

Research Article

Highway Travel Time Prediction Using Sparse Tensor Completion Tactics and K -Nearest Neighbor Pattern Matching Method

Jiandong Zhao ^{1,2}, Yuan Gao,³ Jinjin Tang ¹, Lingxi Zhu,¹ and Jiaqi Ma⁴

¹School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

²MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing, China

³School of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, Beijing, China

⁴Department of Civil and Architectural Engineering and Construction Management, University of Cincinnati, Cincinnati, OH, USA

Correspondence should be addressed to Jiandong Zhao; zhaojd@bjtu.edu.cn

Received 28 September 2017; Revised 8 January 2018; Accepted 31 January 2018; Published 14 March 2018

Academic Editor: Steven I. Chien

Copyright © 2018 Jiandong Zhao 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.

Remote transportation microwave sensor (RTMS) technology is being promoted for China's highways. The distance is about 2 to 5 km between RTMSs, which leads to missing data and data sparseness problems. These two problems seriously restrict the accuracy of travel time prediction. Aiming at the data-missing problem, based on traffic multimode characteristics, a tensor completion method is proposed to recover the lost RTMS speed and volume data. Aiming at the data sparseness problem, virtual sensor nodes are set up between real RTMS nodes, and the two-dimensional linear interpolation and piecewise method are applied to estimate the average travel time between two nodes. Next, compared with the traditional K -nearest neighbor method, an optimal KNN method is proposed for travel time prediction. optimization is made in three aspects. Firstly, the three original state vectors, that is, speed, volume, and time of the day, are subdivided into seven periods. Secondly, the traffic congestion level is added as a new state vector. Thirdly, the cross-validation method is used to calibrate the K value to improve the adaptability of the KNN algorithm. Based on the data collected from Jinggangao highway, all the algorithms are validated. The results show that the proposed method can improve data quality and prediction precision of travel time.

1. Introduction

Travel time is a traffic parameter that can reflect traffic conditions effectively, and it is the most popular traffic information for travelers. How to predict travel time accurately and timely is a classic research question in intelligent transportation system (ITS) research.

In many countries, the travel time estimation and prediction studies are based on the traffic data collected by loop vehicle detectors. In China, although the highway mileage ranks first in the world currently, because of the limitation of early construction budget, no loop vehicle detectors or other traffic detection equipment was set on highways. In recent years, with the development of the remote transportation microwave sensor (RTMS) technology, some highways in China began to put an RTMS every 2 to 5 km to monitor the

traffic flow conditions. RTMS is a fixed traffic information collection device installed by a side mount (the working process of RTMS is described in Section 2). This makes it possible to carry out highway travel time prediction in China.

Abnormal operation, noise interference, or transmission line failure can lead to incomplete traffic data collected by RTMSs, and the large distance between two sensors on a highway can lead to a sparse data problem. Data missing due to equipment failure and other hardware-related problems can be regarded as data missing at random. After introducing the concept of virtual sensor nodes, the sparse data problem caused by the large distance between microwave sensors can be regarded as data missing at the virtual node. These phenomena inevitably affect the quality of the traffic data collected by the remote transportation microwave sensor, thus affecting the accuracy of travel time prediction. In order

to obtain accurate and reliable travel time prediction results, the first step is to solve the above two kinds of traffic data missing.

Aiming at repairing the traffic data missing at random, Coifman [1] proposed a creative interpolation algorithm based on kinetic-wave models to predict travel time. Li et al. [2] pointed out that Coifman's method is deviated, and proposed a new algorithm based on the temporal-spatial queueing model to describe the fast travel time variations using only the speed and headway time series that is measured at upstream and downstream detectors. Zhang et al. [3] pointed out that the mobility and scale limit of probe vehicles may lead to incomplete or inaccurate data and proposed a method to fill the missing data through the weighted average method and exponential smoothing method combined with the characteristics of road network. Zhang and Zhang [4] established a prediction model based on a vector autoregressive model and a general regression neural network with the spatial relationship of adjacent location traffic volume, and concluded that the latter is superior to the former when the missing data ratio is higher. Zhao et al. [5] adopted a two-dimensional linear time-space interpolation method to interpolate microwave vehicle data and predicted travel time based on the K -nearest neighbor pattern matching method. Qu et al. [6] used the matrix completion method to estimate the missing data. Tang et al. [7] estimated missing detector data based on fuzzy C -means and a genetic algorithm. Using weekly data similarity, the data vector structure is transformed into a matrix pattern for completion. Duan et al. [8] designed a denoising stacked autoencoder deep learning model and trained the model hierarchically using the full set of data from all vehicle detector stations to achieve traffic data imputation. Li et al. [9] compared the data completion effect both of the probability and of the kernel probability principal component analyses through the fusion of multidetection point data, showing that imputation errors can be reduced using temporal-spatial dependence. Li et al. [10] pointed out that the PCA trend brings about several advantages to traffic time series analysis and used PCA for missing data completion. Then, they used the multivariable linear regression method for traffic volume prediction. Tan et al. [11] introduced the tensor model and tensor decomposition to the issue of recovering missing traffic data. This model can solve the extreme situation where the data is completely lost. They used the dynamic tensor model to express the traffic data and predict short-term traffic volumes [12]. Considering the traffic volume time-space information, Ran et al. [13] reconstructed traffic volume data in the four-dimensional tensor mode and interpolated the missing data with a low-order tensor completion algorithm. They also converted traffic state estimation to a missing data estimation problem [14]. Zhou et al. [15] recovered the missing OD data by the tensor completion method, taking into account the low-dimensional potential structure and time-space characteristics of volume.

Through the above literature, it can be seen that the existing recovery theories and methods for missing traffic data at random can be attributed to the form of vector, matrix, and tensor, according to the combination of traffic data. The

missing data recovery method based on the structure of "vector or matrix" can be mainly divided into statistical type, interpolation type, and predictive type. This method does not excavate the multimode traffic data characteristics such as time, space, week, and day, and the completion precision is limited. The tensor model is a new method to solve the missing data problem that can completely use the multimode characteristic of data and complete missing data in high-dimensional space [11].

Aiming at solving the sparse data problem, Quiroga and Bullock [16, 17] proposed a travel time estimation method based on the velocity integral model and the position interpolation model using expressway vehicle GPS. Wang et al. [18] proposed a model based on the naïve Bayesian method to achieve traffic volume estimation on the road network where GPS samples are not covered. Xu [19] used a two-dimensional linear interpolation method and a segmentation method to estimate the travel time between expressway stations by remote transportation microwave sensor data. The author of this paper has proposed a solution to the sparse RTMS data problem: Zhao et al. [20] used a two-dimensional linear interpolation method to solve the problem of sparse microwave detector data. This method has had a positive effect on improving the effectiveness of traffic congestion detection.

As a research hotspot of modern intelligent transportation systems, there are many travel time prediction methods. These methods can be divided into model-based methods (ARIMA method [21], Bayesian method [22], multivariable linear regression [10], neural network prediction algorithm [23], Kalman filter algorithm [24], support vector machine [25], etc.), and data-driven methods (k -nearest neighbor matching [5], nonparametric regression [26], etc.). In this paper, the travel time is predicted based on the RTMS data. The sampling interval is short and the sampling frequency is high, so a large amount of data can be collected in a short time. The model-based method has too many parameters when dealing with massive data, and model structure is complicated. The data-driven approach does not need establishing a model or the identification of a large number of parameters. Therefore, it is suitable for handling massive data. The K -nearest neighbor algorithm is a data-driven method, which can make use of multimodal information to cluster data from similar modes together for prediction. The basic idea behind the K -nearest neighbor algorithm is to obtain a complete historical database, extract the data characteristics, and then match the historical data pattern that is most similar to the current data pattern for future situation prediction. The author of this paper has studied travel time prediction based on the K -nearest neighbor algorithm [5]. However, in practical application, it is found that the traditional K -nearest neighbor algorithm has some limitations, such as poor ability to adapt caused by the fixed value of K and low prediction accuracy when there is a large amount of abnormal data or under congested traffic conditions. Therefore, this paper uses the K -nearest neighbor algorithm to predict travel time and further optimizes the K -nearest neighbor algorithm.

