Research Article

Short-Term Traffic Speed Prediction Method for Urban Road Sections Based on Wavelet Transform and Gated Recurrent Unit

Xin Fu,[1,2] Wei Luo [ID],[1] Chengyao Xu,[1] and Xiaoxuan Zhao[1]

[1]*College of Transportation Engineering, Chang'an University, Xi'an 710064, China*
[2]*Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an, Shaanxi 710064, China*

Correspondence should be addressed to Wei Luo; 2018123030@chd.edu.cn

As a core component of the urban intelligent transportation system, traffic prediction is significant for urban traffic control and guidance. However, it is challenging to achieve accurate traffic prediction due to the complex spatiotemporal correlation of traffic data. A road section speed prediction model based on wavelet transform and neural network is, therefore, proposed in this article to improve traffic prediction methods. The wavelet transform is used to decompose the original traffic speed data, and then the coefficients obtained after the decomposition are used to reconstruct the high-frequency random sequences and the low-frequency trend sequence. Secondly, a GRU neural network is constructed to learn the trend of low-frequency sequence. The spatiotemporal correlation between input data is extracted by adjusting the input of the model. Meanwhile, an ARMA model is used to fit unstable random fluctuations of high-frequency sequences. Last of all, the prediction results of the two models are added together to obtain the final prediction result. The proposed prediction model is validated by using road section speed data based on the floating car data collected in Ningbo. The results show that the proposed model has high accuracy and robustness.

1. Introduction

With the socioeconomic development and the acceleration of urbanization, the demand for transportation continues to grow. Though the urban transportation system has been developing as well, trying to match with the increasing demand for transportation, and its supply capacity is improving through the construction of transportation infrastructure, it is still one step behind. Congestion has become a universal problem, causing headache all over the word. Problems brought in by the overcrowding of urban roads include economic, health, and environmental problems, such as stress, fuel consumption, wasted time, and traffic accidents [1]. The intelligent transportation system (ITS) can effectively control and induce urban traffic. It is a key way to solve urban traffic problems, so it puts forward higher requirements for accurate intelligent traffic services [2]. The objects of traffic prediction are traffic parameters such as vehicle flow, speed, and occupancy in a specific area and time period [3]. High-precision traffic prediction can provide

accurate travel information for urban residents, and it helps to realize ITS applications [4]. Real-time traffic data can be obtained effectively with the development of ITS [5]. Therefore, in order to improve the accuracy and reliability of traffic prediction, researchers are committed to develop and improve effective traffic prediction models based on fully mined historical traffic data. In this paper, traffic speed prediction on road sections is studied using historical traffic speed data.

It is the research focus of scholars to consider the spatial correlation between road sections in traffic prediction models due to the complex spatiotemporal correlation between road sections. The traffic conditions of each section of the road network are often affected by the conditions of its upstream and downstream sections. For example, traffic congestion often starts on one or more sections and spreads to other sections after a period of time, resulting in regional congestion [6]. Regarding this trait of congestion, some scholars in early years had built nonparametric models using speed data of the studied road section and its upstream and

downstream sections, which can better capture the spatio-temporal correlation between road sections and thus will improve the accuracy of the prediction models [7].

Due to their high flexibility, good learning, and generalization capabilities, algorithms based on neural networks have been widely used in transportation-related tasks [8]. Recurrent neural network (RNN) is applied to traffic prediction because of its special internal structure capable of effectively processing time series. RNN used for traffic prediction mainly includes Long Short-Term Memory (LSTM) neural networks and Gated Recurrent Unit (GRU). GRU was proposed by Cho et al. in 2014 and it is also reported that GRU achieves equal to or better performance than LSTM [9]. In addition, it is proven that GRU outperformed LSTM on nearly all tasks except language modelling with the naive initialization [10]. Dai et al. used GRU network to make short-term traffic forecasts. In the study, GRU was used to process the spatiotemporal feature information of the internal traffic flow of the matrix to achieve the purpose of prediction [11]. In another hybrid model which predicts lane speed, a GRU network layer was used to achieve the final speed prediction [12]. These studies have shown that GRU is competent on traffic prediction and achieve good results.

At the same time, the traffic speed data generated by each road section also has complex attributes. Firstly, the real-time and dynamic nature of urban traffic results in strong random fluctuations in traffic speed data. The difficulty of traffic prediction is enhanced by this inherent characteristic. Secondly, urban traffic has certain spatiotemporal characteristics and periodic laws, which makes the traffic data of each road section have a relatively constant change trend. For example, commuting sections have low traffic speed during morning and evening rush hours and have high traffic speed during other hours. This makes the trend of traffic data over time more predictable. In order to effectively learn the stable periodic characteristics of traffic data and random fluctuations under real-time dynamic traffic, wavelet analysis theory is applied to perform traffic prediction. The historical traffic data are decomposed into subsequences from high to low in terms of frequencies by wavelet transform (WT). The low-frequency sequence contains characteristics similar to the original data. From a long-term perspective, the volatility trend of low-frequency sequence has a repeating daily periodicity. From a short-term perspective, the data at the moments before and after the sequence are similar and continuous [13]. In the past, some scholars used different models to predict the sequence after WT, including ARIMA model and BP neural network [14–16]. However, most of the inputs of these combined prediction models are single time series, which ignores the correlation between the traffic data generated by spatially adjacent road sections.

Therefore, a hybrid prediction framework has been proposed to make more accurate predictions in this paper. Firstly, the historical traffic speed data are decomposed into subsequences from high to low in terms of frequencies by WT. Moreover, a GRU network is constructed to learn the development trend of low-frequency sequence. By controlling the input of this model, the spatiotemporal correlation of traffic speed data can be extracted effectively. At the same time, an autoregressive moving average (ARMA) model is used to fit the random fluctuations on the high-frequency sequences. Finally, the prediction results of the two models are added together to represent the final prediction result.

The rest of the paper was organized as follows. Previous research studies are discussed in Section 2. And the prediction model proposed will be explained by us in Section 3. In Section 4, the validity and robustness of the model are proven using the speed data set in Ningbo, China. The conclusions and future work of this paper are summarized in the end.

2. Related Work

While long-term prediction predicts the future traffic demand using data such as socioeconomic attributes, the traffic prediction required by ITS is a short-term traffic prediction which mainly focuses on the traffic conditions in next few minutes to several hours [1]. In the past few decades, a series of studies on traffic prediction have been implemented by researchers. The existing methods can be divided into three categories: parametric methods, nonparametric methods, and hybrid model prediction methods.

The structure of prediction model based on the parameter methods is rather simple. Parametric methods such as Autoregressive Integrated Moving Average (ARIMA) [17, 18] and Kalman filtering [19, 20] achieve promising results but they rely on certain physical or statistical assumptions. However, traffic flow has the characteristics of randomness and nonlinearity, and it is difficult to establish a parameter model that can reproduce the characteristics of traffic flow in practice. The gradual popularization of sensors on urban roads and GPS on vehicles enabled the acquisition of real-time traffic data. Therefore, the nonparameter method of modelling using large amount of historical data is widely applied. Nonparametric models mainly include the most traditional statistical machine learning methods and the most popular artificial intelligence algorithms. Support vector machine (SVM) [21, 22], K-nearest neighbor (KNN) [23, 24], and artificial neural network (ANN) are the most widely used ones [25].

