IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 14, 2019, accepted September 8, 2019, date of publication September 13, 2019, date of current version October 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941280

Short-Term Traffic Flow Prediction Method for Urban Road Sections Based on Space–Time Analysis and GRU

GUOWEN DAI[1], CHANGXI MA[1], AND XUECAI XU[2]

[1]School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China
[2]School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding authors: Changxi Ma (machangxi@mail.lzjtu.cn) and Xuecai Xu (xuecai_xu@hust.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 71861023, in part by the Program of Humanities and Social Science of Education Ministry of China under Grant 18YJC630118, in part by the Fundamental Research Fund for the Central Universities under Grant HUST: 2018KFYYXJJ001, and in part by the Foundation of a Hundred Youth Talents Training Program of Lanzhou Jiaotong University.

ABSTRACT Accurate short-term traffic forecasts help people choose transportation and travel time. Through the query data, many models for traffic flow prediction have neglected the temporal and spatial correlation of traffic flow, so that the prediction accuracy is limited by the accuracy of traffic data. This paper proposed a short-term traffic flow prediction model that combined the spatio-temporal analysis with a Gated Recurrent Unit (GRU). In the proposed prediction model, firstly, time correlation analysis and spatial correlation analysis were performed on the collected traffic flow data, and then the spatiotemporal feature selection algorithm was employed to define the optimal input time interval and spatial data volume. At the same time, the selected traffic flow data were extracted from the actual traffic flow data and converted into a two-dimensional matrix with spatio-temporal traffic flow information. The GRU was used to process the spatio-temporal feature information of the internal traffic flow of the matrix to achieve the purpose of prediction. Finally, the prediction results obtained by the proposed model were compared with the actual traffic flow data to verify the effectiveness of the model. The model proposed in this paper was compared with the convolutional neural network (CNN) model and the GRU model, and the results show that the proposed method outperforms both in accuracy and stability.

INDEX TERMS Gated recurrent unit, spatio-temporal analysis, short-term traffic flow prediction, traffic engineering, urban road section.

I. INTRODUCTION

Accurate and real-time short-term traffic flow prediction can correctly and reasonably infer traffic conditions in the future according to the current traffic network change rules, provide convenient route planning for travelers, alleviate traffic congestion and reduce air pollution. Therefore, the study of short-term traffic flow has important practical significance and application prospects. In the past few decades, many scholars have been working on short-term traffic flow prediction. The existing methods mainly include time series model, Kalman filter model, support vector regression model and hybrid combination model [1]. In recent years, with the improvement of computing power of computers, artificial intelligence and deep learning have been rapidly developed in transportation field. Common deep learning algorithms include deep residual networks, cyclic neural networks, and convolutional neural networks, in which self-learning capability highlights the features of short-term traffic flow prediction. Williams *et al.* [2] analyzed urban expressway traffic flow data and used seasonal Autoregressive Integrated Moving Average Model (ARIMA) prediction models to predict urban expressway traffic flow. Wang and Gu [3] proposed a traffic flow velocity prediction model based on single hidden layer convolutional neural network combined with error feedback. Kuremoto *et al.* [4] studied how deep belief networks based on restricted Boltzmann machines can be applied to short-term traffic flow prediction. The convolutional neural network CNN performs very well when processing two-dimensional data such as images. The CNN is usually formed

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

by stacking convolutional layers and sub-sampling layers, that is, pooling layers. The last layer can select different output layers according to different tasks [5]. In the forward calculation, the convolutional layer can simultaneously use multiple convolution kernel parameters to generate a plurality of feature maps reflecting the data distribution at one time. The pooling layer can further reduce the feature map dimensions to reduce the redundant features and reduce the computational cost of the model. CNN training also uses error back propagation algorithm and parameter learning method based on delta rule [5]. Yu *et al.* [6] proposed a data grouping model based on convolutional neural network, which consists of a two-part structure. The Continuous Bag-Of-Words Model (CBOW) is used to find similar spatiotemporal relationships, and the CNN network is used for short-term traffic flow prediction. Zheng *et al.* [7] performed multivariate time series analysis. CNN performs better in short-term traffic flow prediction, but there are several problems in the training process: too much reliance on GPU, and vice versa on the CPU is much lower than the GPU; when the data is insufficient, the predicted result is quite different from the actual result; the presence of the pooling layer leads to the loss of many very valuable information, and also ignores the relationship between the whole and the part.

