

Research Article

Real-Time Corrected Traffic Correlation Model for Traffic Flow Forecasting

Hua-pu Lu,¹ Zhi-yuan Sun,¹ Wen-cong Qu,¹ and Ling Wang^{1,2}

¹*Institute of Transportation Engineering, Tsinghua University, Beijing 100084, China*

²*National Defense Transportation Department, Military Transportation University, Tianjin 300161, China*

Correspondence should be addressed to Ling Wang; rainvgyi@163.com

Received 2 August 2014; Accepted 27 February 2015

Academic Editor: Emilio Insfran

Copyright © 2015 Hua-pu Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper focuses on the problems of short-term traffic flow forecasting. The main goal is to put forward traffic correlation model and real-time correction algorithm for traffic flow forecasting. Traffic correlation model is established based on the temporal-spatial-historical correlation characteristic of traffic big data. In order to simplify the traffic correlation model, this paper presents correction coefficients optimization algorithm. Considering multistate characteristic of traffic big data, a dynamic part is added to traffic correlation model. Real-time correction algorithm based on Fuzzy Neural Network is presented to overcome the nonlinear mapping problems. A case study based on a real-world road network in Beijing, China, is implemented to test the efficiency and applicability of the proposed modeling methods.

1. Introduction

It is of practical significance to predict traffic flow quickly, precisely, and timely. Short-term traffic flow forecasting provides an important basis for traffic guidance and control. Existing studies of short-term traffic flow forecasting can be classified into six categories in transportation literature:

- (a) linear system theory based models, such as Autoregressive Integrated Moving Average (ARIMA) [1] and Kalman Filtering model [2];
- (b) data mining based models, such as Neural Network [3], Nonparametric Regression [4], and Support Vector Machine [5];
- (c) nonlinear system theory based models, such as Wavelet Analysis [6], Catastrophe Theory [7], and Chaos Theory [8];
- (d) simulation based models [9];
- (e) combination model based models [10];
- (f) the other models.

In the era of big data, it brings both opportunities and challenges to short-term traffic flow forecasting. During data

processing, traffic big data meets the same difficulties with the general big data, such as capture, storage, search, sharing, analytics, and visualization. Therefore, short-term traffic flow forecasting method needs to have the capacity to deal with traffic big data. Traffic big data holds several characteristics, such as temporal correlation, spatial correlation, historical correlation, and multistate. Considering the advantages of traffic big data, data-driven based mathematical models can be set up. The physical meaning of these models can be described clearly. In addition, we can put forward real-time correction algorithm to improve the accuracy of traffic flow forecasting.

However, taking into account all the present researches in this field, there is still a lack of consideration of traffic big data and real-time correction for traffic flow forecasting. Further researches remain to be conducted on the direction of traffic big data analysis. In this paper, the method of short-term traffic flow forecasting is proposed in detail. The remainder of this paper is organized as follows. Section 2 presents basic mathematical model. In Section 3, real-time corrected traffic correlation model is established. A case study based on a real-world road network is carried out in Section 4 to demonstrate the performance and applicability of the proposed method. Finally, conclusions are drawn in Section 5.

2. Basic Mathematical Model

2.1. Big Data Driven Based Traffic Correlation Model. Traffic big data has a strong temporal-spatial-historical correlation as follows.

- (i) In the temporal series, the traffic flow of last moment can be regarded as a continuation of current traffic flow. Dynamic traffic flow data continuously change over time with a certain trend.
- (ii) In the spatial series, the traffic flow of downstream sections can be seen as a continuation of the upstream traffic flow. There exists a spatial association between traffic flow data of neighboring junctions or sections and that of target junctions or sections.
- (iii) In the historical series, the traffic demand characteristics determine that traffic flow characteristics of the same day in the same period are similar. The law of traffic flow cycle is especially evident.