Intelligent algorithms have attracted widespread attention not only in academia and industry but also in the field of transportation in recent years. RNN is widely recognized as a suitable method to capture the temporal evolution of traffic flow. LSTM, as a typical representative of RNN, was used by many scholars for short-term traffic prediction. Ma et al. applied LSTM to road traffic speed and volume prediction for the first time. By introducing the forget gate, the network can internally connect time series data and increase prediction accuracy [26]. Experimental results show that this network model is superior to ordinary neural networks. The variant GRU of LSTM, which was proposed in 2014, has also been used in traffic prediction because of its simpler structure and similar effects with LSTM [11, 12]. Convolutional neural networks (CNN) algorithm is widely used

in computer vision and image classification [27]. In 2017, CNN was proven to be suitable for traffic speed prediction of the entire road network by learning traffic as images [28]. However, due to the complexity of the topological structure of urban road networks, it is difficult for traditional CNN networks to obtain the spatial characteristics of irregular grid structures. A model called graph convolutional network (GCN) is used to extract the spatial correlation between the road sections. For example, a spatiotemporal GCN (STGCN) model was proposed to extract the spatiotemporal dependence of road network traffic speed data and make predictions [29]. In addition, a model called diffusion convolutional CNN (DCRNN), which combines GCN and RNN at the same time, models the spatial dependence of traffic as a diffusion process on a directed graph and uses RNN to fit temporal correlation [30]. An unsupervised algorithm named sparse autoencoder (SAE) was firstly used to identify and predict traffic state [31]. Furthermore, a deep belief network (DBN) trained with a greedy unsupervised method is also used to predict the traffic speed of an arterial in Beijing [32]. Most neural network models do not provide a reasonable explanation, which is different from statistical machine learning algorithms. However, neural network models have higher prediction accuracy, especially when dealing with large amounts of data.

Short-term traffic prediction can be affected by many factors due to different prediction scenarios. A single prediction model may not be suitable for all scenarios. Fusco et al. built a two-layer model combining Bayesian network and neural network and verified it with floating vehicle data [33]. There are also scholars who combined unsupervised learning algorithms with supervised learning algorithms for model. A prediction model combining SAEs and LSTM is proposed, which uses SAEs and LSTM to extract the spatial and temporal correlations of traffic data, respectively [34]. In the research of Duan, CNN for extracting spatial features and LSTM for capturing temporal information were combined [35]. The method based on wavelet transform has also been applied to predicted the traffic. After the data is decomposed and reconstructed by WT, the neural network optimized by particle swarm is used to predict the sequence separately [36]. A speed prediction study down to the lane scale has also been proposed recently. The researchers built a two-layer deep learning framework that combines LSTM and GRU to predict the speed of different lanes [12]. The combined prediction model combines different prediction algorithms, which can give full play to the advantages of each model to obtain better predicted results.

3. Methodology

3.1. Framework Overview.

A W-GRU-ARMA model was constructed for short-term traffic speed prediction in this section. As its name implies, the model is composed of three parts: wavelet transform (W), GRU, and ARMA. The spatiotemporal relationship of urban traffic speed data is taken into account in this model. It focuses on the short-term correlation of traffic speed data for the predicted road sections in the time dimension and the spatial correlation

with the upstream and downstream sections. Figure 1 shows the overall architecture. After the wavelet transform of the traffic speed time series data, different low-frequency and high-frequency sequences will be obtained. Therefore, a GRU model for predicting low-frequency trend sequence and an ARMA model for predicting high-frequency random sequences are established. The final prediction result is obtained by summing the prediction results of each model.

The input of the constructed GRU prediction model contains the traffic speed data for predicting road sections and its upstream and downstream sections. For example, the road section A is selected as a research object, and the speed of the road section A at $t + 1$ is used as a prediction target. Then, the input vector X of the GRU model is as follows:

$$X = \left[x_{i+1,t-m}, \ldots, x_{i+1,t-1}, x_{i+1,t}, x_{i,t-m}, \ldots, x_{i,t-1}, x_{i,t}, \right.$$
$$\left. x_{i-1,t-m}, \ldots, x_{i-1,t-1}, x_{i-1,t} \right], \quad (1)$$

where i represents the predicted road section, $i + 1$ and $i - 1$ represent the downstream and upstream sections of the predicted road section, respectively, t represents current time, and $t - m$ represents the m previous time. The value of m is determined according to the algorithm performance during the model training phase.

3.2. Wavelet Transform.

Wavelet transform is a method for processing nonstationary and nonlinear signals with the advantages of multiresolution and multiscale. Wavelets provide an output in terms of the time-frequency scale, which can approach the original signal at any scale and capture the details of the original signal [37]. In nonstationary data analysis, wavelet transform is often used to extract the trend information of sequence changes [38]. The discrete wavelet transform (DWT) can decompose the original traffic speed data into a series of multiple frequencies. The Mallat algorithm is efficient in nonstationary traffic speed time series into sequences of different frequencies through high-pass filters and low-pass filters. The outputs of the low-pass filter and the high-pass filter are defined by dA and dD in equations (2) and (3) which are called approximate coefficients and detail coefficients:

$$dA = \sum_{k=-\infty}^{\infty} X[k] \varphi_l [2n - k], \quad (2)$$

$$dD = \sum_{k=-\infty}^{\infty} X[k] \varphi_h [2n - k], \quad (3)$$

where X is the original signal, φ represents the filter, and the subscribe l and h represent the low-pass filter and high-pass filter, respectively.

Figure 2(a) demonstrates the process of Mallat algorithm for two-level decomposition. The original time sequence data X is put through both low-pass filter and high-pass filter, and the outputs are $dA1$ and $dD1$ of the first layer, respectively. Then, the obtained $dA1$ is passed through two
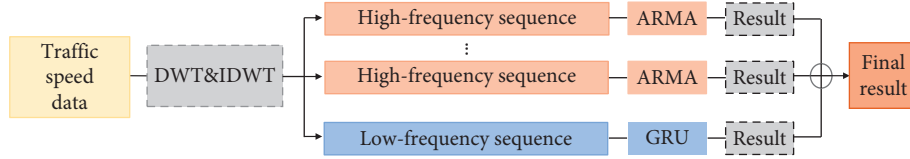
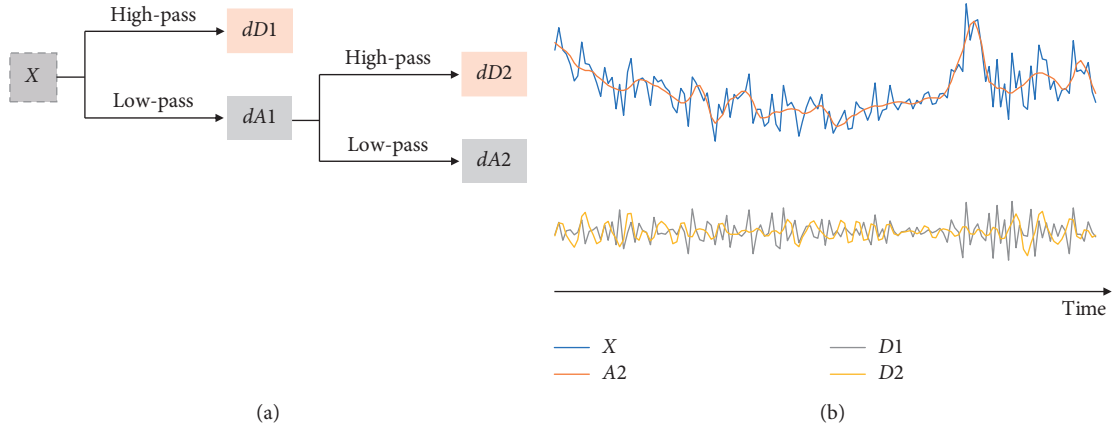Figure 1: The framework of the proposed model.

Figure 2: Discrete wavelet transform. (a) Decomposition process. (b) Sequence reconstruction.

filters again to obtain two coefficients $dA2$ and $dD2$. After decomposition, time series components with different frequencies can be obtained, but the lengths of the components after decomposition are not equal. The sequence length is reduced by half after the decomposition. Therefore, inverse discrete wavelet transform (IDWT) is used to reconstruct the data using approximate coefficients and detail coefficients to obtain sequences that are equal to the length of the original sequence but with different frequencies. As shown in Figure 2(b), the approximate coefficient $dA2$ is used to reconstruct the low-frequency component to form a sequence $A2$; the detail coefficients $dD1$ and $Dd2$ are used to reconstruct the high-frequency component to obtain the sequences $D1$ and $D2$.