In summary, the major purpose of this paper is similar to the literature [10]: to solve the issues caused by missing data

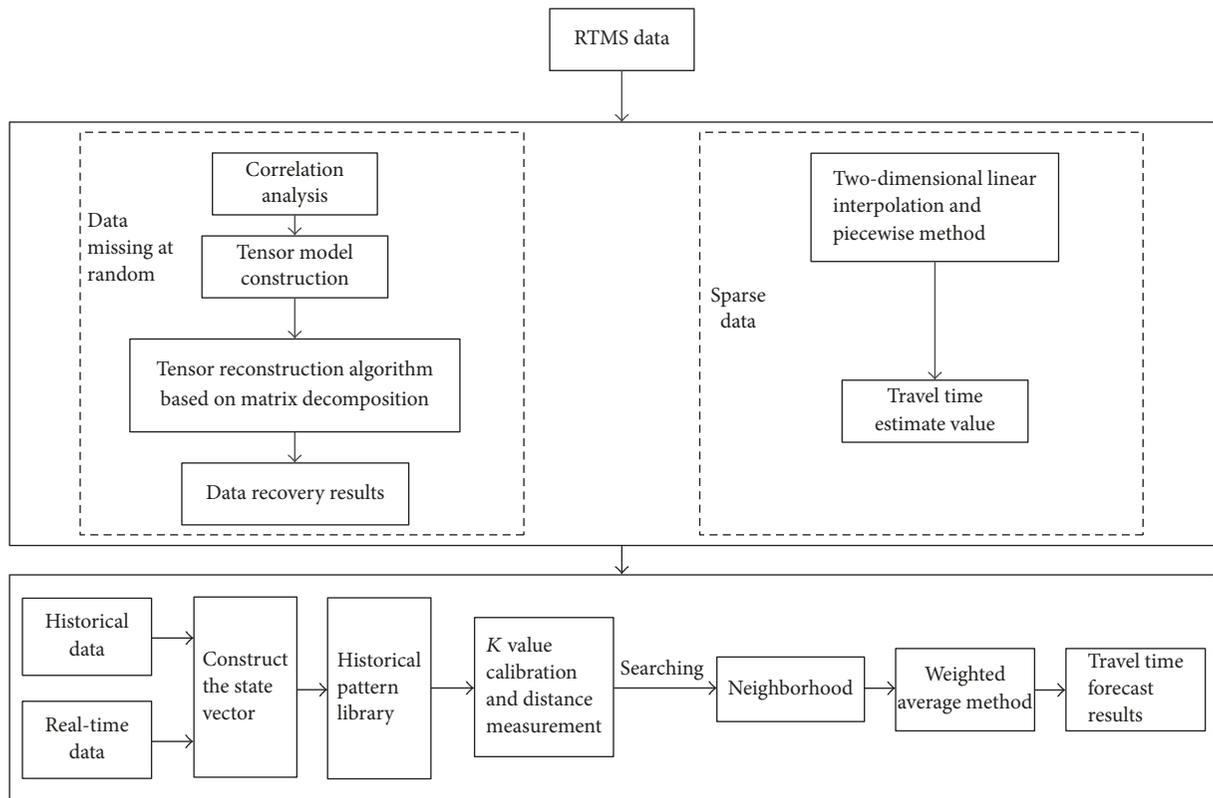


FIGURE 1: Framework of the paper.

and sparsity problem and then improve the accuracy of travel time estimation and prediction. Therefore, this paper firstly uses the data completion method based on the tensor form to complete the missing RTMS data. Then, the concept of virtual nodes is introduced, and the two-dimensional linear interpolation method is used to fill the completely missing data at the virtual node to estimate the travel time. The traditional travel time prediction method based on the K -nearest neighbor algorithm is optimized. The state vector (speed, traffic volume, and time of day) is subdivided. The degree of traffic congestion is increased with the new state vector. The K value is calibrated by the cross-validation method. Finally, the method was verified using the remote transportation microwave sensor data of Jinggangao highway. The framework of this paper is shown in Figure 1.

2. Data Preprocessing

Because RTMS has the characteristics of sublane detection, the collected data cannot be directly used for the travel time prediction. It should be sorted into section data first. Due to equipment failure, transmission failure, and other reasons, there are some abnormalities in RTMS data, which seriously affected the quality of the data. This section describes the data preprocessing process. Firstly, the sublane data is integrated into section data. Secondly, an abnormal data judgment method is based on the threshold method and traffic flow theory.

For the preprocessing of RTMS data, the author of this paper has proposed a method in the literature [19]. For the convenience of the reader's understanding, this part introduces the data preprocessing method in the literature [19].

The data in this paper was collected from the Jinggangao highway, Dujiakan station to Jingliang Road station, between January 1, 2016, and March 15, 2016. The experiment segment has three toll stations. The number of sensors is 11. The experiment segment has three lanes in each direction. The highest speed limit of the freeway is 120 km/h and the length of this highway segment (Dujiakan station to Jingliang Road station) is 9.6 km. Jinggangao Expressway is an expressway linking Beijing with major southern cities such as Guangzhou, Zhuhai, Hong Kong, and Macao. The section studied in this paper is located in the capital of China, Beijing. It is one of the main lines to get in and out of Beijing in the southwest. The traffic on the experiment segment is heavy every day and it experiences traffic congestions during peak hours.

As shown in Figure 2, the RTMS [20] is a fixed-type traffic information collection device based on digital radar technology, which can obtain real-time traffic information data. Take traffic speed detection as an example; when a vehicle enters the monitoring area in Figure 2, the RTMS will detect and record its entry time. When the vehicle leaves the detection area, the remote transportation microwave sensor will also record its departure time. The vehicle speed can be

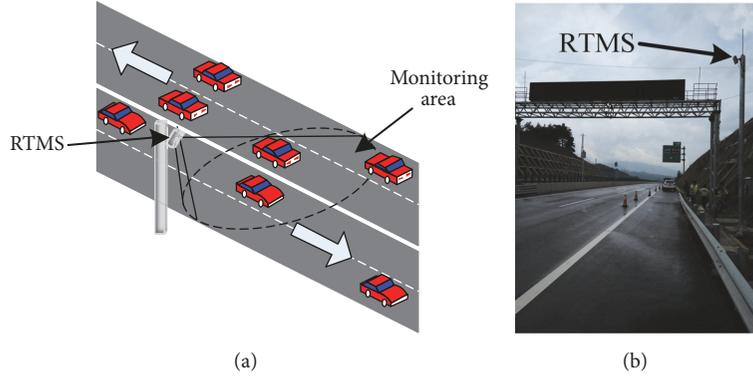


FIGURE 2: Remote transportation microwave sensor. (a) Detection principle. (b) Field application.

TABLE 1: Traffic information collected by microwave detectors.

Number	Definition	Description
(1)	<i>occur_time</i>	<i>Detection time</i>
(2)	<i>device_id</i>	<i>ID number of device</i>
(3)	<i>intensity</i>	<i>Traffic volume</i>
(4)	<i>speed</i>	<i>Speed of vehicle</i>
(5)	<i>occupy</i>	<i>Time occupancy</i>
(6)	<i>car_class</i>	<i>Type of vehicle</i>
(7)	<i>direction_id</i>	<i>Direction of lane</i>
(8)	<i>lane_id</i>	<i>Number of lane</i>
(9)	<i>road_id</i>	<i>Number of road</i>
(10)	<i>transferred</i>	<i>Flag of transfer</i>

calculated using the time difference and detection distance. Therefore, the detected traffic speed is the average speed of all the vehicles that pass through the detection zone during a certain period for a certain lane. RTMS can detect other traffic data including intensity, occupancy, and car style. Table 1 shows the traffic information that can be collected by RTMS. This paper uses 7 italic items in Table 1.

2.1. Data Integration. RTMS has the characteristic of sublane detection. RTMS detects traffic data of six lanes. However, the basis of traffic flow prediction is instantaneous section speed and intensity. Taking into account the contribution rate of traffic volume on the section speed calculation, a weighted method that can convert all the data of each lane to section data is presented [5]. The data of six lanes was integrated into available section average speed and total intensity data, as shown in (1), and the data integration period is 5 minutes.

$$\begin{aligned}
 v &= \frac{\sum V_s \times Q_i}{q}, \\
 q &= \sum Q_i,
 \end{aligned} \tag{1}$$

where v represents the average speed of all vehicles passing through the detection section; V_s represents the average speed of the vehicle passing through the detection section in each

lane during the detection period; Q_i represents the traffic volume of each lane passing through the detection section during the detection period; q represents the total traffic volume through the detection section during the detection period.

2.2. Abnormal Data Identification. The abnormal data judgment method is based on the threshold method and traffic flow theory. The threshold method takes into account the fact that the value of the traffic volume parameter must change within a reasonable threshold interval at a certain time interval. Traffic volume theory holds that the relationship between the three parameters of traffic volume satisfies a certain regularity. Based on the theory of traffic volume, this paper obtains the discriminant rules of threshold form and realizes the identification and judgment of abnormal data. According to the characteristics of the test section, the rules were determined as follows.

Step 1 (recognition based on the threshold method). The rules determined according to the threshold method are shown in (2). The collected traffic volume data that does not meet (2) can be recognized as abnormal data.

$$\begin{aligned}
 0 &\leq Q \leq f_c \times C \times \frac{T}{60}, \\
 0 &\leq v \leq f_v \times V_f, \\
 0 &\leq O \leq 100\%,
 \end{aligned} \tag{2}$$

where Q is the vehicle volume; C is the road capacity; T is the data acquisition cycle; f_c is the correction factor (the value of f_c varies between 1.3 and 1.5); v is the location average speed; V_f is the limit speed of the road; O is the occupancy; f_v is the correction factor (the value of f_v varies between 1.3 and 1.5).