Therefore, the basic form of traffic correlation model [11] is expressed as

$$V_s(t) = \gamma_T V_T(t) + \gamma_S V_S(t) + \gamma_H V_H(t), \quad (1)$$

s.t. $\gamma_T + \gamma_S + \gamma_H = 1, \quad \gamma_T, \gamma_S, \gamma_H \geq 0,$

where $V_s(t)$ is the traffic flow parameter of section s at time t , representing flow $q_s(t)$, speed $v_s(t)$, or occupancy $\rho_s(t)$. $V_T(t)$, $V_S(t)$, and $V_H(t)$ are the estimated value of $V_s(t)$. γ_T , γ_S , and γ_H are coefficients of these three variables.

$V_T(t)$ is calculated by temporal correlation analysis, generally based on Regression Analysis Model [12]. $V_S(t)$ is calculated by spatial correlation analysis, generally based on Neighbor Regression Model [13]. $V_H(t)$ is calculated by historical correlation analysis, generally based on Discrete Fourier Transform Model [14].

Thus, simplified equation of traffic correlation model is obtained:

$$V_s(t) = \sum \sum \omega_{jd} V_j(t-d), \quad (2)$$

s.t. $\sum \omega = 1, \quad \omega > 0, \quad \eta \in N^*,$

where ω_{jd} is regression coefficient of $V_j(t-d)$. η is the number of ω_{jd} .

2.2. Correction Coefficients Optimization Algorithm. It is found that the speed and accuracy of data processing are both important for big data driven method. To improve the speed of traffic correlation model, the number of unknown variables in formulation (2) must be reduced. However, variables reduction may decrease the accuracy of the model.

Therefore, this paper defines a threshold of computing speed and derives the maximum of acceptable number of variables. Thus, the number of unknown variable η can be achieved. The correlation coefficients (r_{jd}) between

the studied section s and the related section j are used to choose variables, the number of which is η :

$$r_{jd} = \left(\sum_{t=1}^{n-d} V_s(t) V_j(t-d) - \frac{1}{n-d} \sum_{t=1}^{n-d} V_s(t) \sum_{t=1}^{n-d} V_j(t-d) \right) \cdot \left(\left[\sum_{t=1}^{n-d} V_s^2(t) - \frac{1}{n-d} \left(\sum_{t=1}^{n-d} V_s(t) \right)^2 \right] \times \left[\sum_{t=1}^{n-d} V_j^2(t-d) - \frac{1}{n-d} \left(\sum_{t=1}^{n-d} V_j(t-d) \right)^2 \right] \right)^{-1/2},$$

s.t. $d \geq 0.$ (3)

In addition, for $V_j(\cdot)$, a unique value of $V_j(t-d)$ is determined. Therefore, a lot of variables are reduced. When the max of correlation coefficient (r_j) corresponding to each section j is calculated, corresponding time delay (d) and the unique value of $V_j(t-d)$ are obtained:

$$r_j = \max r_{jd}. \quad (4)$$

If the value of r_j is large enough, the variable $V_j(t-d)$ is preserved. Otherwise, the variable $V_j(t-d)$ is reduced. After several data tests, it is found that the values of r_j are not very different with relatively high values when η is less than 4; the rapid decay of r_j is observed with relatively low values when η is more than 8. Therefore, $\eta \in [4, 8]$.

Variables reduction makes ω_{jd} meaningless. So, a new variable ω_j , which is normalized r_j , is selected to replace the variable ω_{jd} :

$$\omega_j = \frac{r_j}{\sum_{j=1}^{\eta} r_j}, \quad \sum_{j=1}^{\eta} \omega_j = 1, \quad (5)$$

$\omega_j > 0.$

Since the alternative process will bring some errors, which are likely to be systematic, a linear correction algorithm is present. Two correction variables φ and ε are introduced for calibration error. Simplified traffic correlation model is

$$\min \quad |V_s(t) - \widehat{V}_s(t)|,$$

s.t. $V_s(t) = \varphi \sum_{j=1}^{\eta} \omega_j V_j(t-d) + \varepsilon,$ (6)

where $\widehat{V}_s(t)$ is the actual value of $V_s(t)$.