3.3. Gated Recurrent Unit. RNN has a wide range of applications in the field of time series analysis. It can implement a mechanism similar to the human brain and maintain a certain memory of the processed information. However, traditional RNN models are prone to vanishing gradients and gradient explosions during training [39]. A variation of RNN called LSTM is proposed to solve the problem effectively. The cells of hidden layers for LSTM have a special structure compared with traditional neuron nodes, which is the key to the long-term dependence of LSTM learning time series. Figure 3(a) shows the cell structure of the hidden layer of LSTM. Information inflows, outflows, and previous status updates can be achieved by adding input gates, output gates, and forget gates to this cell structure. The forget gate is responsible for determining how much of the previous cell state is retained in the current cell state, the input gate is responsible for determining how many inputs are retained in

the current cell state, and the output gate is responsible for determining the output of the current cell state.

Gated Recurrent Unit (GRU) is a variation of LSTM networks. It inherits the advantages of RNN model: it automatically learns features and effectively models long-term-dependent information. It has been applied to short-term traffic prediction successfully [11, 12]. Figure 3(b) shows the cell structure of the hidden layer of GRU network, and it is more simply compared with LSTM, obviously. Intuitively, the input and forget gates in LSTM were integrated as a reset gate in GRU [9], which determines how to combine the new input information with the information from the previous time. Another gate in GRU is called update gate; it determines how much of the information from the previous time can be saved to the current time. Therefore, GRU is one gate less compared to LSTM. In addition, the cell state and hidden state in LSTM have been integrated as one hidden state in GRU. These changes make the GRU network have fewer parameters and faster training speed and require less data to generalize the model effectively [40]. The calculation formula of GRU is as follows:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right), \tag{4}$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right), \tag{5}$$

$$\overline{h}_t = \tanh\left(W \cdot [r * h_{t-1}, x_t]\right), \tag{6}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \overline{h}_t. \tag{7}$$

Formulae (4) and (5) show how the update gate z_t and reset gate r_t are calculated in GRU neurons. W_z denotes the weight of z_t, W_r denotes the weight of r_t, and σ denotes the

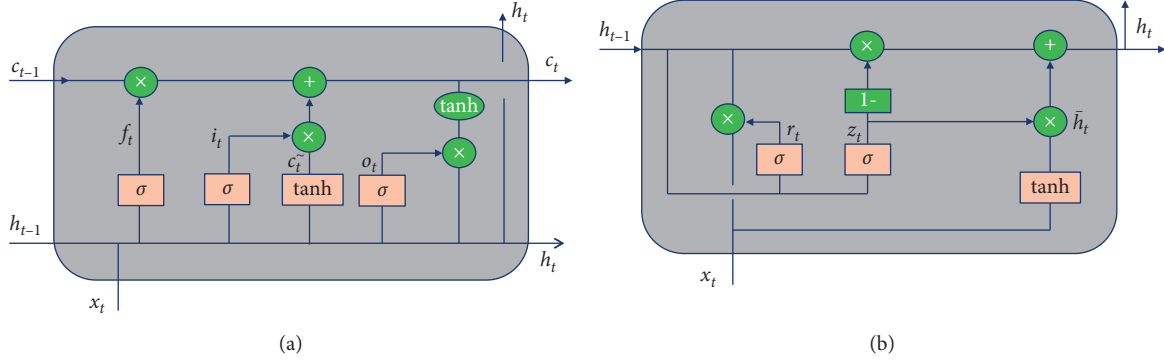(a)   (b)

FIGURE 3: The structure of LSTM and GRU. (a) LSTM. (b) GRU. x_t is input. h_t and h_{t-1} are the output of the current cell and previous cell. In (a), i_t, f_t, and o_t are the input, forget, and output gates, respectively. c_{t-1} and c_t denote the memory cell and the new memory cell content. In (b), r_t and z_t are the reset and update gates, respectively. $\bar{h}_t$ is the pending output.

sigmoid function. The innermost term $[h_{t-1}, x_t]$ represents the sum of vectors h_{t-1} and x_t. A larger value of z_t indicates that more information has been maintained by the current cell while the less for the previous cell. r_t suggests that when the value of the equation is equal to 0, the information from the previous cell is discarded.

Formulae (6) and (7) show the calculation of the pending output value $\bar{h}_t$ and final output value h_t of the GRU neural network. h_{t-1} represents the output from previous cell, W denotes the weight of the z_t, and tanh denotes the hyperbolic tangent function. $\bar{h}_t$ is obtained by multiplying h_{t-1} of the previous cell by r_t, plus x_t, multiplying by the W, and using the hyperbolic tangent function. h_t is the sum of two vectors. One is obtained by multiplying $1-z_t$ by h_{t-1} and another one is obtained by multiplying zt by $\bar{h}_t$.

3.4. Autoregression Moving Average Model.

Autoregressive moving average (ARMA) model is the most common type of time series models used for stationary random process analysis [41]. This method does not require strong similarity between the data at the predicted time and the data at the previous time. The method could smooth the predicted values at excessive fluctuations by averaging several measured data. The high-frequency sequences generated by WT have smooth fluctuations. Therefore, this article makes use of ARMA to predict the high-frequency sequences and simulate the random fluctuation of original data caused by the real-time and dynamic nature of traffic. The basic model of the autoregressive moving average model is ARMA (p, q), which consists of two parts, namely, the autoregressive model (AR) and the moving average model (MA). The basic expression is as follows:

$$\tilde{y}_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i}^{high} + \sum_{i=1}^{q} \lambda_i \varepsilon t - i, \tag{8}$$

where c is a constant, ε_t denotes the random error of the Gaussian white noise distribution, φ and λ are the parameters of the models AR and MA, and p and q refer to the orders of the models AR and MA. On the left side of the equation, $\tilde{y}_t$ denotes the predicted result of the ARMA

model, corresponding to the predicted value of the traffic speed on the high-frequency sequences at time t.

4. Experiments

4.1. Data.

The data used to evaluate the performance of the proposed model were the floating car speed data collected in Ningbo. The raw data were uploaded by the GPS equipment on about 4,300 taxis every day from July 1, 2017, to July 30, 2017. The GPS equipment records the running status of the vehicle every fifteen seconds. Each recorded piece of floating car speed data included record time, vehicle speed (instantaneous speed), location (latitude and longitude), and direction of travel. These data-high-frequency floating vehicle data-can reflect detailed vehicle dynamics [42]. In this study, a representative busy area in Ningbo was selected for the study, and the road network was divided into 283 sections based on the existence of intersections (Figure 4). The vector road network data were obtained from Open Street Map.