Step 2 (recognition based on traffic volume theory). According to the traffic volume theory and the vehicle driving conditions on the road, the abnormal data determination rule is as follows:

(1) The location average speed V is 0; the volume Q is not 0.

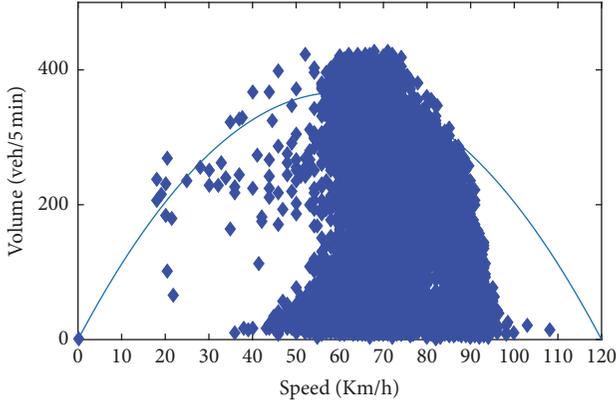


FIGURE 3: Experiment data distribution.

- (2) The volume Q is 0, and the occupancy rate O and the location average speed v are not 0 at the same time.
- (3) The occupancy rate is 0, and the volume rate is greater than the set value.
- (4) The relationship between speed and volume is satisfied in

$$Q > \frac{4Q_m}{V_f^2} \times v \times (V_f - V) \times f_Q, \quad (3)$$

where Q_m is the maximum volume and f_Q is a correction factor.

2.3. Validation and Analysis. According to the “Technical Standard of Highway Engineering (JTG B01-2014),” it is known that the road capacity of expressways in China’s plains area is 2,000 vehicles per hour and the speed limit is 120 km/h. Therefore, according to the traffic engineering theory, the relationship between speed and volume should meet

$$\begin{aligned} Q &\leq \frac{4Q_m}{V_f^2} \times v \times (V_f - v) \times f_Q \\ &\leq \frac{5v(120 - v)}{9} \times f_Q \cdot \text{veh/h} \\ &\leq \frac{11v(120 - v)}{108} \cdot \text{veh/5 min.} \end{aligned} \quad (4)$$

Figure 3 is a speed–volume scatter plot drawn using the data from one sensor on a test road section. On the y -axis, the unit is veh/5 min in one direction. The points outside the curve are recognized as abnormal data.

3. Random Missing Traffic Data Recovery with Tensor Reconstruction

Abnormal operation, noise interference, transmission line failure, or abnormal data removal can lead to a random data-missing problem. The data-missing problem inevitably affects the quality of the traffic data collected by the RTMS. In this section, a missing data recovery method based on the static

TABLE 2: The relevance of each pattern.

Pattern	Correlation
Space	0.9763
Week	0.9910
Day	0.9011
Time	0.2656

tensor theory is described. The reconstruction procedure of traffic data based on the static tensor model includes a correlation analysis of traffic data, establishment of the traffic tensor model, a low-rank analysis of traffic data, and the recovery method based on a multilinear low full-rank decomposition model.

3.1. Correlation Analysis of Traffic Data. Similarity is one of the factors impacting imputation performance. Combining the multidimensional characteristics of traffic data, mining the multimode similarities will make a great contribution to imputing missing values [11]. In this paper, the correlation analysis is used to reveal the similarity of traffic data.

The multimode characteristic is the result of people’s understanding and the description of characteristics from different angles. The analysis result is different when people analyze one thing from different angles or using different methods. For the traffic data, it showed significant multimode characteristics when observed from the day, week, space, or time mode, and the researchers found that traffic data showed a strong multicorrelation in these modes. Figure 4 shows the distribution of remote transportation microwave sensor data in each mode. Figure 4(a) shows the distribution of data in day mode and time mode; Figure 4(b) is a three-dimensional diagram which shows the distribution of data in space mode, where the x -axis represents the moment, the y -axis represents the location, and the z -axis represents the volume. Figure 4(c) is a three-dimensional diagram which shows the distribution of data in week mode. The x -axis represents the moment, the y -axis represents the week, and the z -axis represents the volume. It can be seen that RTMS data has similar characteristics in different modes. It can be seen from Figure 4(a) that the volumes between 8 and 9 AM on Thursday were lower than other periods. As 8-9 AM is the peak of the weekday, traffic congestion during this time of Thursday caused a drop in traffic volume.

In this paper, the following equation in the literature [10] is used to give the correlation of the various modes of traffic data:

$$S = \frac{\sum_{n \geq i \geq j \geq 1} R(i, j)}{n(n-1)/2}, \quad (5)$$

where n refers to the whole data points and $R(i, j)$ refers to the correlation coefficient matrix.

Table 2 shows the correlation of traffic data tensors in day mode, week mode, space mode, and time mode.

It can be seen from Table 2 that the traffic data show a high correlation in space, week, and day modes. However,

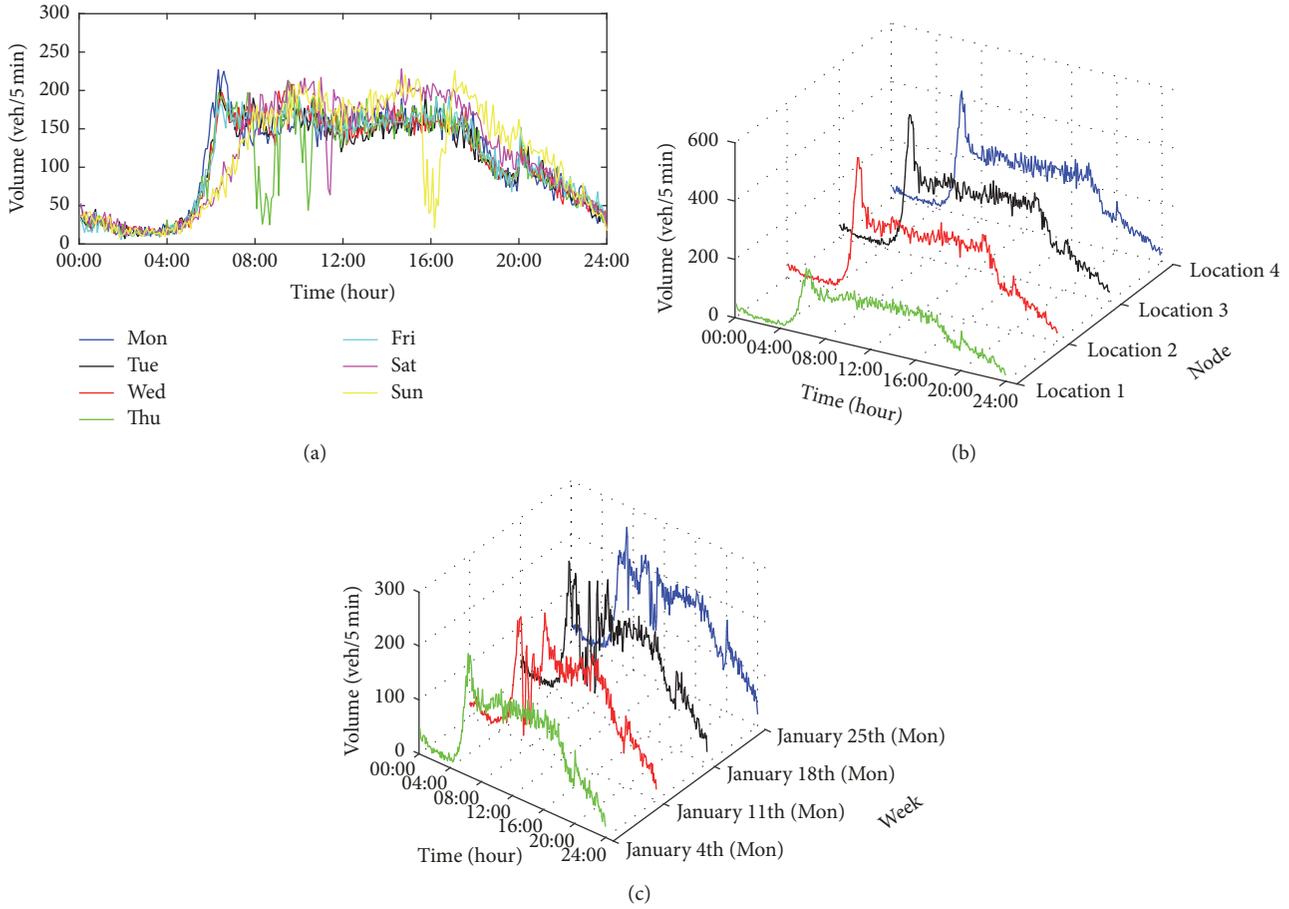


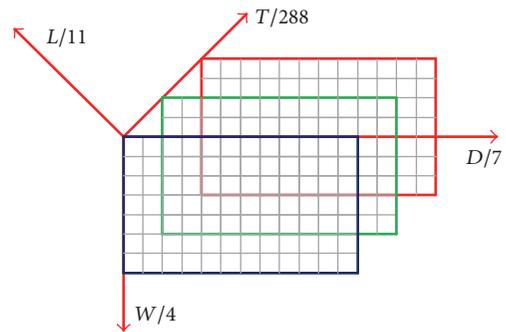
FIGURE 4: The trend of the various modes of traffic data. (a) Day mode and time mode. (b) Space mode. (c) Week mode.

in the time mode, the traffic volume shows different states at different times, which makes the traffic volume tensor exhibit a low correlation in time mode. It can also be seen that, due to the weak correlation between weekdays and weekends, compared to the “week” mode, the correlation in the “day” mode is low.