3. Model Improvement

3.1. Real-Time Correction Problem Statement. Basic mathematical model can be used for traffic flow forecasting. The error of traffic flow forecasting is written as

$$\begin{aligned} \Delta(t) &= \widehat{V}_s(t) - V_s(t) \\ &= \widehat{V}_s(t) - \left[\varphi \sum_{j=1}^{\eta} \omega_j V_j(t-d) + \varepsilon \right]. \end{aligned} \quad (7)$$

Therefore, $\widehat{V}_s(t)$ is achieved:

$$\widehat{V}_s(t) = \varphi \sum_{j=1}^{\eta} \omega_j V_j(t-d) + \varepsilon + \Delta(t). \quad (8)$$

The variation range of $\Delta(t)$ reflects the accuracy of traffic flow forecasting. To improve the accuracy of traffic flow forecasting, $\Delta(t)$ is made as a compensation variable for $V_s(t)$. Then, $\widehat{V}_s(t)$ is replaced by $V_s(t)$ in formulation (8):

$$V_s(t) = \varphi \sum_{j=1}^{\eta} \omega_j V_j(t-d) + \varepsilon + \Delta(t). \quad (9)$$

For different traffic state, the propagation of traffic congestion is different. So, the temporal-spatial-historical correlation variables are dynamic. As shown in Section 2, basic mathematical model for traffic flow forecasting is put forward based on temporal-spatial-historical correlation characteristic of traffic big data. However, the multistate characteristic is largely ignored; the temporal-spatial-historical correlation variables are seen as static variables. Although a linear fitting method is used to improve accuracy, the error which is the main part of $\Delta(t)$ still exists.

Effective analysis of traffic correlation model is shown in Figure 1. The error of traffic congestion stage is larger than that of traffic smooth stage. The error of last moment may affect the error of current moment.

$\Delta(t)$ responds to changes with traffic flow state and temporal series. It is difficult for mathematical model to describe the characteristics of $\Delta(t)$. This paper tries to present real-time correction algorithm based on nonlinear mapping.

It assumes that $\Delta(t)$ is influenced by level of service at time $t-1$ ($\text{LOS}(t-1)$) and $\Delta(t-1)$. Then, the nonlinear mapping can be written as

$$f: [\Delta(t-1), \text{LOS}(t-1)] \longrightarrow \Delta(t). \quad (10)$$

3.2. Real-Time Correction Algorithm. The main goal of real-time correction algorithm is to calculate the error term ($\Delta(t)$) in formulation (9). Because of nonlinear mapping, Fuzzy Neural Network (FNN) can be used to overcome this problem. The structure of Fuzzy Neural Network is shown in Figure 2.

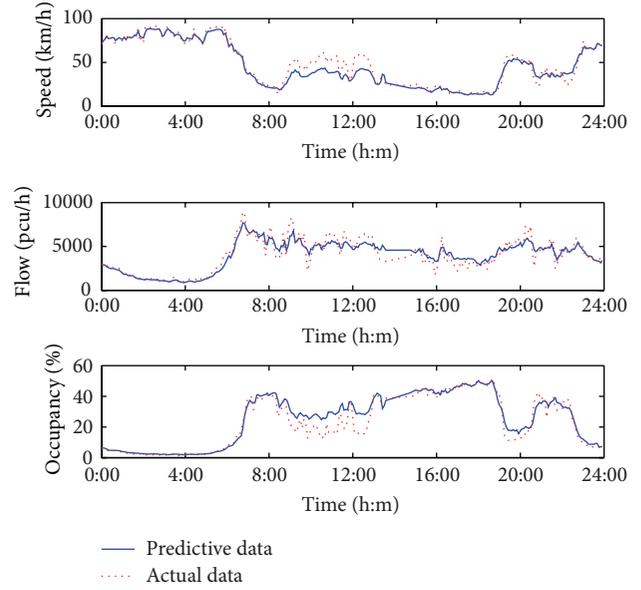


FIGURE 1: Effective analysis of traffic correlation model.