In order to use the raw data to estimate the average road speed, data cleaning was performed in the first place. The erroneous data with incorrect time and location have been deleted, and the abnormal speed values were identified and removed by the interquartile range method [43]. Since the speed of urban sections will be affected by the intersection [42, 44], the data uploaded by GPS when the vehicle is suspended at the intersection temporarily were kept, which makes the final estimated results more realistic. Then, the Feature Manipulate Engine (FME) platform is used to estimate the speed of the road sections. FME is a set of solution customization software for spatial and nonspatial data analysis, processing, and conversion. Through this platform, a geometric map matching algorithm [45] is used to match the cleaned spatial data with direction attributes to the road network. Meanwhile, an algorithm for estimating the average vehicle speed at the interval of ten minutes is designed. In the algorithm, the average value of consecutive track points of the same vehicle on the same road section within the same ten minutes represents the average speed of a single vehicle (SV). And the final average speed of a road section (RV) is the mean of the all SV. After all these, the missing values were imputed by the linear interpolation [43]. Finally,
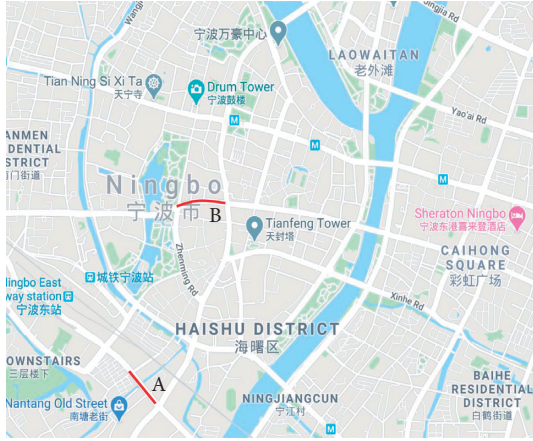
FIGURE 4: Road networks of the target area.

FIGURE 5: Speed estimation for some road sections of the target area.

148 speed values are obtained every day and we choose data from 6 a.m. to 24 p.m. (108 time steps in 1 day) as experimental data. A time-space traffic diagram, in which x-axis is time, y-axis is space, and the color inside represents speed [46], is used to demonstrate the processed data, as shown in Figure 5.

Next, road section B, which is adjacent to the hospital, and the main channel section A of the city are the main research objects, as shown in Figure 4. The calculated results of the two road sections are shown in Figure 6. Road section B has a low overall speed during the daytime, and the pattern is not obvious. However, road section A has a more obvious morning and evening peak trend. According to the demand of the prediction model and the abovementioned data processing steps, the average speed of the road section in every driving direction from 06:00 to 24:00 every 10 min was obtained. After that, a time series of road section speeds with 108 data per day and a total of 3240 data over 30 days were generated. The road section speed data with an effective length of 30 days were divided into two parts according to a ratio of 8 : 2. The data of the first 24 days were used to train the model, and the data of the next 6 days were used to test the model.

4.2. Accuracy Indicators and Experimental Setup.
Two measurements are used as performance indicators for the accuracy of the proposed model of short-term prediction. They are Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), as shown in Formulae (8) and (9). MAPE is the relative error of the prediction, and RMSE provide the deviation of the predicted value from the actual value. These measurements help us to better understand the prediction results:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\widetilde{x}_i - x_i|}{x_i}, \tag{9}$$

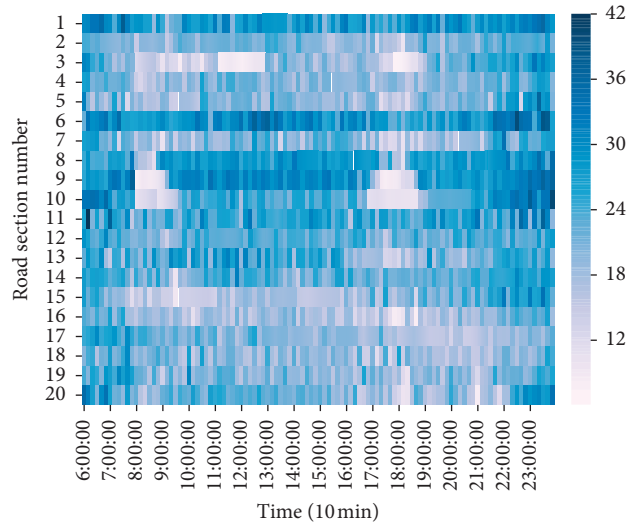$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\widetilde{x}_i - x_i)^2}{n}}, \tag{10}$$

where $\widetilde{x}$ is the predicted value at time i, x is the actual value at time i, and n is the number of predicted values.

According to the characteristics of wavelet decomposition, DB4 is used as the mother wavelet of DWT to decompose and reconstruct the original data in two stages. Two high-frequency sequences and one low-frequency sequence can be obtained for training and prediction. For low-frequency sequences, the minimum-maximum normalization method is used to scale the input data to the [−1, 1] range before training the model. The predicted output values by the model will be readjusted to normal values. After several experiments, a GRU network with two hidden layers and each hidden layer with 256 units is used to predict low-frequency sequences. All neural network approaches were implemented using Tensorflow.

4.3. Results and Analysis.
In this section, a speed data set for urban road sections is used to evaluate the W-GRU-ARMA model. The validity of the model is verified by predicting the speed value of the road section in the next 10 minutes.

First of all, we applied the model to the experimental data of two road sections (A and B) with different traffic patterns. In this prediction experiment, it is important to choose the appropriate time steps for input. The best input is determined after performing prediction experiments on different inputs of time steps from 1 to 8 in the GRU network. Experimental results are shown in Figure 7.

Figure 7 shows that when the input time step increases from 1, the prediction errors decrease rapidly and the best prediction results on both A and B were obtained when the input time step is 2. The RMSE and MAPE values of prediction for road section A are 1.585 and 6.014, respectively, and those of road section B are 1.361 and 5.459, respectively. On the other hand, as the number of input step size continues to increase, the prediction errors do not decrease significantly. This means that the traffic status at a certain time has a strong correlation with the historical time closer
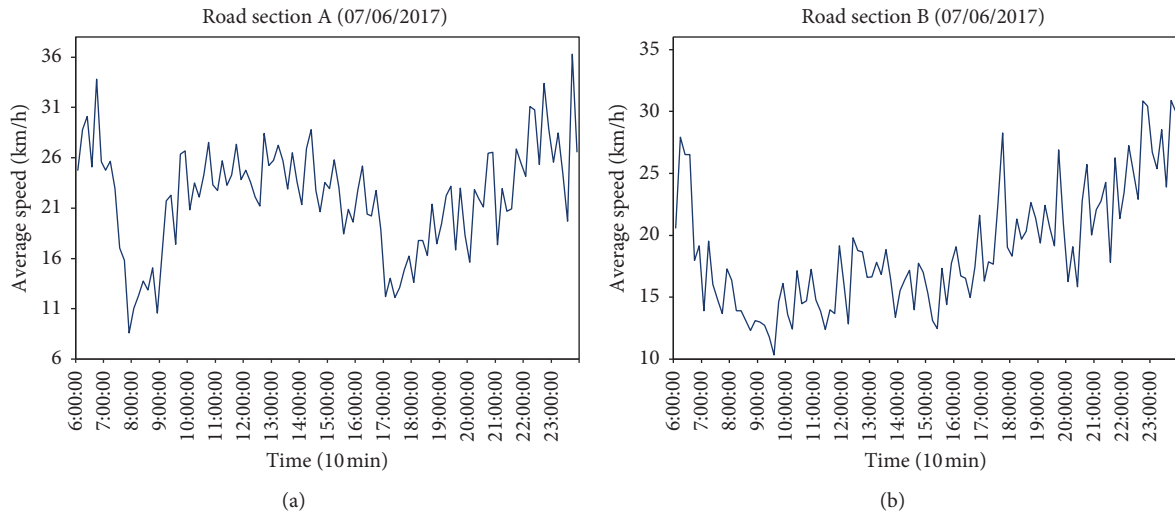
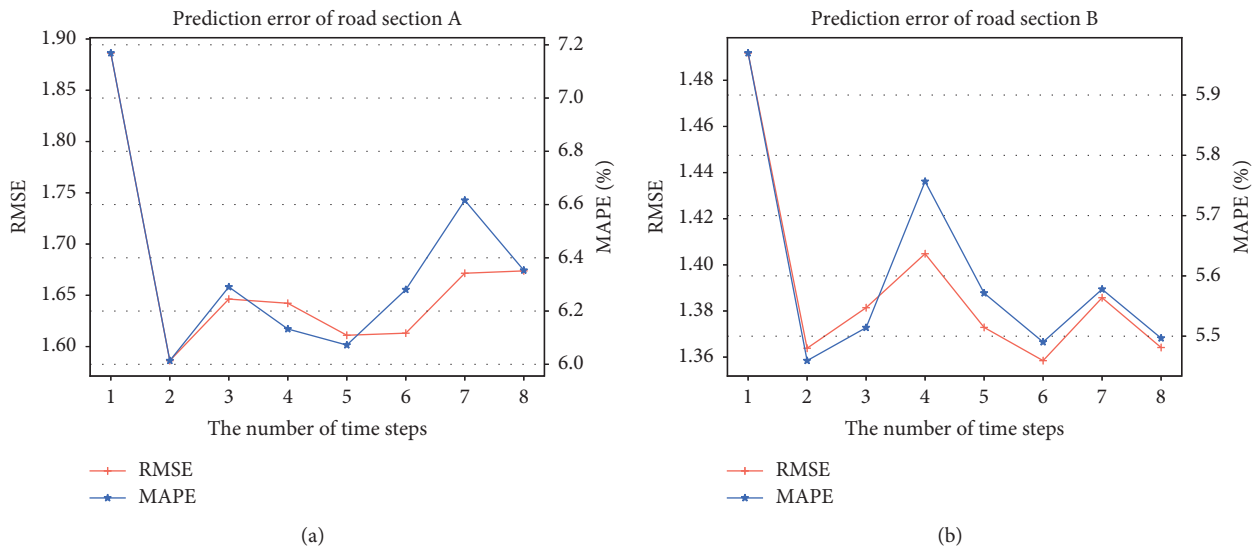Figure 6: The traffic patterns. (a) Road section A. (b) Road section B.