In recent years, relevant scholars have begun to study the prediction methods based on the interaction and organic connection of multiple road segments in the road network [8]. Cheng *et al.* [9] quantitatively analyzed the temporal and spatial correlation of traffic flow, and established a hybrid process neural network model to achieve short-term traffic flow prediction. Min *et al.* [10] proposed an Generalized Space-Time Autoregressive Integrated Moving Average (GSTARIMA) model by mining the spatio-temporal characteristics of urban traffic flow. Gao *et al.* [11] fully exploited the weekly similarity and spatial correlation of traffic flow time series, combined with the characteristics of large-scale data fusion of RBF neural network, proposed short-term traffic based on traffic flow spatiotemporal correlation and radial basis neural network. Flow prediction method. Min *et al.* [12] used the traffic flow time series on several sections in the road network as the research object, and established the GARCH prediction model. Combined with the example analysis, the prediction result is better than the prediction result based on the single-section traffic flow data alone. Cui [13] proposed a parallel online time-space composite prediction method based on traffic flow information on multiple sections, and established two online adaptive prediction models. Dong and Shao [14] combined with the spatio-temporal characteristics of traffic flow, taking traffic flow under free-flow state as the research object, and establishing a prediction model based on state space to realize traffic flow prediction of multiple sections of road network. Qiu *et al.* [15] proposed time-series data prediction method and spatial regression estimation method based on time-space data flow data respectively, and used the least squares dynamic weighted fusion algorithm to fuse the prediction results of two traffic flow prediction methods to obtain more Precise predictions. Li *et al.* [16] comprehensively studied the time and space factors of traffic flow, combined with support vector machine to establish online traffic flow prediction model, and used grid search method to optimize model parameters. At present, the studies on the prediction method based on the temporal and spatial correlation characteristics of traffic flow is still in its infancy, which leaves some gap. Although it can utilize the spatio-temporal characteristics of traffic flow to predict short-term traffic flow, its accuracy is limited and its dependence on data is large. When the data is insufficient, the prediction accuracy will be greatly reduced.

The Recurrent Neural Network (RNN) is an artificial neural network with a hidden layer node oriented connection and closed loop. Unlike other feed-forward neural networks, RNN can use internal memory cells to handle timing inputs of any length. For the standard RNN architecture, the effect of a given input on the hidden layer unit and the network output will disappear or explode with the periodic connection of the RNN network, which is the gradient disappearance problem [17]. In order to solve this problem, LSTM (Long-Short Term Memory Neural Network), a specially designed RNN network, is designed to give memory units the ability to determine when to remember or when to forget certain information, thus solving the time series. When the problem is solved, the optimal time lag can be automatically determined, showing long-term memory ability of historical data or samples [16]. Ma *et al.* [18] proposed a LSTM model for urban travel time prediction. The optimal input length and hidden layer nodes were determined by controlling variables. The implementation of four LSTM is compared with four models, such as time series ARIMA model and BPNN. The results show that the performance of LSTM considering spatial features is the best. Ma et al. proposed a method for training LSTM cyclic neural networks using remote microwave sensor data to predict traffic speed and achieved good results. Tian and Pan [19] discussed the performance of the LSTM cyclic neural network for predicting short-term traffic flow and compared it with several other commonly used models. Fu *et al.* [20] used the LSTM and GRU neural network models to predict short-term traffic flow and confirmed that the cyclic neural network model is superior to ARIMA. The LSTM controls the state of the transmission through the gated state, remembering that it takes a long time to memorize, forgetting the unimportant information; unlike the normal RNN, there is only one way of memory overlay. It is especially useful for many tasks that require ''long-term memory''. However, because of the introduction of various contents, and more parameters, it also makes the training more difficult.