Every input unit in Fuzzy Neural Network is corresponding to certain fuzzy subset of the input variables, which are $\text{LOS}(t-1)$ and $\Delta(t-1)$. The format of input signal is

$$\begin{aligned} &[\mu_A(\text{LOS}), \mu_B(\text{LOS}), \mu_C(\text{LOS}), \mu_D(\text{LOS}), \mu_E(\text{LOS}), \\ &\mu_F(\text{LOS}), \mu_{VS}(\Delta), \mu_S(\Delta), \mu_M(\Delta), \mu_L(\Delta), \mu_{VL}(\Delta)]. \end{aligned} \quad (11)$$

Every output unit in Fuzzy Neural Network is corresponding to certain fuzzy subset of the output variable, which is $\Delta(t)$. The format of output signal is

$$[\mu_{VS}(\Delta), \mu_S(\Delta), \mu_M(\Delta), \mu_L(\Delta), \mu_{VL}(\Delta)]. \quad (12)$$

To get membership degree model of LOS and Δ , SAGA-FCM is used to get the clustering centers. LOS is divided into six types [15], as $\text{LOS} = [A, B, C, D, E, F]$. The membership degree of LOS is calculated by

$$\begin{aligned} \mu_i(\text{LOS}(t)) &= \left(\frac{\|x(t) - a_i\|}{\|x(t) - a_A\|} + \frac{\|x(t) - a_i\|}{\|x(t) - a_B\|} \right. \\ &\quad \left. + \dots + \frac{\|x(t) - a_i\|}{\|x(t) - a_F\|} \right)^{-1}, \end{aligned} \quad (13)$$

$$\text{s.t. } i \in \{A, B, C, D, E, F\},$$

where $\mu_i(\text{LOS}(t))$ is the membership degree of which sample data $x(t)$ that belongs to LOS_i , $x(t) = [v(t), \rho(t)]$. a is the set of clustering centers for LOS , $a = \{a_A, a_B, a_C, a_D, a_E, a_F\}$, and $a_i = (v_i, \rho_i)$.

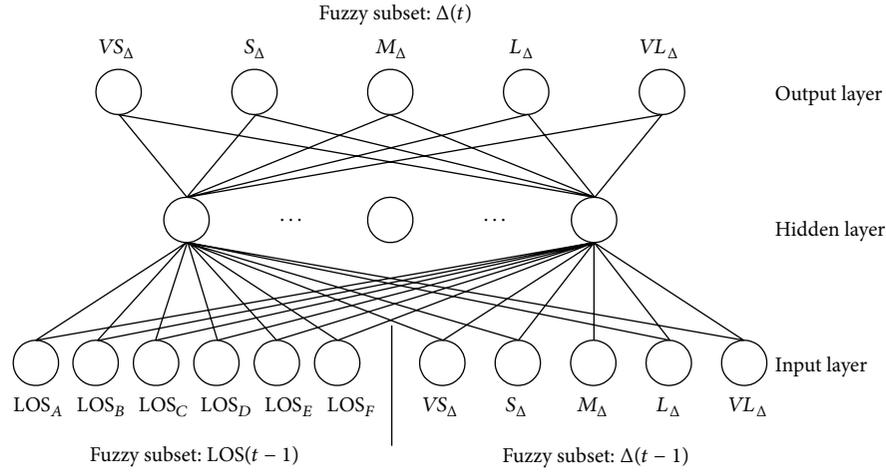


FIGURE 2: The structure of Fuzzy Neural Network.