Figure 7: Prediction error of input at different time steps. (a) Road section A. (b) Road section B.

to it. In addition, the prediction errors of the two research sections have similar changes with the increase of the input time step, which may indicate that the traffic data generated by different road sections have similar short-term dependencies. Therefore, the time step of the input was set to 2 in subsequent experiments.

In this prediction experiment, a total of six days of traffic speeds were predicted, including weekdays and weekends. The actual traffic speed, the predicted traffic speed, and the associated residual for road sections A and B on July 26 (Wednesday) and July 30 (Sunday) are shown in Figure 8 to illustrate the prediction results of proposed model.

Figure 8 show that the actual value and the predicted value are in good agreement. The traffic pattern of road section A is more complicated on weekend, which makes the prediction difficult. The proposed model, therefore, has

slightly better prediction performance during the weekdays than at the weekends. On the other hand, the position of the standard red rectangles also reflects that the proposed model can better catch sudden changes of traffic speed. In addition, the prediction result drawn in Figures 8(c) and 8(d) reflects that the model performs better on road section B, possibly because similar traffic pattern B presents every day.

In order to find more details from the prediction results, the prediction errors of the two road sections per hour within 6 days are calculated and shown by box diagrams in Figure 9.

As shown in Figure 9, the prediction error of a road section at different times of the day is inequable. No matter for road section A or B, the prediction performance is more stable at the nonpeak hours, and there is a certain fluctuation of the performance at the morning and evening travel peaks,
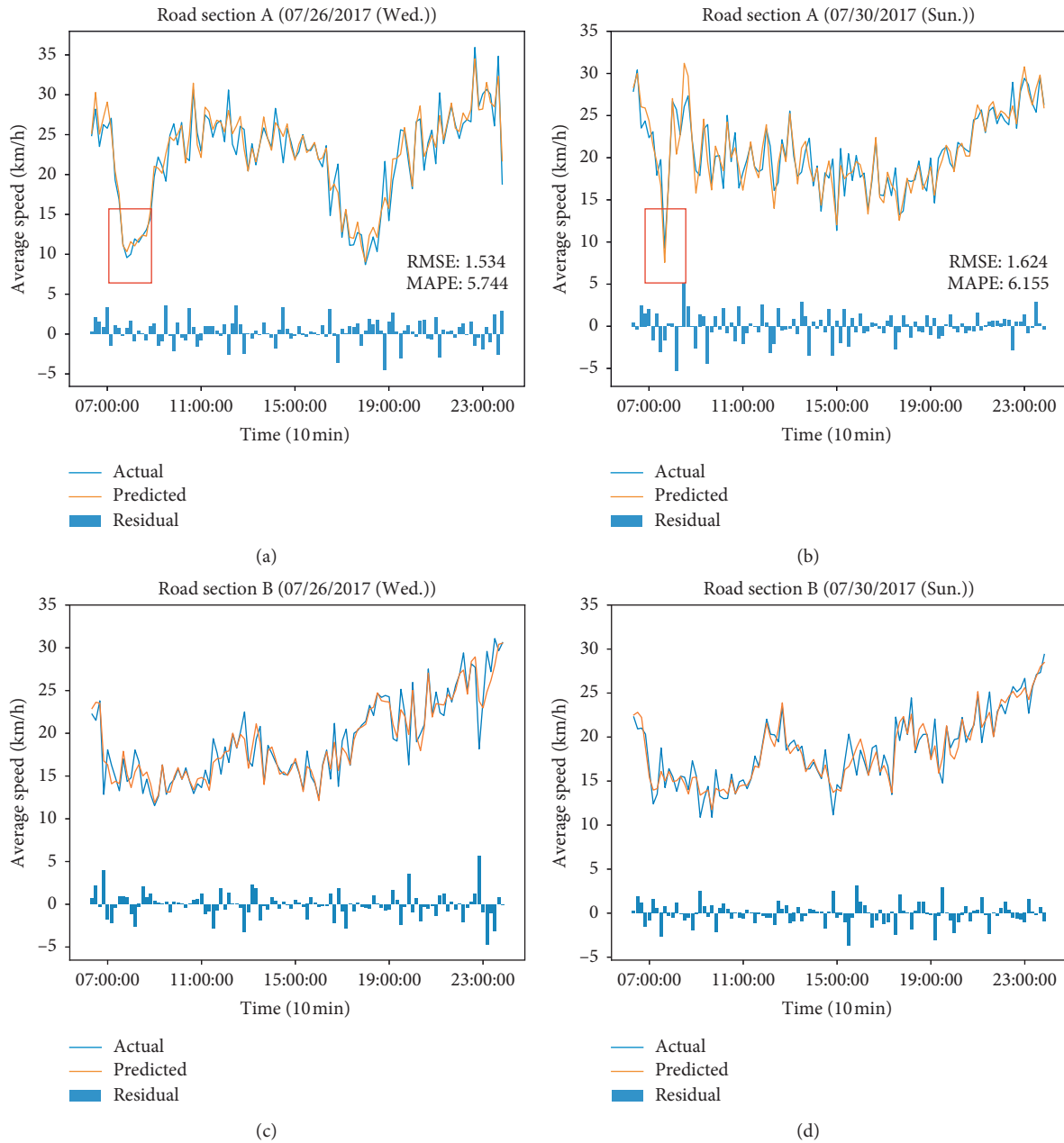
FIGURE 8: The predicted result of road sections A and B. (a) Road section A on 07/26/2017. (b) Road section A on 07/30/2017. (c) Road section B on 07/26/2017. (d) Road section B on 07/30/2017.

especially on road section A. For road section A, the prediction performance is better at noon and evening. Furthermore, the regularity of the prediction result for road section B in different time periods is not significant, because road section B is relatively busy all the time. But it is worth mentioning that the prediction error in different periods is always within a smaller range.

To verify the effectiveness of the proposed model, the prediction performance is compared with the GRU model, LSTM model, SAEs model, and ARIMA model. Tables1 and 2 demonstrate the prediction performance of different models.