By taking the advantages of spatiotemporal analysis and long-term and short-term memory networks, this paper designs a GRU short-term traffic flow prediction model based on spatio-temporal analysis. The GRU effect is equivalent to LSTM but has fewer parameters, which is suitable for constructing a large training model. The time-space analysis of the traffic flow data that needs to be trained and tested to select the appropriate input data has a positive effect on improving the accuracy of the prediction model.

In summary, the emergence of deep learning has brought tremendous opportunities for the development of traffic flow prediction. But overall, the application of deep learning methods is still in the initial exploration stage. Based on the summary of previous studies, this study integrates space-time analysis with gated loop unit neural network algorithm, applies in the field of short-term traffic flow prediction, and conducts in-depth research and discussion.

II. SPATIO-TEMPORAL CORRELATION ANALYSIS AND DATA REPRESENTATION OF TRAFFIC FLOW DATA

When considering predictive problems, it is necessary to determine the data characteristics associated with traffic prediction and how to select and use the data. From a time domain perspective, traffic flow data is a sequence that changes over time, related to its previous time state and the latter time state. In the spatial domain, the traffic flow states on adjacent road sections interact, and the traffic flow state of the upstream section or the downstream section affects the traffic flow state of the research section to varying degrees. Therefore, changes in traffic flow have certain temporal and spatial characteristics.

Traffic data can contain a lot of information, but choosing information related to research questions is critical. In this study, by analyzing the spatiotemporal correlation in traffic flow data, we can sort the data according to the correlation between the data and the current forecasting problem, which is convenient for subsequent selection. After selected the reasonable input data size to construct the prediction model, the data is ranked according to the relevance of the data, the prediction performance of the model is verified, and the spatio-temporal feature selection algorithm (STFSA) algorithm is used to select the optimal input data size. By analyzing the traffic flow data, the prediction model constructs a matrix using the selected data as known information, and predicts the future traffic flow based on the matrix information. This study uses the Pearson correlation coefficient as a criterion for evaluating correlation strength:

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

where x_i and y_i represent two traffic flow sequences, and the magnitude of the Pearson correlation coefficient value indicates the correlation between the two time series. The closer the coefficient values are, the higher the correlation between the two sequences.

A. TIME CORRELATION ANALYSIS OF TRAFFIC FLOW DATA

The traffic state at the current time of a certain section will be affected by the state at the previous moment, and will also affect the state at the next moment. In addition, traffic flow is a continuous process of stochastic change, which essentially reflects the overall behavior of social activities. Its continuity and periodicity reflected by social activities also indicate that traffic flow also interacts with each other in the time dimension. An absolute correlation of the correlation coefficient greater than 0.68 is a strong correlation. Therefore, in order to weigh the size of the input data and the prediction accuracy of the model, the time when the autocorrelation coefficient is greater than 0.68 is selected as the time lag of initializing the input data. The specific algorithm of STFSA will be given in Section 2.4.

B. SPATIAL CORRELATION ANALYSIS OF TRAFFIC FLOW DATA

For the urban road network, each roadway segment is not isolated, but interconnected. The roadway network connectivity indicates that the traffic flow attached to the roadway segment interacts in the spatial dimension. The traffic flow is attached to the roadway, and the traffic flow of adjacent roadway sections will also affect each other. When the roadway is unblocked, the traffic flow in the upstream section will have an impact on the traffic state of the downstream section. At the same time, when the downstream section is in a congested state, it will also affect the travel behavior of the upstream section, thus affecting the upstream traffic state. Due to the strong correlation between road traffic monitoring points, in the road traffic analysis, excessive traffic data input complicates the traffic prediction model and increases the parameters of the neural network. In order to simplify the prediction model and effectively use the flow data while ensuring the prediction accuracy, the initial selection rule of the data in the spatial dimension of the simulation data is to select the data of the six monitoring points with the highest correlation with the predicted points. In addition, the optimal input spatial data for the model using the STFSA algorithm is redefined by the algorithm based on specific prediction problems. The specific algorithm of STFSA will be given in Section 2.4.