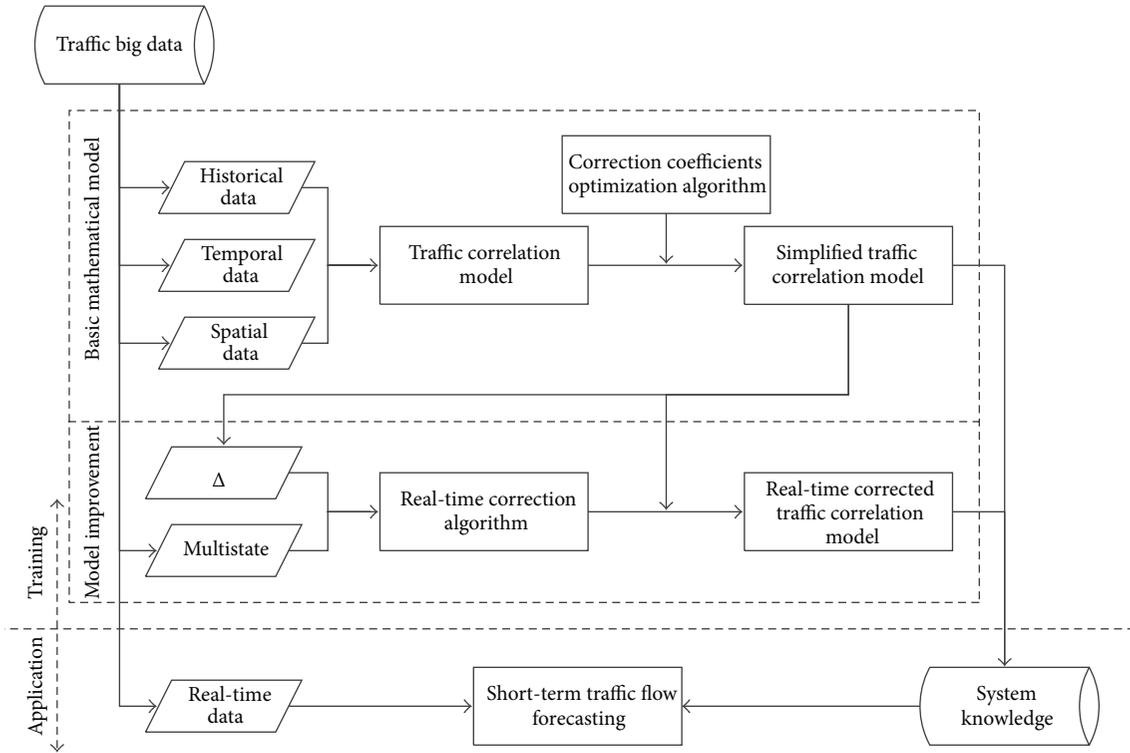


FIGURE 3: Traffic flow forecasting framework.

Δ is divided into five types, as $\Delta = [VS, S, M, L, VL]$. The membership degree model of Δ is calculated by

$$\mu_k(\Delta(t)) = \left(\frac{\|\Delta(t) - b_k\|}{\|\Delta(t) - b_{VS}\|} + \frac{\|\Delta(t) - b_k\|}{\|\Delta(t) - b_S\|} + \dots + \frac{\|\Delta(t) - b_k\|}{\|\Delta(t) - b_{VL}\|} \right)^{-1}, \quad (14)$$

s.t. $k \in \{VS, S, M, L, VL\}$,

where $\mu_k(\Delta(t))$ is the membership degree of which sample data $\Delta(t)$ belongs to Δ_k . b is the set of clustering centers for Δ , $b = \{b_{VS}, b_S, b_M, b_L, b_{VL}\}$.

Historical data is applied to train the Fuzzy Neural Network. Input signal is corresponding to target output. After training, the network can be seen as a container of fuzzy relations and if we want to get other conclusions from the network, the only thing that needs to be done is to input the real value after defuzzification, as shown in

$$\Delta(t) = \frac{\sum \mu_k(\Delta(t)) \cdot b_k}{\sum \mu_k(\Delta(t))}. \quad (15)$$