From Tables 1 and 2, it can be seen that the prediction performance of the proposed model is better than ARIMA, GRU, SAEs, and LSTM models on the two experimental road sections, especially compared with ARIMA. This is owing to the ARIMA model which assumes that the traffic condition is a stationary process while this is not always true in reality. In general, the prediction error of each model on road section B is lower than that on road section A. In the prediction for road section B, the performance of LSTM is better than GRU. Based on this finding, this experiment was expanded on 12 road sections. The experimental results of the following four models are compared:
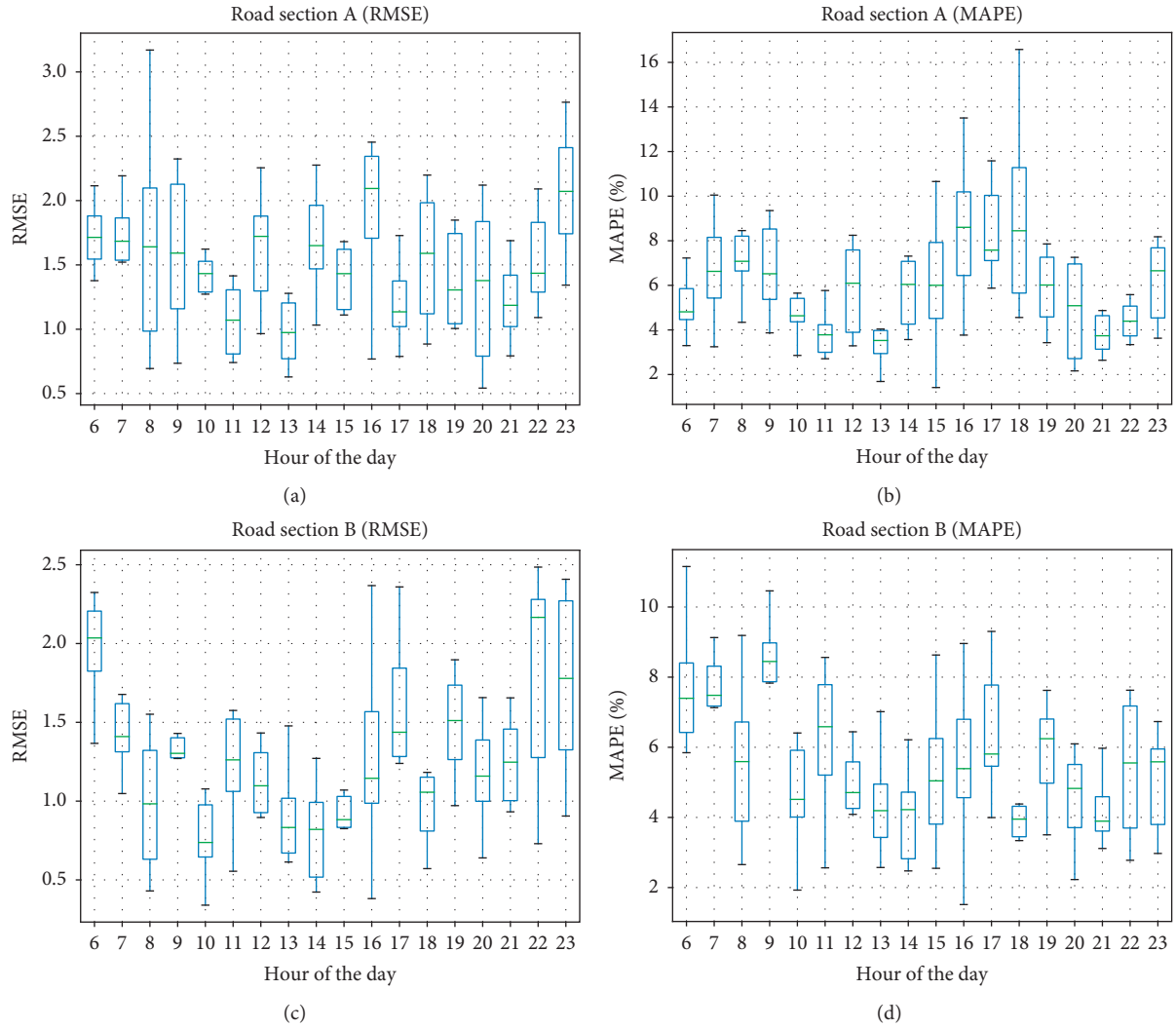
FIGURE 9: The error distribution per hour. (a) The RMSE of road section A, (b) the MAPE of road section A, (c) the RMSE of road section B, and (d) the MAPE of road section B.

TABLE 1: Performance comparison of different models on road section A.

Model	RMSE	MAPE (%)
ARIMA	2.104	9.593
SAEs	1.874	7.721
LSTM	1.946	8.132
GRU	1.860	7.656
W-GRU-ARMA	**1.585**	**6.014**

W + GRU + ARMA, W + LSTM + ARMA, GRU, and LSTM. The prediction results are presented in Figure 10.

As can be seen in Figure 10, W + GRU + ARMA achieves better results than W + LSTM + ARMA when the prediction results of the ARMA model are the same. This means that GRU networks, which are simpler in structure than LSTM, have the potential to predict the low-frequency sequence. Looking at it another way, the prediction effect of W + LSTM + ARMA model is not as good as that of the pure LSTM on some road sections. This is because pure LSTM can

better capture the sudden changes in traffic data [26]. When LSTM is used instead of GRU in our proposed model, it is possible that the smoothness of the data makes LSTM lose its advantage in some cases.

To verify the robustness of the proposed model, W-GRU-ARMA prediction was applied to the experimental data of 81 road sections in the study area. The prediction performance of different models is shown in Figure 11.

Prediction results shown in Figure 11 have proven that the proposed model is robust. The prediction error of the

TABLE 2: Performance comparison of different models on road section B.

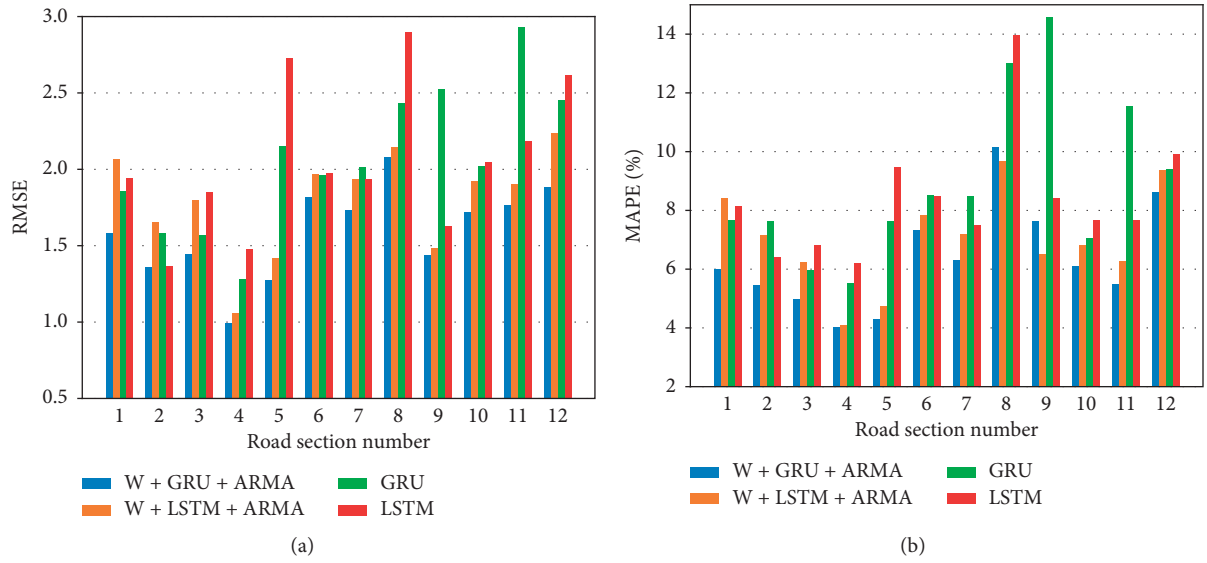| Model	RMSE	MAPE (%)
ARIMA	2.220	8.799
SAEs	1.376	6.077
LSTM	1.367	6.398
GRU	1.582	7.636
W-GRU-ARMA	**1.361**	**5.459**