C. MATRIX CONVERSION OF TRAFFIC FLOW DATA

According to the analysis of the temporal and spatial correlation of traffic flow data, the traffic flow data are extracted according to the time dimension and the spatial dimension to construct a space-time correlation matrix with traffic flow information, which is defined according to the following rules:

$$X_W = \begin{bmatrix} w_{1,t-(Q-1)} & w_{1,t-(Q-2)} & \cdots & w_{1,t} \\ w_{2,t-(Q-1)} & w_{2,t-(Q-2)} & \cdots & w_{2,t} \\ \vdots & \vdots & \vdots & \vdots \\ w_{P,t-(Q-1)} & w_{P,t-(Q-2)} & \cdots & w_{P,t} \end{bmatrix} \quad (2)$$

In this matrix, Q represents the time lag of the input data, and P represents the number of detection points selected by the spatial correlation analysis. The constructed spatio-temporal traffic data matrix is considered as the input data of the model, and the traffic flow at the predicted time of the predicted point is used as the output of the model.

D. SPACE-TIME FEATURE SELECTION ALGORITHM

Feature selection methods fall into three broad categories: filter, wrapper, and embedding.

Filter: Firstly select the feature of the dataset, and then train the learner. The feature selection process has nothing to do with the subsequent learning process. The famous one is the relieve (relevant features) method, which designs a "correlated statistic" to measure the importance of the feature. The specific approach is to find the nearest neighbor sample x_j of the same classification for each training sample x_i, and the nearest neighbor sample x_k which is not a classification. If diff(x_i, x_j)t represents the difference between x_i and x_j on the attribute t, then the correlation statistic calculates: the difference between the square of diff(x_i, x_k) and the square of diff(x_i, x_j) at all The average on the sample.

Embedding: Embedded feature selection combines the feature selection process with the learner training process.; that is to say, the feature selection is automated in the learner training process.

Wrapper: The wrap feature selection directly employs the model performance that will ultimately be used as the evaluation criteria for the feature subset. Therefore, it is better than the filter selection, but because the learner is trained multiple times during the feature selection process, the overhead is usually much larger than the filter type, and the LVW method is more typical. It uses a random strategy for subset search under the Las Vegas Method framework, and uses the error of the final classifier as a feature subset evaluation criterion.

The purpose of the wrapped feature selection is to select a subset of features that are "tailor-made" for a given learner that is most beneficial to its performance. Since the wrap feature selection method is optimized directly for a given learner, the wrap feature selection is better than the filter feature selection from the final learner performance. On the other hand, the learner needs to be trained multiple times during the feature selection process, so the computational overhead of the wrapped feature selection is usually much larger than the filtered feature selection.

The LVW algorithm is simple and straightforward. As the number of samples increases, the probability of getting the correct result gradually increases. If the correct result has been found during the random sampling process, the method can discriminate and report.

In this study, a STFSA based on the wrapped feature selection method (LVW) [21] was chosen.

As shown in equation (2), it is assumed that the combination of different time-delay Q of the input data and the number of different spatial data collection points P constitutes an input space I, and each element R in the input space is a specific input delay Q* Consists of the number of data collection points P*. STFSA can be expressed by the following formula:

$$R^* = \arg\min((f(R) - y) + \alpha\,|d|), \quad R \in I \tag{3}$$

STFSA is an algorithm for finding network inputs that exhibit the highest prediction accuracy on a validated data set. The addition of item $\alpha\,|d|$ is a compromise between network complexity and prediction accuracy. The specific calculation process is as follows:

III. CONSTRUCTION OF SHORT-TERM TRAFFIC FLOW PREDICTION GRU MODEL

Whether it is a convolutional neural network or an artificial neural network, the premise is that the elements are independent of each other, and the inputs and outputs are independent, e.g. cats and dogs. But in the real world, many elements are connected to each other. Some tasks need to be able to better process the sequence information, that is to say, the previous input and the subsequent input are related, for example, changes in stocks over time; The RNN network introduces directional loops, handling the contextual correlation between those inputs. The RNN network remembers the previous information and uses the previous information to influence the output of the following nodes.