3.2. *Establishment of Traffic Data Tensor Model.* The tensor is a multidimensional array, which is a high-order extension of the first-order vector and the second-order matrix. The position of the element needs to be represented by three or more variables. The N -order tensor is expressed as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$.

Section 3.1 illustrated that the expressway traffic data show strong temporal and spatial correlation, which provides a sufficient base for the construction of a traffic data tensor model. Based on the analysis results of Section 3.1, this paper constructs the tensor form of traffic data $\mathcal{A} \in \mathbb{R}^{4 \times 11 \times 7 \times 288}$, as shown in Figure 5, by combining the multimode traffic volume spatial-temporal characteristics.

3.3. *Low-Rank Analysis of Traffic Data.* The reason why the tensor completion algorithm can successfully estimate the missing data in the traffic tensor is that the low-rank



Model structure:
 L/11 indicates that the model contains 11 sensors
 W/4 represents a total of 4 weeks of data
 D/7 means 7 days a week
 T/288 represents a day of statistical traffic 288 times
 Existing similarity among sensors, days, and hours

FIGURE 5: Tensor model of traffic data.

hypothesis can successfully approximate the multimode correlation of the traffic data. In essence, when the tensor is of low rank or has sparse characteristics, it can be said that

the tensor is redundant and can be expressed in a more compact form, so that a complete tensor can be reconstructed with fewer sampling tensors. To solve the problem of lost data recovery of the tensor, it is generally assumed that the tensor to be recovered has a low-rank structure; that is, its data is distributed in a low-dimensional linear subspace. Obviously, the traffic data tensor model constructed by the actual data without any human intervention is not of low rank, and most of the time it is full rank. Therefore, a low-rank approximation is needed. In order to reveal the working principle of the low rank, the high-order singular value decomposition (HOSVD) on the tensor is used for a multimode low-rank approximation. HOSVD can use a tensor with a low-rank structure to approximate a known tensor. The HOSVD model of the tensor decomposes the tensor into a kernel tensor, multiplying a matrix of n -mode in each mode as shown in

$$\mathcal{A} \approx G \times_1 U_1 \times_2 U_2 \times \cdots \times_N U_N. \quad (6)$$

The tensor G is called a kernel tensor, and its elements represent the relationship between the principal components of various models. U_N is the expansion matrix of N -mode. $G \times_N U$ is the N -mode product of tensor G and matrix U .

Figure 6 shows the partial results of the low-rank approximation of the traffic tensor using the n -mode rank $[3, 3, 3, 3]$. As shown in Figure 6, the low-rank approximation of the tensor with rank $[3, 3, 3, 3]$ can maintain the main change of $4 \times 11 \times 7 \times 288$ traffic volume tensor data. The appropriate multi-low-rank hypothesis can effectively capture the multirelevance of traffic volume, gain the main changes of traffic volume data, and thus estimate the missing data.

In this paper, the tensor recovery algorithm based on low-rank and multilinear full-rank decomposition has the same purpose as the HOSVD model. They all make a low-rank approximation for a given tensor.

3.4. Tensor Completion Based on Matrix Decomposition. In this paper, the missing tensor data recovery via multilinear low/full-rank decomposition model is used to estimate the missing data in a tensor. For the convenience of the reader's

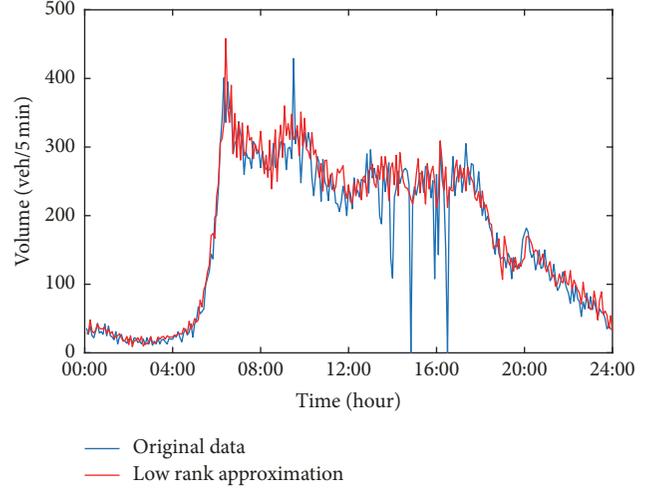


FIGURE 6: Low-rank approximation results of traffic data.

understanding, this part first introduces the method taking the fourth-order tensor in this part of the paper as an example.

As shown in (7), the tensor completion method uses the optimization algorithm to recover the missing data by minimizing the tensor rank as low as possible. That is, the recovered tensor has a low-rank structure.

$$\begin{aligned} \min_{\mathcal{A} \in R^{I_1 \times I_2 \times I_3 \times I_4}} \quad & \text{rank}(\mathcal{A}), \\ \text{s.t.} \quad & \mathcal{A}_\Omega = \mathcal{B}_\Omega, \end{aligned} \quad (7)$$

where $\text{rank}(\mathcal{A})$ represents the rank of the tensor and Ω is the set of known element subscripts. \mathcal{B} is the original volume (or speed) tensor and \mathcal{A} is the recovered tensor. The constraint is that the value with the subscript Ω is equal to the value of the corresponding data in \mathcal{B} .

However, the formula is not solvable. Assuming that the given tensor has an n -mode low-rank structure, (7) can be transformed to minimize the n -ranks of the given tensor, as shown in

$$\begin{aligned} F \left(\min_{\mathcal{A} \in R^{I_1 \times I_2 \times I_3 \times I_4}} (\text{rank}(A_{(1)}), \text{rank}(A_{(2)}), \text{rank}(A_{(3)}), \text{rank}(A_{(4)})) \right) \\ \text{s.t.} \quad (A_{(i)})_\Omega = (B_{(i)})_\Omega, \quad i = 1, 2, 3, 4, \end{aligned} \quad (8)$$

where $A_{(i)}$ and $B_{(i)}$ are the mode- i unfolding of \mathcal{A} and \mathcal{B} .

In order to make full use of the information of the tensor patterns, the rank of each mode of the tensor is simplified by the weighted summation strategy, as shown in

$$\sum_{i=1}^4 \lambda_i \left(\min_{A_{(i)}} \text{rank}(A_{(i)}) \right),$$

$$\text{s.t.} \quad (A_{(i)})_\Omega = (B_{(i)})_\Omega$$

$$\lambda_i = \frac{I_i}{I_1 + I_2 + I_3 + I_4}, \quad i = 1, 2, 3, 4. \quad (9)$$

In order to simplify the tensor completion problem, it is simplified by minimizing an F norm using the full-rank

decomposition of the matrix. The tensor completion problem can be transformed into

$$F(G_{(i)}, H_{(i)}, A_{(i)}) = \min_{G_{(i)}, H_{(i)}, A_{(i)}} \frac{1}{2} \|G_{(i)}H_{(i)} - A_{(i)}\|_F^2 \quad (10)$$

s.t. $(A_{(i)})_\Omega = (B_{(i)})_\Omega$

for $i = 1, 2, 3, 4$.

$G_{(i)}$, $H_{(i)}$, and $A_{(i)}$ of the function can be obtained by solving the function $F(G_{(i)}, H_{(i)}, A_{(i)})$, as shown in

$$\begin{aligned} G_{(i)} &\leftarrow \text{orth}(A_{(i)}H_{(i)}^T) \leftarrow A_{(i)}H_{(i)}^T \\ &\leftarrow A_{(i)}H_{(i)}^{-1} \leftarrow A_{(i)}H_{(i)}^T (H_{(i)}H_{(i)}^T)^{-1}, \\ H_{(i)} &\leftarrow (\text{orth}(A_{(i)}H_{(i)}^T))^T A_{(i)} \leftarrow G_{(i)}^{-1} A_{(i)} \\ &\leftarrow (G_{(i)}G_{(i)}^T)^{-1} G_{(i)}^T A_{(i)}, \\ A_{(i)} &\leftarrow \text{orth}(A_{(i)}H_{(i)}^T) (\text{orth}(A_{(i)}H_{(i)}^T))^T \\ &\quad + \mathcal{P}_\Omega(B_{(i)} - G_{(i)}H_{(i)}) \end{aligned} \quad (11)$$

for $i = 1, 2, 3, 4$ respectively,

$$\mathcal{P}_\Omega(B) = \begin{cases} B_{xy} & \text{if } (x, y) \in \Omega \\ 0 & \text{otherwise,} \end{cases}$$

where H^T represents the transpose of matrix H , H^{-1} represents the Moore-Penrose inverse of the matrix H , and $\text{orth}(A_{(i)}H_{(i)}^T)$ is a set of standard orthogonal bases formed by the columns of the matrix $A_{(i)}H_{(i)}^T$. The core strategy of the algorithm is that when a set of variables are optimized, other group variables can be treated as constants. In the process of optimizing the tensor completion problem, three sets of variables are $G_{(i)}$, $H_{(i)}$, and $A_{(i)}$, respectively. The $G_{(i)}$, $H_{(i)}$, and $A_{(i)}$ group includes 4 variables, respectively. In order to obtain the final optimization results, the solution method used in this paper is to optimize $G_{(i)}$, $H_{(i)}$, and $A_{(i)}$ alternately, and for each optimization, the specific value of $G_{(i)}$, $H_{(i)}$, and $A_{(i)}$ will be updated once. The first optimization of the algorithm uses the initial value, and once the algorithm reaches the convergence criterion, the optimization is stopped and the result is obtained. The algorithm flow chart is shown in Figure 7.