(a)

(b)

FIGURE 10: Prediction performance of four models on 12 road sections. (a) RMSE and (b) MAPE.
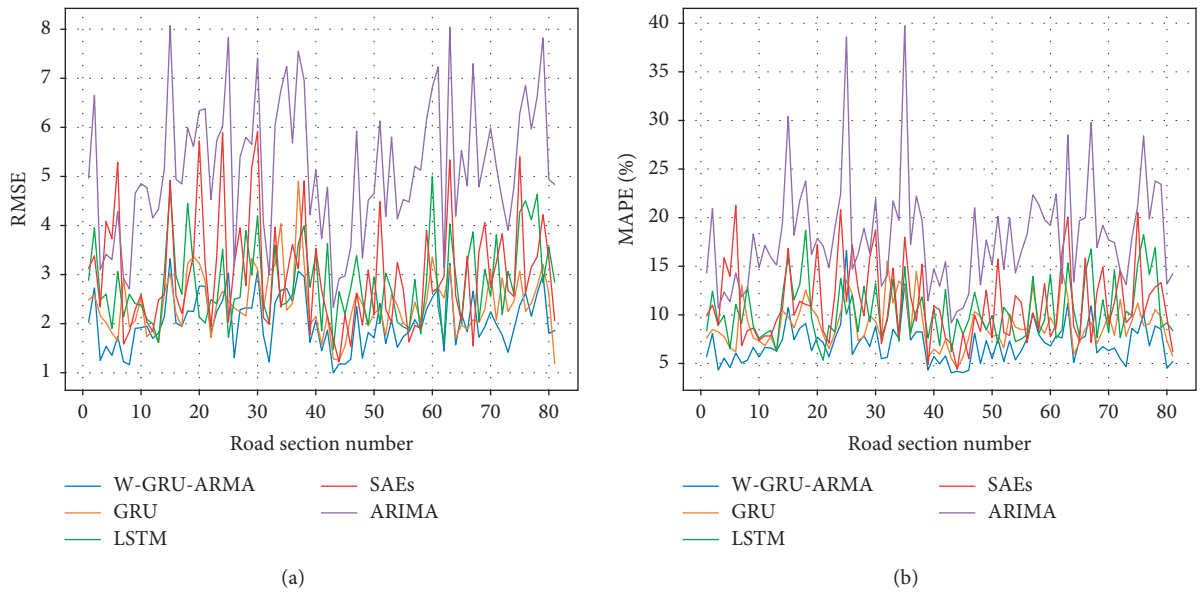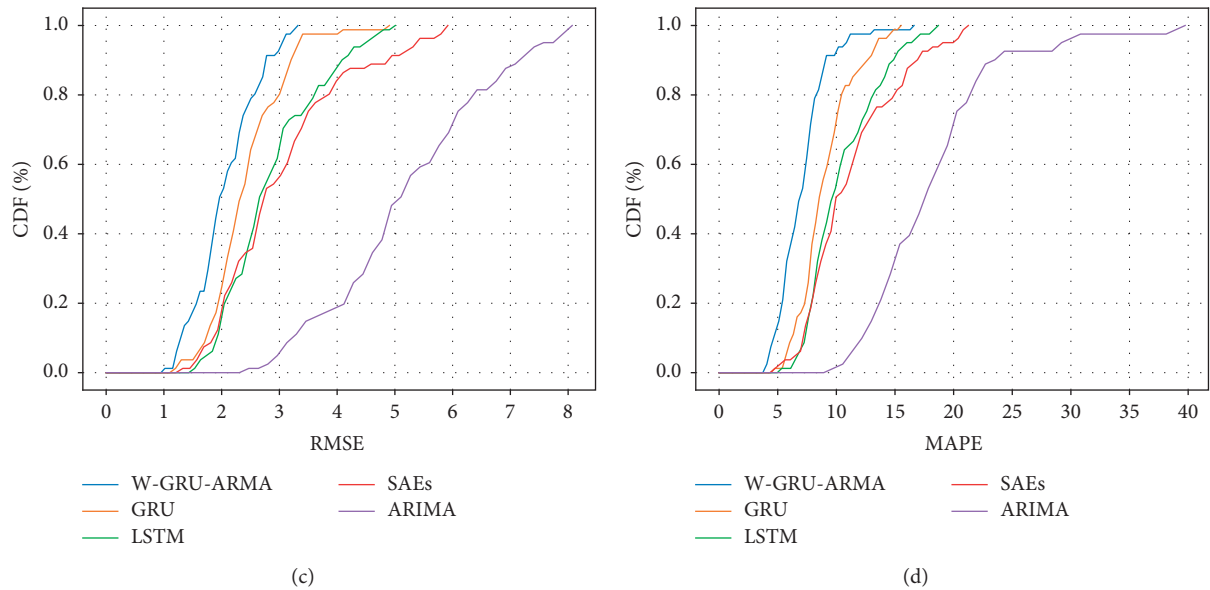

(a)

(b)

FIGURE 11: Continued.

FIGURE 11: Prediction performances comparison of different models. (a) The RMSE of 81 road sections, (b) the MAPE of 81 road sections, (c) the Cumulative Distribution Function of RMSE, and (d) the Cumulative Distribution Function of MAPE.

proposed model always fluctuates within a lower range, while the performance is better than other models on most road sections. On more than 80% road sections, according to Figure 11(c), the RMSE of prediction is below 2.5. The MAPE, as shown in Figure 11(d), is below 10% on more than 90% of the road sections. These results verify the effectiveness and superiority of our model for short-term traffic prediction on road sections.

5. Conclusions

This paper proposed a new combination model for short-term vehicle speed prediction based on the spatiotemporal correlation of traffic data and unstable random fluctuations. The model consists of three components, namely, WT, GRU network, and ARMA model. For one thing, the historical traffic speed data are decomposed into subsequence from high to low in terms of frequencies by WT. Furthermore, there are multiple low-frequency sequences that have stable trends and spatiotemporal relationships between one another. A GRU model is constructed so that it takes in these sequences and then predicts the speed value of next moment. At the same time, the ARMA model is used to predict high-frequency sequences with unstable randomness. This enables our proposed model to simultaneously fit the steady trend and randomness of traffic speed data. The prediction experiment using the real traffic speed data generated by floating vehicle was carried out. Experimental results showed the advantage of the proposed model compared with the previous models in terms of two performance measures: MAPE and RMSE.

We also obtained the following key content from the experimental results. When GRU networks and LSTM networks are used to predict low-frequency sequences, GRU networks perform better, which means that GRU networks

have great potential for smooth sequence prediction. In addition, although the traffic speed data has a strong time dependence, the prediction performance of the model does not increase with the input time step. It makes sense to choose the best input before conducting a prediction experiment. It is also found that the prediction error of the proposed model in different periods is slightly different, implying that effective analysis of traffic data at different time periods and the establishment of corresponding models may improve the accuracy of traffic predictions. Therefore, this will be the main content of our further research.

Data Availability

The taxi trajectory data used to support this study were provided by the Ningbo public transportation passenger transport administration and were made nonpublic.