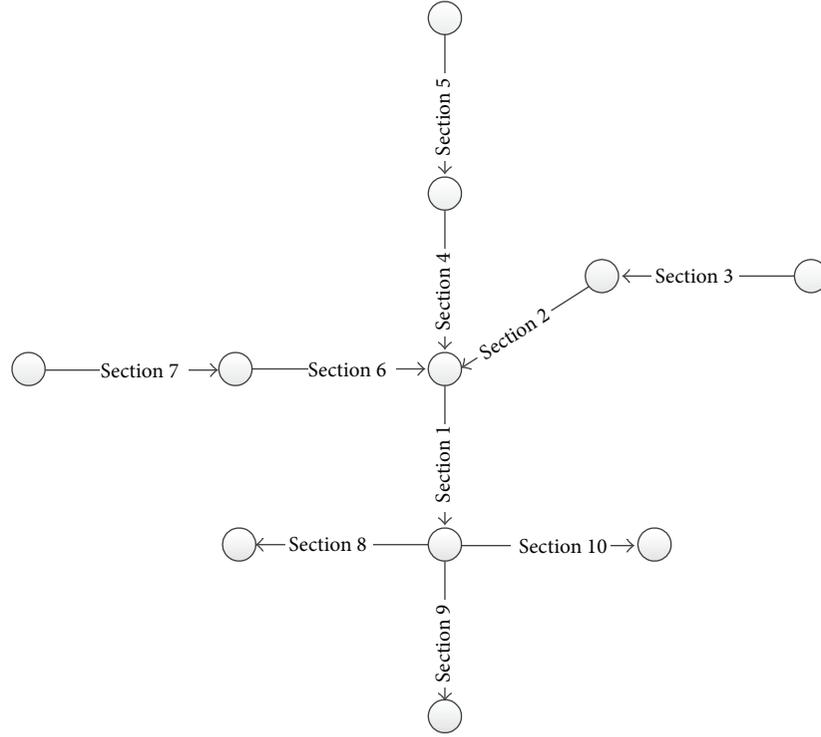


FIGURE 4: Spatial location of research object.

3.3. Real-Time Corrected Traffic Correlation Model. Real-time corrected traffic correlation model, which is seen as the improved traffic correlation model, is composed of static part and dynamic part. The static part is “ $\varphi \sum_{j=1}^n \omega_j V_j(t-d) + \varepsilon$,” which shows the physical meaning of traffic correlation. The dynamic part is “ $\Delta(t)$,” which shows the physical meaning of traffic multistate characteristic. Traffic flow forecasting framework is outlined in Figure 3.

The steps of traffic flow forecasting are as follows.

Step 1 (traffic correlation model). Based on historical data, temporal data, and spatial data, basic traffic correlation model, as shown in formulation (1), is built.

Step 2 (simplified traffic correlation model). Simplified traffic correlation model, as shown in formulation (6), is proposed based on correction coefficients optimization algorithm. Thus, “ $\varphi \sum_{j=1}^n \omega_j V_j(t-d) + \varepsilon$ ” is obtained.

Step 3 (real-time correction algorithm). Based on Fuzzy Neural Network, real-time correction algorithm is put forward. Thus, “ $\Delta(t)$ ” is obtained.

Step 4 (real-time corrected traffic correlation model). Real-time corrected traffic correlation model, as shown in formulation (8), is stored in the database as system knowledge.

Step 5 (short-term traffic flow forecasting). Based on system knowledge, real-time data is processed to calculate the forecasting results.

4. Case Study

4.1. Data Characteristics. Taking a section of the Second Ring Road (Section 1, as shown in Figure 4) and its surrounding trunk road in Beijing, China, as the object of study, it verifies the effectiveness and feasibility of the proposed method. Basic traffic flow data are detected by microwave detectors. Interval time of traffic flow data is 5 min.

4.2. Effective Analysis

4.2.1. Basic Mathematical Model. Based on historical data, basic mathematical model, including traffic correlation model and simplified traffic correlation model, is built. The static part of real-time corrected traffic correlation model, which is “ $\varphi \sum_{j=1}^n \omega_j V_j(t-d) + \varepsilon$,” is achieved, as shown in

$$q_1(t) = 1.0629 \times [0.2246 \cdot q_9(t) + 0.2151 \cdot q_1(t-5) + 0.2145 \cdot q_1^H(t-5) + 0.1787 \cdot q_2(t) + 0.1670 \cdot q_6(t-10)] + 22.652,$$

$$v_1(t) = 1.029 \times [0.2677 \cdot v_9(t) + 0.2630 \cdot v_1(t-5) + 0.2461 \cdot v_1^H(t) + 0.2232 \cdot v_2(t)]$$

$$+ 1.8173,$$

$$\rho_1(t) = 1.0148 \times [0.2791 \cdot \rho_9(t) + 0.2747 \cdot \rho_1(t-5) + 0.2294 \cdot \rho_2(t) + 0.2168 \cdot \rho_1^H(t)]$$

$$+ 0.4793.$$

(16)

TABLE 1: Goodness of Fit.