LSTM is the Long Short Memory Network, which is actually a variant of RNN. It can learn long-term dependence information. The state calculation formula of RNN according to the chain derivation method will cause the gradient to become a multiplication form, and the sigmoid less than 1 will make the multiplication much smaller. In order to solve this problem, scholars have adopted an accumulated form, and the derivatives are also accumulated to avoid the disappearance of the gradient. LSTM uses the cumulative form, but its implementation is more complicated. After the RNN model is expanded, the hidden layers are interconnected at multiple times, and all of the cyclic neural networks have a duplicate network module. RNN's repetitive network module is simple, with only one tanh layer. The structure of the LSTM repeating network module is much more complicated. It Includes three gates, namely the forget gate, the input gate and the output gate. Each door is responsible for the difference. The forget gate is responsible for determining how many unit states of the previous moment are retained to the current state of the unit. The input gate is responsible for determining how many current moments of input are retained to the current state of the unit, while the output gate is for determining how much output the unit state has at the current time.

GRU, a type of RNN, is a very good variant of the LSTM network. It not only inherits the advantages of the RNN model to automatically learn features and effectively model long-distance dependent information, but also keeps the RNN prediction performance and has a significant increase in speed, while maintaining the effect and making the structure simpler. GRU is simpler and more effective than the LSTM network. Three gate functions are introduced in LSTM: input gate, forget gate, and output gate. There are only two gates in the GRU model: the update gate and the reset gate.

Algorithm 1 Spatio-Temporal Feature Selection Algorithm (STFSA)

Require: d: Enter the amount of spatio-temporal data
Require: $R_{initial}$: Initialized correlation time length and space related data
Require: E: The verification set prediction error of the input data evaluated using MAPE.
Require: α: Regularization.
Input: input data set I; input one element in the data set R; prediction algorithm f; algorithm stop control parameters T, P, Q
Output: Best input data R*
process:
1: E$\leftarrow$ Validation (f($R_{initial}$))
2: d$\leftarrow$ $|R_{initial}|$;
3: R$*$ $\leftarrow$ $R_{initial}$;
4: t $\leftarrow$0;
5: while t<T do
6: p$\leftarrow$0;
7: while p<P do
8: Changing the time lag does not repeat the search for R' in data set I;
9: $d' < | R' |$;
10: $E' \leftarrow$ Validation(f ($R_{initial}$));
11: if (E +a|d$'$ | $<$ E +a|d|) then
12: p$\leftarrow$ 0;
13: E$\leftarrow$ E$'$
14: d$\leftarrow$d$'$;
15: R$*$ $\leftarrow$R$'$;
16: else
17: p$\leftarrow$p+1;
18: end if
19: end while
20: q$\leftarrow$0;
21: while q<Q do
22: Change the number of links added to the calculation. In data set I, do not repeat the search for R';
23: $d' < |R'|$;
24: $E' \leftarrow$ Validation (f (R$'$));
25: if (E$'$ +a|d$'$ | $<$ E+a|d|) then
26: q$\leftarrow$0;
27: E$\leftarrow$E$'$;
28: d$\leftarrow$d $'$;
29: R$*$ $\leftarrow$R $'$;
30: else
31: q$\leftarrow$q+1;
32: end if
33: end while
34: t = t+1;
35: end while
36: Return R*

Compared with LSTM, GRU has one less gate, which reduces some matrix multiplication. Although GRU is closer to LSTM, it saves a lot of time in training. In text class processing, it is more recommended to use GRU to train than LSTM. What is special about these two gating mechanisms is that they are capable of storing information in long-term sequences that are not cleared over time or removed because they are irrelevant to prediction [22].The GRU model is described below.

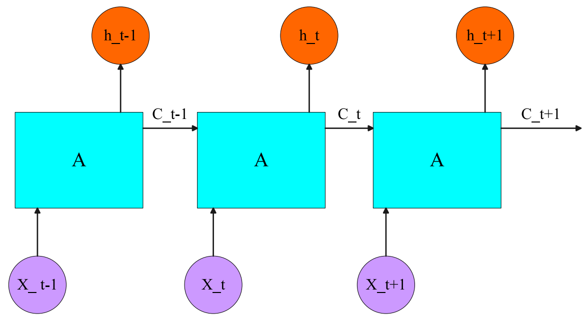


FIGURE 1. GRU neural network structure.

