., Zhen, F., Xiong, L., & Zhu, S. (2013). Methods in urban temporal and spatial behavior research in the Big Data Era. *Advances in Geography, 32*(9), 1352–1360.
Shao, C., Zhao, Y., & Wu, Y. (2006). Study on road traffic data acquisition technology. *Modern transportation technology*.
Shi, M. (2012). *Study on short-term traffic flow forecasting method based on Kalman filter.* Chengdu: Southwest Jiaotong University.
Simini, F., González, M. C., Maritan, A., & Barabási, A. L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96.
Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa. *The Review of Economic Studies, 83*(3), 1263–1295.
Subhani, A., Stephens, D., Kumar, R., & Vovsha, P. (2013). Incorporating cycling in Ottawa-Gatineau travel forecasting model. *The 2013 Conference and Exhibition of the Transportation Association of Canada-Transportation: Better-Faster-Safer*.
Vapnik, V. N., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control, 24*, 774–780.
Vythoulkas, P. C. (1993). Alternative approaches to short term traffic forecasting for use in driver information systems. *Transportation and traffic theory, 12*, 485–506.
Wang, F., Xiong, J., Zuo, Q., & Meng, X. (2012). *Research on the real-time transmission of high-capacity traffic video information based on mobile internet*.
Wang, P., González, M. C., Hidalgo, C. A., & Barabási, A. L. (2009). Understanding the spreading patterns of mobile phone viruses. *Science, 324*(5930), 1071–1076.
Wang, P., Huang, Z. R., & Gong, H. (2013). Transportation engineering in the big data era. *Journal of the University of Electronic Science and Technology of China, 42*(6), 806–816.
Wen, S., Zhou, P., & Kang, H. (2010). Study of combining traffic forecasting model based on gray theory and BP neural network. *Journal of Dalian University of Technology, 50*(4), 547–550.
Wu, J., Yang, S., & Liu, C. (2009). Short-time load forecasting method for support vector machine based on genetic algorithm optimization. *Journal of Central South University (Natural Science Edition), 40*(1), 180–184.
Xu, L., & Fu, H. (2009). *Theories and methods of traffic information intelligent prediction.* Beijing: Science Press.
Xue, J., & Shi, Z. (2008). Study on short-term traffic flow prediction based on chaotic time series analysis. *Journal of Transportation Systems Engineering and Information, 8*(5), 68–72.
Yan, X. Y., Wang, W. X., Gao, Z. Y., & Lai, Y. C. (2017). Universal model of individual and population mobility on diverse spatial scales. *Nature Communications, 8*(1), 1639.
Yang, L., Jia, L., He, L., & Kong, Q. (2005). Study on traffic flow prediction algorithm based on chaotic wavelet network. *Journal of Shandong University (Engineering Edition), 35*(2), 649–653.
Yang, Z. (2017). *Smart city: Big data, the Internet of things and cloud computing.* Beijing: Tsinghua University Press.
Yang, Z., Wang, Y., & Guan, Q. (2006). A short-term traffic flow forecasting method based on support vector machine. *Journal of Jilin University (Engineering Science Edition), 36*(6), 881–884.
Yao, Z., Shao, C., & Xiong, Z. (2007). Study on short-time traffic flow combination forecasting method based on wavelet packet and least squares support vector machine. *China Management Science, 15*(1), 64–68.
Yin, Y., Lu, H., Peng, X., & Kou, F. (2008). Study on traffic mode split based on LOGIT model and BP neural network. *Gansu Science and Technology, 24*(2), 76–78.
Zhang, X. (2007). The forecasting approach for short-term traffic flow based on wavelet analysis and neural network. *Information and Control, 36*(4), 467–470.
Zhao, G. (2013). *Big data technology and application practice guide.* Beijing: Electronic Industry Press.
Zhou, M., & Wand, H. (2002). A gray GM (1, 1) model based on genetic algorithm. *Journal of Nanchang University (Natural Science Edition), 26*(4), 331–333.