The convergence conditions in Figure 7 are shown as

$$\begin{aligned} \text{sc1} &= \frac{\|\mathcal{P}_\Omega(B_{(i)} - G_{(i)}^k H_{(i)}^k)\|_F}{\|\mathcal{P}_\Omega(B_{(i)})\|_F} \leq \text{tol}, \\ \text{sc2} &= \left| 1 - \frac{\|\mathcal{P}_\Omega(B_{(i)} - G_{(i)}^k H_{(i)}^k)\|_F}{\|\mathcal{P}_\Omega(B_{(i)} - G_{(i)}^{k-1} H_{(i)}^{k-1})\|_F} \right| \leq \frac{\text{tol}}{2}, \end{aligned} \quad (12)$$

where tol is the convergence threshold. This method converts the missing tensor recovery problem into a weighting multilinear low-rank decomposition form. Then, a differential

gradient strategy and a block iterative strategy are used to solve this problem.

Figure 8 shows the traffic volume and speed data collected by a microwave detector on January 5, 2016, and the result of missing data completion.

4. Solution of the Data Sparseness Problem Based on a Two-Dimensional Linear Interpolation Method

So far, using the tensor theory, the recovery of missing data is completed. Due to the large distance between every two sensors on the highway, the collected traffic data is obviously sparse, which leads to a large calculation error when the travel time is estimated. This paper introduces virtual nodes between real nodes and solves the sparse data problem using the two-dimensional linear interpolation algorithm. Then, the velocity data after interpolation is used to estimate the travel time between two toll stations.

4.1. Virtual Node Settings. As shown in Figure 9, this paper adds some virtual microwave detection nodes on the road, which makes the distance between two neighboring nodes less than 1 km after adding the virtual node. The actual node location is where RTMS is installed.

From Figure 9, it can be seen that the sparse data is as data missing at the virtual node. In order to solve this problem, this paper uses a two-dimensional linear interpolation method [5, 20] to interpolate the velocity data at the virtual node.

4.2. Two-Dimensional Linear Interpolation Method. For the convenience of the reader's understanding, this part introduces the two-dimensional linear interpolation method in the literature [5, 20].

The principle of the two-dimensional linear interpolation method is shown in Figure 10. The first step of the method is to calculate average speeds in time of adjacent detectors utilizing the temporal linear interpolation. The equations are as follows:

$$V(d, t) = V(d, h-1) + \frac{t-t_0}{\Delta T} \cdot [V(d, h) - V(d, h-1)], \quad (13)$$

$$V(d+1, t) = V(d+1, h-1) + \frac{t-t_0}{\Delta T} \cdot [V(d+1, h) - V(d+1, h-1)],$$

where $V(d, t)$ and $V(d+1, t)$ represent average speeds of detection points d and $d+1$ at time t , respectively; $V(d, h)$ represents the speed of detection point d in time period h ; t_0 is the start of time period h ; ΔT is the duration of h .

The second step is to calculate the average speeds of detectors by the spatial linear interpolation which can be expressed as

$$V(x, t) = V(d, t) + \frac{x(t) - x_d}{l} \cdot [V(d+1, t) - V(d, t)], \quad (14)$$

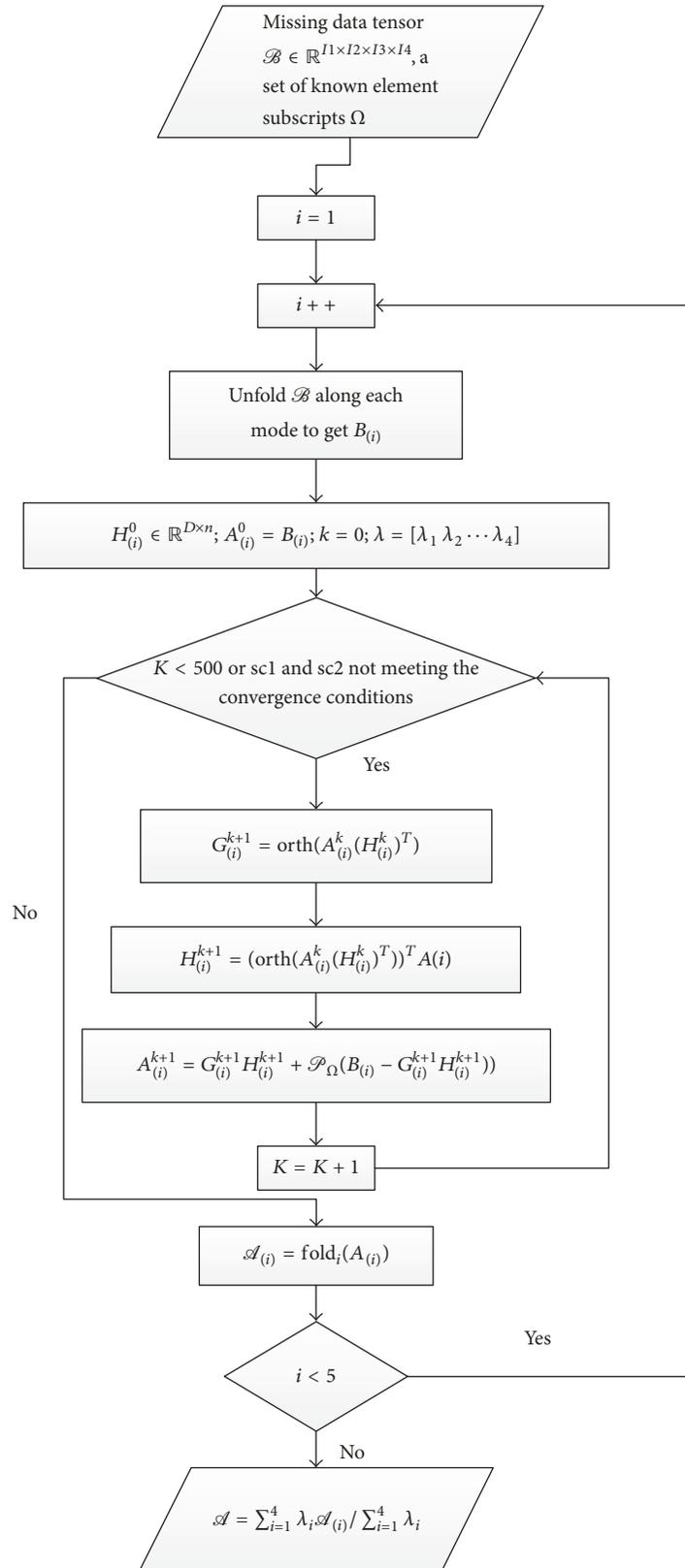


FIGURE 7: Flow chart of the algorithm.

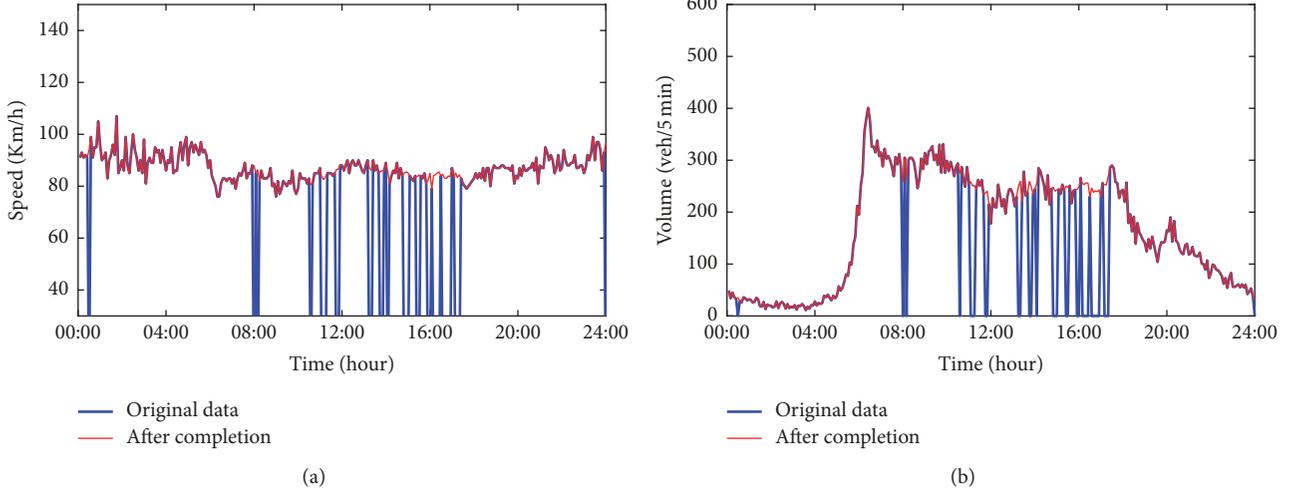


FIGURE 8: Tensor completion results (January 5, 2016). (a) Speed. (b) Volume.