Conflicts of Interest

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

Acknowledgments

Supplementary Materials

Supplementary files are the original data of the experimental part of the paper. These data are used to train and test the model to verify its validity and accuracy. File "road_A.xlsx" is the training data of road section A, and file "road_-A_test.xlsx" is the test data of road section A. File

"road_B.xlsx" is the training data of road section B, and file "road_B_test.xlsx" is the test data of road section B. Each file contains four valid fields: "time," "up," "mid," and "down," where "up" represents the speed data of the upstream section, "mid" represents the speed data of the study, and "down" represents the speed data of the downstream section. (*Supplementary Materials*)

References

[1] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we'e going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.

[2] X. Zhang, G. Xiong, L. Xiao et al., "A design of intelligent route guidance system based on shortest path algorithm," in *Proceedings of the 2015 IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI)*, pp. 12–17, IEEE, Hammamet, Tunisia, November 2015.

[3] U. Ryu, J. Wang, T. Kim, S. Kwak, and J. U, "Construction of traffic state vector using mutual information for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 55–71, 2018.

[4] J. Tang, L. Li, Z. Hu, and F. Liu, "Short-term traffic flow prediction considering spatio-temporal correlation: a hybrid model combing type-2 fuzzy C-means and artificial neural network," *IEEE Access*, vol. 7, 2019.

[5] A. Garcia-Ortiz, S. M. Amin, and J. R. Wootton, "Intelligent transportation systems—enabling technologies," *Mathematical and Computer Modelling*, vol. 22, no. 4–7, pp. 11–81, 1995.

[6] X. Ma, H. Yu, Y. Wang et al., "Large-scale transportation network congestion evolution prediction using deep learning theory," *PloS One*, vol. 10, no. 3, 2015.

[7] B. Yu, X. Song, F. Guan et al., "k-Nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *Journal of Transportation Engineering*, vol. 142, no. 6, Article ID 4016018, 2016.

[8] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.

[9] K. Cho, B. Van Merrinboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, https://arxiv.org/abs/1406.1078.

[10] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proceedings of the International Conference on Machine Learning*, International Machine Learning Society (IMLS), Lille, France, pp. 2342–2350, July 2015.

[11] G. Dai, C. Ma, and X. Xu, "Short-term traffic flow prediction method for urban road sections based on space-time analysis and GRU," *IEEE Access*, vol. 7, pp. 143025–143035, 2019.

[12] Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, "Short-term prediction of lane-level traffic speeds: a fusion deep learning model," *Transportation Research Part C: Emerging Technologies*, vol. 106, pp. 1–16, 2019.

[13] N. Zhang, X. Guan, J. Cao, X. Wang, and H. Wu, "Wavelet-HST: a wavelet-based higher-order spatio-temporal framework for urban traffic speed prediction," *IEEE Access*, vol. 7, pp. 118446–118458, 2019.

[14] Y. Gao and F. Chen, "Wavelet analysis-based npr prediction of short-term traffic flow," *Journal of University of Science & Technology of China*, vol. 38, no. 12, pp. 1427–1431, 2008.

[15] L. C. Ma, W. L. Xu, and D. X. Liu, "Prediction model of traffic flow along typical roads in city urban district based on wavelet transform," *Control and Decision*, vol. 26, no. 5, pp. 789–793, 2011.

[16] C. Yanchong, H. Darong, and Z. Ling, "A short-term traffic flow prediction method based on wavelet analysis and neural network," in *Proceedings of the 2016 Chinese Control and Decision Conference (CCDC)*, pp. 7030–1034, IEEE, Yinchuan, China, May 2016.

[17] K. I. Wong and Y. C. Hsieh, "Short-term traffic flow forecasting for urban roads using space-time ARIMA," in *Proceedings of the Transportation and Urban Sustainability*, Hong Kong Society for Transportation Studies, Hong Kong China, pp. 583-584, December 2010.

[18] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, p. 21, 2015.

[19] L. L. Ojeda, A. Y. Kibangou, and C. C. De Wit, "Adaptive Kalman filtering for multi-step ahead traffic flow prediction," in *Proceedings of the 2013 American Control Conference*, pp. 4724–4729, IEEE, Washington, DC, USA, June 2013.

[20] S. V. Kumar, "Traffic flow prediction using kalman filtering technique," *Procedia Engineering*, vol. 187, pp. 582–587, 2017.

[21] M. T. Asif, J. Dauwels, C. Y. Goh et al., "Spatiotemporal patterns in large-scale traffic speed prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 794–804, 2013.

[22] B. Yao, C. Chen, Q. Cao et al., "Short-term traffic speed prediction for an urban corridor," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 2, pp. 154–169, 2017.

[23] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.

[24] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, 2016.

[25] C. Gang, W. Shouhui, and X. Xiaobo, "Review of spatio-temporal models for short-term traffic forecasting," in *Proceedings of the 2016 IEEE international conference on intelligent transportation engineering (ICITE)*, pp. 8–12, IEEE, Singapore, August 2016.

[26] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, Neural Information Processing Systems Foundation, Inc. (NIPS), December 2012, pp. 1097–1105, Lake Tahoe, NV, USA.

[28] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[29] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, AAAI, Stockholm, Sweden, July 2018.

[30] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," in

Proceedings of the International Conference on Learning Representations (ICLR'18), International Conference on Learning Representations, Vancouver, BC, Canada, May 2018.

[31] Y. Lv, Y. Duan, W. Kang et al., "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.

[32] Y. Jia, J. Wu, and Y. Du, "Traffic speed prediction using deep learning method," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1217–1222, Rio de Janeiro, Brazil, IEEE, November 2016.

[33] G. Fusco, C. Colombaroni, and N. Isaenko, "Short-term speed predictions exploiting big data on large urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 73, pp. 183–201, 2016.

[34] F. Lin, Y. Xu, Y. Yang et al., "A spatial-temporal hybrid model for short-term traffic prediction," *Mathematical Problems in Engineering*, vol. 2019, Article ID 4858546, 12 pages, 2019.

[35] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018.

[36] L. Ouyang, F. Zhu, G. Xiong et al., "Short-term traffic flow forecasting based on wavelet transform and neural network," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, IEEE, Yokohama, Japan, October 2017.

[37] C. Parameswariah and M. Cox, "Frequency characteristics of wavelets," *IEEE Transactions on Power Delivery*, vol. 17, no. 3, pp. 800–804, 2002.

[38] A. Jensen and A. la Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer Science & Business Media, Berlin, Germany, 2001.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] L. Kuan, Z. Yan, W. Xin et al., "Short-term electricity load forecasting method based on multilayered self-normalizing GRU network," in *Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–5, IEEE, Beijing, China, November 2017.

[41] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic timeseries data by using box-jenkins techniques," *Transportation Research Record*, vol. 773, no. 722, pp. 1–9, 1979.

[42] Z. He, G. Qi, L. Lu, and Y. Chen, "Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 320–339, 2019.

[43] H. Yu, N. Ji, Y. Ren, and C. Yang, "A special event-based K-nearest neighbor model for short-term traffic state prediction," *IEEE Access*, vol. 7, pp. 81717–81729, 2019.

[44] Y. Yue, H. X. Zou, and Q. Q. Li, "Urban road travel speed estimation based on low sampling floating car data," in *Proceedings of the ICCTP 2009: Critical Issues In Transportation Systems Planning, Development, and Management*, pp. 1–7, American Society of Civil Engineers, Reston, VA, USA, July 2009.

[45] J. S. Greenfeld, "Matching GPS observations to locations on a digital map," in *Proceedings of the 81th Annual Meeting of the Transportation Research Board*, Transportation Research Board (TRB), Washington, DC, USA, pp. 164–173, April 2002.

[46] Z. He, Y. Lv, L. Lu, and W. Guan, "Constructing spatio-temporal speed contour diagrams: using rectangular or nonrectangular parallelogram cells?" *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 44–60, 2019.