Parameter of traffic flow	Flow	Speed	Occupancy
R^2	0.8885	0.9655	0.9664

TABLE 2: Clustering centers of LOS.

i	Speed (km/h)	Occupancy (%)
A	68.1	2.1
B	58.3	3.1
C	51.2	11.4
D	46.0	14.9
E	38.0	20.5
F	23.4	35.1

TABLE 3: Clustering centers of Δ .

k	Flow (pcu/h)	Speed (km/h)	Occupancy (%)
VS	-1043	-9.7998	-7.4513
S	-498	-2.5634	-2.2416
M	-2.9	0.2030	-0.2448
L	291.6	2.3692	1.6348
VL	812.2	9.9043	7.0243

Goodness of Fit is shown in Table 1, in which the value of R^2 shows the high effectiveness of formulation (16).

4.2.2. Model Improvement. Real-time correction algorithm is presented to obtain the dynamic part of real-time corrected traffic correlation model, which is " $\Delta(t)$." The steps are as follows.

Step 1 (clustering center search). Based on SAGA-FCM, LOS is divided into six types; Δ is divided into five types. Clustering Centers are shown in Tables 2 and 3.

Step 2 (calculation of membership degree). Formulations are shown in Section 3.2.

Step 3 (training Fuzzy Neural Network). Input layer has 11 nodes; the transmission function of the nerve cells in the hidden layer is transig; the output layer has 5 nodes and the transmission function of the nerve cells in the output layer is logsig, while the training function is traingdx. Historical data is used to train Fuzzy Neural Network.

4.2.3. Short-Term Traffic Flow Forecasting. Making one day as an example, the result is shown in Figure 5.

Mean Absolute Percentage Error (MAPE) is used to evaluate the accuracy of the proposed model:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\widehat{V}_s(t) - V_s(t)}{\widehat{V}_s(t)} \right| \times 100\%. \quad (17)$$

Evaluation result is shown in Table 4. Comparing with these two kinds of model, real-time corrected traffic correlation model is better than the basic mathematical model. That is, the static part of real-time corrected traffic correlation

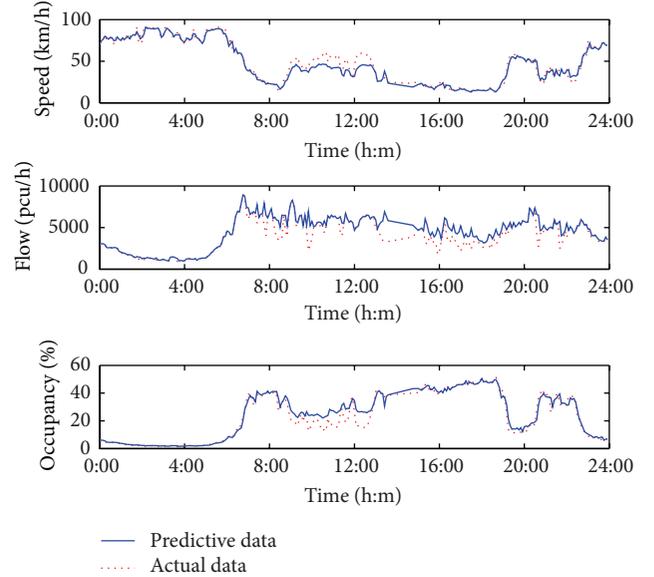


FIGURE 5: Effective analysis of real-time correction algorithm.

model reflects the temporal-spatial-historical correlation by way of mathematical model; the dynamic part of real-time corrected traffic correlation model modifies the mathematical model based on multistate characteristics of traffic big data.

5. Conclusions

Traffic big data strongly shows temporal-spatial-historical correlation and multistate characteristic. Traffic correlation model is established based on temporal-spatial-historical correlation. Correction coefficients optimization algorithm is put forward to reduce parameters and ensure calculation accuracy. In order to improve the effectiveness of short-term traffic flow forecasting, real-time correction algorithm is presented based on multistate characteristic. Fuzzy Neural Network is used to overcome the problem of nonlinear mapping. Case study shows that real-time correction algorithm can improve the effectiveness of traffic correlation model.

The core of this paper is to present a short-term traffic flow forecasting based on traffic big data analysis. The advantages of real-time corrected traffic correlation model for traffic flow forecasting are as follows.