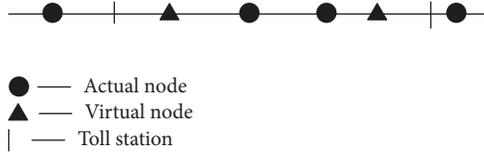


FIGURE 9: Virtual node diagram.

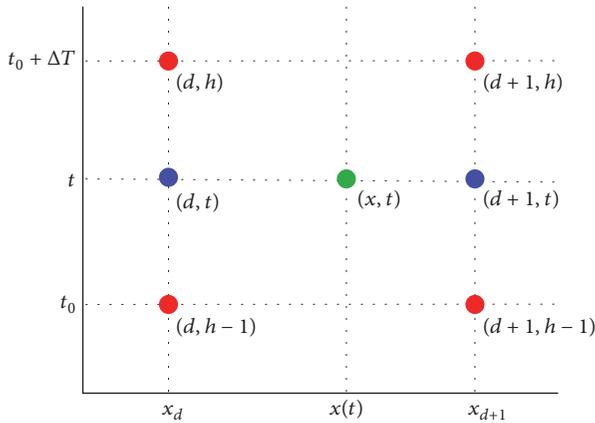


FIGURE 10: Principle of two-dimensional linear interpolation methods.

where $V(x, t)$ is the average speed of the detection point x at time t ; $x(t)$ is the position where the vehicle is at time t ; x_d is the position of detector d ; l is the distance between the detector d and the detector $d + 1$.

4.3. Travel Time Estimation. Then, the travel time is estimated based on the piecewise method as follows.

Based on Figure 11, the procedure of travel time estimation between two toll stations on the expressway is as follows.

Step 3. Using the average speed detected by RTMS before and after the toll station, the location average speed at the toll station is calculated by the two-dimensional linear interpolation method.

Step 4. The road sections between two neighboring RTMS nodes are defined as L_i . Calculate the travel time T_i on L_i .

If $L_i \leq 1000$ m, then

$$\bar{v} = \frac{(v_i + v_{i+1})}{2}, \quad (15)$$

$$T_i = \frac{L_i}{\bar{v}}.$$

If $L_i > 1000$ m, then the virtual nodes are added to make the distance between the nodes less than 1 km. The travel time between two nodes can be calculated using (15).

Step 5. Calculate the travel time between toll stations A and B, $t = \sum T_i$;

Figure 12 shows the distribution of the actual detector nodes and virtual detector nodes of Jinggangao highway (Dujiakan station to Jingliang Road station).

Take the travel time estimate of Dujiakan to Jingliang Road station at 8:00 on January 4 as an example. The velocity at the virtual node and the toll station is given by the two-dimensional linear interpolation method as shown in Table 3.

Finally, the travel time calculation result is 419 s.

5. Travel Time Prediction Based on the K -Nearest Neighbor Algorithm

In this paper, the authors developed a travel time forecasting algorithm based on the optimal K -nearest neighbor algorithm. The traditional KNN algorithm includes five procedures: data acquisition and input, state vector construction, pattern library establishment, K value calibration, and

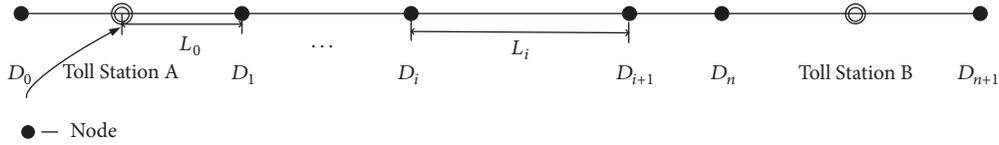


FIGURE 11: Toll station and RTMS node diagram.

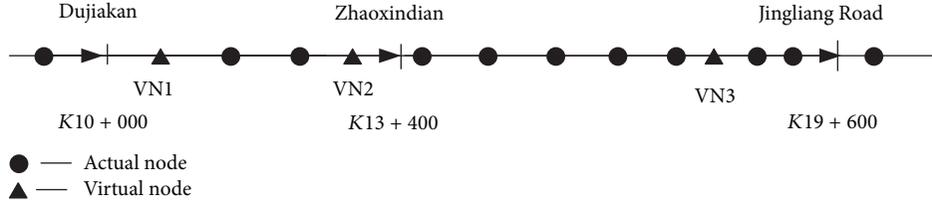


FIGURE 12: Relative position of detectors and toll stations. “▶” refers to the driving direction; “|” refers to the toll station.

TABLE 3: The section speed at the virtual node and the toll station.

Section	Dujiakan	Jingliang Road	VN1	VN2	VN3
Section speed (km/h)	46	93	64	85	90

prediction based on weighted average functions. Taking into account the deficiencies of the traditional KNN method, optimization is made in three aspects. Firstly, in addition to the three original state vectors of speed, volume, and time of day, the congestion levels are introduced as a new state vector. Secondly, by analyzing the change characteristic of traffic volume, the whole day is divided into seven periods. Thirdly, considering the low adaptability and precision caused by the fixed K value of the traditional KNN algorithm, the cross-validation method is used to calibrate the K value. The procedure of optimal K -nearest neighbor algorithm is introduced in this section.

5.1. State Vector Construction

5.1.1. Speed and Traffic Volume. The average speed and traffic volume of road sections were selected as the state vector to describe the road traffic status. As the traffic volume at the toll station cannot be measured directly, the traffic volume obtained by the first detector after the toll station consistent with vehicle heading direction was selected as a state parameter. Taking into account the continuity of the traffic state in the time dimension, the speed and volume of the first three cycles before the current state were selected as the state vector of the historical model library. A velocity vector and a traffic volume vector in the pattern library can be expressed as

$$\mathbf{M}(\bar{v}_{t-3T}, q_{t-3T}, \bar{v}_{t-2T}, q_{t-2T}, \bar{v}_{t-T}, q_{t-T}), \quad (16)$$

where t is the start time of the next travel time forecast period; T is the length of the predicted cycle; \bar{v}_{t-nT} indicates the average speed of the vehicle during the period $[t - nT, t -$

$(n-1)T]$; q_{t-nT} indicates the traffic volume during the period $[t - nT, t - (n-1)T]$.

5.1.2. Time of the Day. Time of day affects the travel time on the road; travel time is impacted by social activities and traveler habits, and the ratio of traffic volume to car type is different at different periods, so the travel times are not the same. According to [27], by analyzing the change of traffic volume at different times of the day, this paper divides the whole day into seven periods, as shown in Table 4.

5.1.3. Traffic Congestion Degree. Different weather, maintenance, accidents, and other events will affect the traffic conditions and travel times, so the road traffic condition needs to be joined into the state vector of the pattern library.

This paper uses the degree of the traffic congestion to characterize the impact of different events on road traffic conditions.

It was found that the change in road section average speed can reflect the fluctuation of the traffic volume condition. In this paper, the road section average speed was used as a standard to assess the degree of traffic congestion. Refer to the Ministry of Transport of the People's Republic of China, “Highway Network Operation Monitoring and Service Interim Technical Requirements.” The degree of traffic congestion was divided into four degrees as shown in Table 5.

5.2. Pattern Library Establishment. The state vector of the historical model library includes period, speed, traffic volume, and congestion degree. The decision attribute is travel time (estimation by method in Section 4). The state vectors and decision attributes were coded as shown in Table 6.

TABLE 4: Time of division in one day.

Time of the day	1	2	3	4	5	6	7
	Flat in the morning	Peak in the morning	Trough at noon	Peak at noon	Flat at night	Peak at night	Trough at night
Start time	6:00	9:00	11:00	14:00	17:00	19:00	21:00
End time	9:00	11:00	14:00	17:00	19:00	21:00	The next day 06:00

TABLE 5: Crowded classification standard.

Traffic congestion degree	1	2	3	4
Degree	Smooth	Slightly crowded	Crowded	Seriously crowded
Speed (km/h)	>70	(50, 70]	(30, 50]	≤30

TABLE 6: Property and encoding of each state component.

Parameter type	Basic parameters (average speed and traffic volume)	Time of the day	Traffic congestion level	Decision attribute (travel time)
Attributes and coding	$\bar{v}_{t-3T}, \bar{v}_{t-2T}, \bar{v}_{t-T}$ $q_{t-3T}, q_{t-2T}, q_{t-T}$	1: flat in the morning 2: peak in the morning 3: trough at noon 4: peak at noon 5: flat at night 6: peak at night 7: trough at night	1: smooth 2: slightly crowded 3: crowded 4: seriously crowded	Travel time

5.3. *K Value Calibration.* The basic idea of KNN algorithm is to set a quantity K to search up K -nearest points to the current state in the search area, and use these points to make predictions. Use all those points if there are fewer than K points in the search area, and keep the K -nearest points if there are more than K points in the search area.

Appropriate K values play a key role in accurate predictions [5]. In order to determine the best K value of the historical data library in each period, this paper uses a cross-validation method. The specific steps are as follows.

Step 6. Assume that the minimum and maximum K values are K_{\min} and K_{\max} , respectively.

Step 7. The historical data of every period are randomly divided into P parts, with each being recorded as DB_1, DB_2, \dots, DB_P . $DB_A, A = 1, 2, \dots, P$ are used as test data, and the other $P - 1$ parts are combined as the new historical data library.

Step 8. Calculate the average value of the prediction performance index using the test data under different K values.

Step 9. When the performance index is optimal, the corresponding K_T is the optimal K value of this period.

The optimal K value of the test road section was calibrated as follows. The minimum and maximum K values are 1 and 50, respectively, and the historical data library in each period is divided into 10 parts on average. Taking the historical data

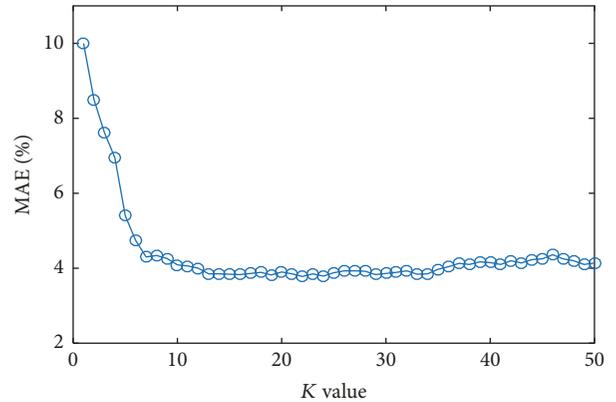


FIGURE 13: MAE under different values of K .

for the segments from Dujiakan to Jingliang Road station as an example, the results of the optimal K value calibration in different periods are shown in Table 7. Taking the peak period at night as an example, the average percentage change for absolute error in the case of different K values is shown in Figure 13. When $K = 22$, the mean absolute error is minimized.

5.4. *Travel Time Prediction Weighted Average Function.* In this paper, the weighted average method is used to make predictions based on the nearest neighbor data. Assuming

TABLE 7: Optimal values of K .

Time of the day (coding)	1	2	3	4	5	6	7
K value	20	14	13	11	16	22	7

that the searched K -nearest neighbors are $M_1 \sim M_K$, the distance between them and the current state vector is $D_1 \sim D_K$, and their decision attributes (the measured values of time) are $t_1 \sim t_K$, respectively. The predicted value of travel time (TP) is

$$TP = \sum_{i=1}^K \beta_i t_i, \quad (17)$$

$$\beta_i = \frac{D_i^{-1}}{\sum_{i=1}^K D_i^{-1}}.$$

D is the Euclidean distance; the Euclidean distance between two n -dimensional vectors $\mathbf{X}_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $\mathbf{X}_2 = (x_{21}, x_{22}, \dots, x_{2n})$ is calculated as

$$D = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2}. \quad (18)$$

6. Test Case and Error Analysis

Taking travel time prediction for the Dujiakan to Jingliang Road station as an example, the distance between the two stations is 9.6 km. Remote transportation microwave sensor data collected from January 4, 2016, to February 29, 2016, were selected as training data. Remote transportation microwave sensor data collected from March 1 to March 15 were used as test data.

Firstly, the travel time was predicted based on original training data. The state vector is constructed based on the original training data. The travel time is estimated by the two-dimensional linear interpolation method and the segmentation method in Section 4 as the decision attribute. Then, the model library is constructed. The travel time is predicted by the K -nearest neighbor algorithm based on the test set data.

Secondly, the travel time was predicted based on data after completion. The static tensor filling method was used to fill the original data, and the state vector is constructed based on the completion data. The travel time is reestimated by two-dimensional linear interpolation and segmentation method as the decision attribute. Then, the model library is constructed. The K -nearest neighbor algorithm is used for prediction based on the test set data to obtain the travel time results.

Thirdly, the two-dimensional linear interpolation method and the segmentation method are used to estimate the travel time (ranges from March 1 to March 15) based on the test data and the result is used as the comparison data to calculate the prediction error.

In order to evaluate the travel time prediction results, the relative error (RE) and the mean absolute percentage

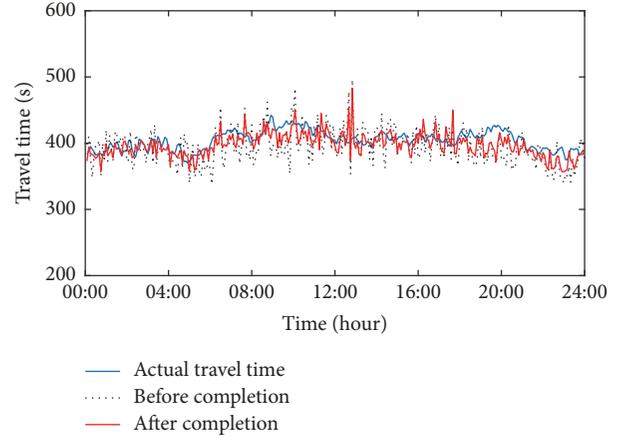


FIGURE 14: Travel time prediction result curves (03/02/2016).

error (MAPE) are used to measure the accuracy. The error is calculated as follows:

$$RE(t_p) = \frac{TP(t_p) - T(t_p)}{T(t_p)} \times 100\%, \quad (19)$$

$$MAPE = \frac{1}{L} \sum_L \frac{|TP(t_p) - T(t_p)|}{T(t_p)} \times 100\%,$$

where $RE(t_p)$ is the relative error for the prediction period t_p ; $TP(t_p)$ is the forecast travel time for the period t_p ; $T(t_p)$ is the actual travel time of the period t_p ; MAPE is the average absolute relative percentage error; L is the total number of validation forecast periods.

The prediction results on March 2, 2016, March 4, 2016, and March 5, 2016, are selected as an example to analyze the prediction error of our method.

The actual travel time curve and the predicted travel time before and after data completion on March 2, 2016, March 4, 2016, and March 5, 2016, are shown in Figures 14, 16, and 18. It can be seen from Figures 14, 16, and 18 that the K -nearest neighbor algorithm can predict the travel time accurately, and the predicted travel time after the missing data is filled is closer to the actual travel time. The relative errors before and after completion are shown in Figures 15, 17, and 19. The errors before and after completion are shown in Table 8. It can be seen that the error is improved after data completion, and the mean absolute percentage error of the whole day before and after optimization is reduced.

7. Conclusion

This paper studies the travel time prediction for expressways based on RTMS data.

TABLE 8: Travel time prediction error.

Evaluation index	03/02/2016		03/04/2016		03/05/2016	
	RE	MAPE	RE	MAPE	RE	MAPE
Before completion	(-19.2, 15.3%)	7.4%	(-17.4%, 21.9%)	6.4%	(-14.9%, 12.1%)	6.1%
After completion	(-14.3, 10.8%)	3.1%	(-12.3%, 18.7%)	3.0%	(-11.1%, 9.6%)	2.8%

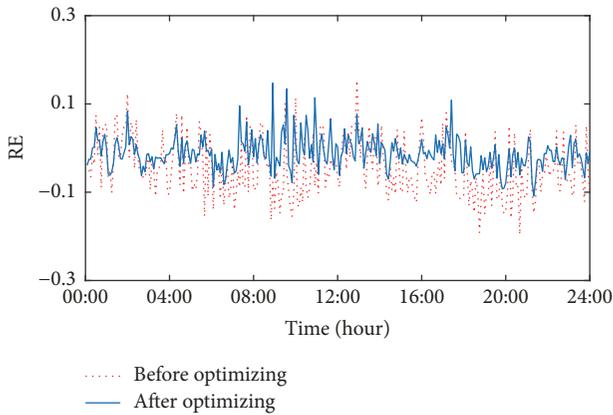


FIGURE 15: RE curves (03/02/2016).