- (1) The temporal-spatial-historical correlation, which is considered in the static part of model, explains the physical meaning of traffic flow forecasting by mathematical model.
- (2) The multistate, which is considered in the dynamic part of model, explains the dynamic characteristic of traffic flow.
- (3) Real-time correction algorithm improves the accuracy of traffic flow forecasting. Case study shows the high efficiency and applicability of the proposed methods.

TABLE 4: Evaluation result of proposed models.

Model	MAPE of flow (%)	MAPE of speed (%)	MAPE of occupancy (%)
Basic mathematical model	13.14%	9.18%	14.48%
Model improvement	10.93%	7.17%	10.54%

Moreover, the proposed methods can be extended to 15- or 30-minute-ahead forecasting. The next steps of this work are to study traffic incident and its influence for short-term traffic flow forecasting. In addition, how to deal with the long period traffic flow forecasting like one hour or even longer can also be focused on.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to the editor and anonymous reviewers for their valuable suggestions. This research was funded by the “Twelfth Five-Year” National Science & Technology Pillar Program (2014BAG01-B04) and Beijing Science and Technology Plan (no. Z121100000312101).

References

- [1] Y. Peng, M. Lei, J. B. Li, and X. Y. Peng, “A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting,” *Neural Computing and Applications*, vol. 24, no. 3-4, pp. 883–890, 2014.
- [2] Y.-C. Zhao, J.-F. Wu, Y. Zhang, and B. Zhou, “Wavelet-based kalman filter for traffic flow forecasting in sensor networks,” *Information Technology Journal*, vol. 11, no. 10, pp. 1518–1522, 2012.
- [3] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, “Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg-marquardt algorithm,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, 2012.
- [4] X. L. Zhang and H. P. Lu, “Non-parametric regression and application for short-term traffic flow forecasting,” *Journal of Tsinghua University*, vol. 49, no. 9, pp. 1471–1475, 2009.
- [5] G. Fu, G.-Q. Han, F. Lu, and Z.-X. Xu, “Short-term traffic flow forecasting model based on support vector machine regression,” *Journal of South China University of Technology (Natural Science)*, vol. 41, no. 9, pp. 71–76, 2013.
- [6] J. Wang and Q. X. Shi, “Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet analysis-support vector machine theory,” *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 219–232, 2013.
- [7] Y. P. Zhang and Y. L. Pei, “Research on traffic flow forecasting model based on cusp catastrophe theory,” *Journal of Harbin Institute of Technology*, vol. 11, no. 1, pp. 1–5, 2004.
- [8] X. W. Liu, X. M. Fang, Z. H. Qin, C. Ye, and M. Xie, “A Short-term forecasting algorithm for network traffic based on chaos theory and SVM,” *Journal of Network and Systems Management*, vol. 19, no. 4, pp. 427–447, 2011.
- [9] S. Ricci and A. Tieri, “Check and forecasting of railway traffic regularity by a PETRI Nets based simulation model,” *Ingegneria Ferroviaria*, vol. 64, no. 9, pp. 723–767, 2009.
- [10] Y. Zhang and Y. C. Liu, “Analysis of peak and non-peak traffic forecasts using combined models,” *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 21–37, 2011.
- [11] H.-P. Lu, Z.-Y. Sun, and W.-C. Qu, “Big data-driven based real-time traffic flow state identification and prediction,” *Discrete Dynamics in Nature and Society*, In press.
- [12] H. H. Dong, X. L. Sun, L. M. Jia, and Y. Qin, “Multimode traffic volume prediction model,” *Journal of Jilin University*, vol. 41, no. 3, pp. 645–649, 2011.
- [13] H. P. Lu, Y. G. Sui, M. Guo, and R. M. Li, *Urban Mixed Traffic Flow Analysis Model and Method*, China Railway Publishing House, Beijing, China, 2009.
- [14] L. G. Sun, R. M. Li, S. Dong, and H. P. Lu, “Study on short-term traffic flow combined forecast method,” *Journal Wuhan University of Technology (Transportation Science & Engineering)*, vol. 34, no. 5, pp. 874–881, 2010.
- [15] Transportation Research Board, *2010 Highway Capacity Manual*, Transportation Research Board of the National Academies, Washington, DC, USA, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