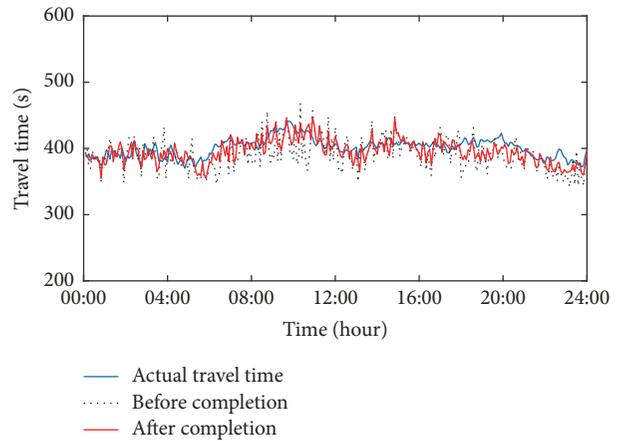


FIGURE 18: Travel time prediction result curves (03/05/2016).

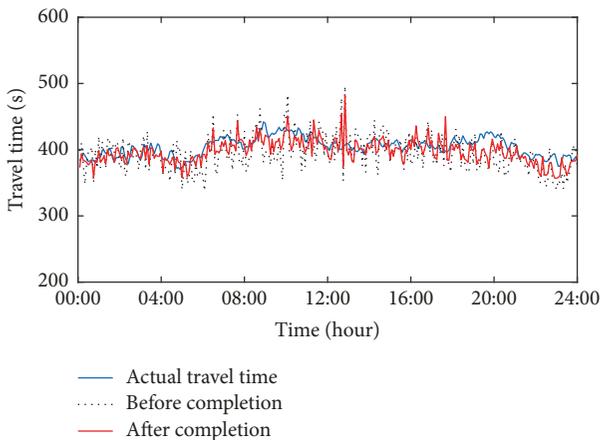


FIGURE 16: Travel time prediction result curves (03/04/2016).

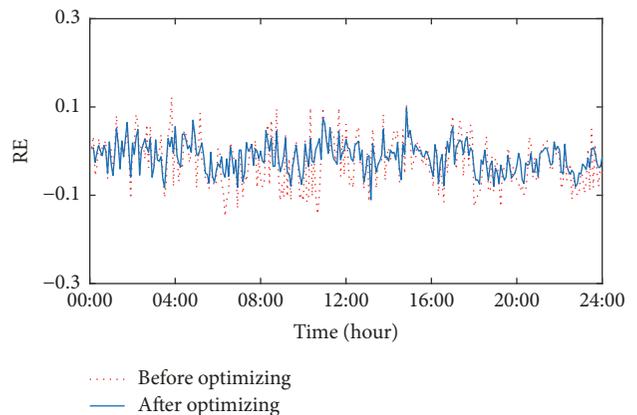


FIGURE 19: RE curves (03/05/2016).

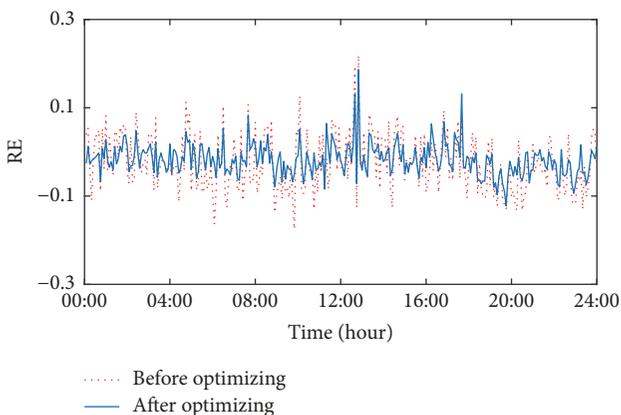


FIGURE 17: RE curves (03/04/2016).

(1) Considering the influence of the randomly missing data on prediction accuracy, the correlation and low-rank characteristics of the microwave vehicle data are analyzed, and then the traffic tensor model is established. Based on multilinear low-rank decomposition, a tensor completion algorithm is used to fill the random missing traffic data.

(2) Virtual nodes are introduced between real nodes, and the sparse data problem is solved using the two-dimensional linear interpolation algorithm. Then, the velocity data after interpolation is used to estimate the travel time between toll stations.

(3) The optimal K -nearest neighbor algorithm is proposed for travel time prediction, with the travel time as the decision attribute and the speed, traffic volume, time of the day, road section, and traffic congestion level as the state vector. The historical model library is established. Taking

the Euclidean distance as the criterion, the similarity pattern is searched using the K -nearest neighbor strategy, and the weighted average method is used to predict travel time. Practical data collected from Jinggangao highway in Beijing is used to verify the algorithm. The results show that the optimal K -nearest neighbor algorithm can accurately predict travel time, and the accuracy of prediction results using data after completion is higher than when using original data.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Jiandong Zhao and Yuan Gao presented the algorithms and co-wrote the paper; Jinjin Tang, Lingxi Zhu, and Jiaqi Ma analyzed the data.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (2016JBM053).

References

- [1] B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 351–364, 2002.
- [2] L. Li, X. Chen, Z. Li, and L. Zhang, "Freeway travel-time estimation based on temporal-spatial queueing model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1536–1541, 2013.
- [3] Z. Zhang, D. Yang, T. Zhang, Q. He, and X. Lian, "A study on the method for cleaning and repairing the probe vehicle data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 419–427, 2013.
- [4] Y. Zhang and Y. Zhang, "A comparative study of three multivariate short-term freeway traffic flow forecasting methods with missing data," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, pp. 1–14, 2016.
- [5] J. Zhao, F. Xu, W. Liu, J. Bai, and X. Luo, "Travel time prediction based on pattern matching method," *International Journal On Smart Sensing and Intelligent Systems*, vol. 8, no. 1, pp. 658–676, 2015.
- [6] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-based missing data imputation for traffic flow volume: a systematical approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 512–522, 2009.
- [7] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 29–40, 2015.
- [8] Y. Duan, Y. Lv, Y.-L. Liu, and F.-Y. Wang, "An efficient realization of deep learning for traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 168–181, 2016.
- [9] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 108–120, 2013.
- [10] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: an integrated study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3430–3439, 2015.
- [11] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.
- [12] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2123–2133, 2016.
- [13] B. Ran, H. Tan, Y. Wu, and P. J. Jin, "Tensor based missing traffic data completion with spatial-temporal correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 446, pp. 54–63, 2016.
- [14] B. Ran, L. Song, J. Zhang, Y. Cheng, and H. Tan, "Using tensor completion method to achieving better coverage of traffic state estimation from sparse floating car data," *PLoS ONE*, vol. 11, no. 7, Article ID e0157420, 2016.
- [15] H. Zhou, D. Zhang, K. Xie, and Y. Chen, "Spatio-temporal tensor completion for imputing missing internet traffic data," in *Proceedings of the 34th IEEE International Performance Computing and Communications Conference, IPCCC*, China, December 2015.
- [16] C. A. Quiroga, *An Integrated GPS-GIS Methodology for Performing Travel Time Studies*, Louisiana state University, Baton Rouge, La, USA, 1997.
- [17] C. A. Quiroga and D. Bullock, "Travel time studies with global positioning and geographic information systems: an integrated methodology," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 1-2, pp. 101–127, 1998.
- [18] X. Wang, L. Peng, and T. Chi, "A method of urban traffic flow speed estimation using sparse floating car data," *Cehui Xuebao/Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 7, pp. 866–873, 2016.
- [19] F. F. Xu, *Highway Traffic Congestion Detection and Travel Time Prediction Based on Microwave Data*, Beijing Jiaotong University, Beijing, China, 2016.
- [20] J. D. Zhao, F. F. Xu, Y. J. Guo, and Y. Gao, "Traffic congestion detection based on pattern matching and correlation analysis," *Advances in Transportation Studies*, no. 40, pp. 27–40, 2016.
- [21] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, "Hybrid short-term freeway speed prediction methods based on periodic analysis," *Canadian Journal of Civil Engineering*, vol. 42, no. 8, pp. 570–582, 2015.
- [22] X. Fei, Y. Zhang, K. Liu, and M. Guo, "Bayesian dynamic linear model with switching for real-time short-term freeway travel time prediction with license plate recognition data," *Journal of Transportation Engineering*, vol. 139, no. 11, pp. 1058–1067, 2013.
- [23] J. J. Tang, F. Liu, Y. J. Zou, W. B. Zhang, and Y. H. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 2340–2350, 2017.
- [24] J. Zhao, H. Wang, W. Liu, and J. Bai, "Highway travel time prediction between stations based on toll ticket data," *Tongji Daxue Xuebao/Journal of Tongji University*, vol. 41, no. 12, pp. 1849–1854, 2013.

- [25] Y.-X. Sun, W. Guan, Y. Ge, G.-H. Zhang, and X. Yang, "Short-time forecasting of passenger travel time based on SVM method," *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 34, no. 6, pp. 1587–1592, 2014.
- [26] L. C. Huang, *Research on Link Travel Time Estimation and Prediction Based on Data Driven*, Beijing Jiaotong University, Beijing, China, 2015.
- [27] X. Wang, X.-H. Chen, and X.-M. Yang, "Short term prediction of expressway travel time based on K nearest neighbor algorithm," *Zhongguo Gonglu Xuebao/China Journal of Highway and Transport*, vol. 28, no. 1, pp. 102–111, 2015.