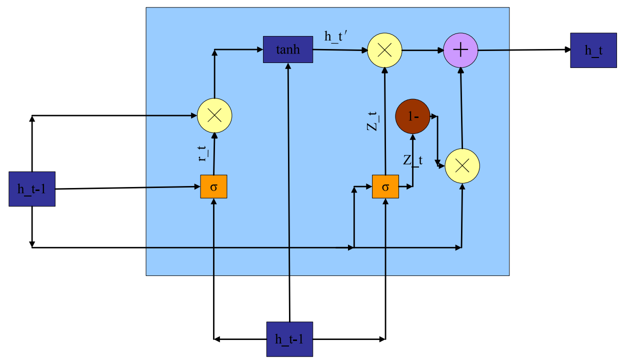


FIGURE 2. GRU neural network neuron structure.

Forget gate, input gate, and output gate are employed in the LSTM neural network. These "gate" are designed to control the state of the cells by eliminating or enhancing the input information into the nerve cell unit. The GRU neural network has improved the "gate" design, integrating the forgetting gate and the input gate in the LSTM into an Update Gate. That is, the cell structure originally composed of three "gates" is optimized into two "gates" into the cellular structure. At the same time, the Cell State has been merged and other improvements have been made [23]–[29]. The overall GRU neural network structure is shown in Figure 1.

As the traditional recurrent neural network, the GRU neural network is also a chain model composed of multiple neural unit modules. In the traditional recurrent neural network, neuron A may be just the simplest tanh function or ReLU function. But in GRU, neuron A is a more complex threshold structure. The detailed structure of a single neuron is shown in Figure 2.

Neurons of the GRU neural network are expressed using mathematical formulas as:

$$Z\_t = \sigma(W\_z \cdot [h\_t - 1, X\_t]) \tag{4}$$
$$r\_t = \sigma(W\_r \cdot [h\_t - 1, X\_t]) \tag{5}$$

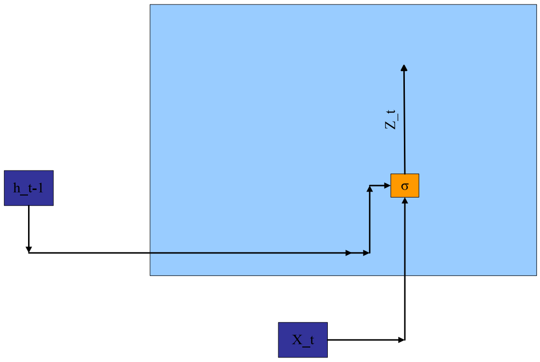


FIGURE 3. Update gate model.

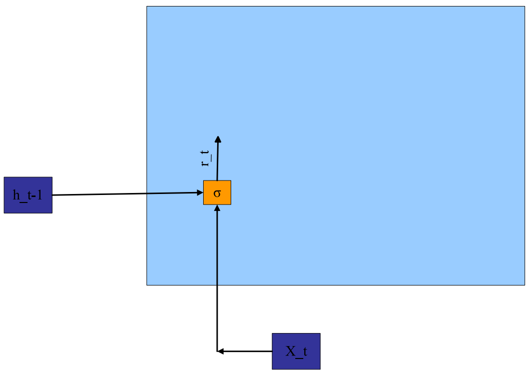


FIGURE 4. Reset gate model.

$$h\_t' = \tanh(W \cdot [r\_t * h\_t - 1, X\_t]) \tag{6}$$

$$h\_t = ([1 - Z\_t] * h\_t - 1) + Z\_t * h\_t' \tag{7}$$

The individual nerve cells in the GRU neural network will now be decomposed in detail. First, as shown in Figure 3, is the representation of the update gate in the GRU neural network. According to Z_t in Fig. 3 and formula (4), an update gate is shown, h_t-1 represents the output of the previous neuron, X_t represents the input of the current neuron, W_z represents the weight of the update gate, and σ represents the sigmoid function. The update gate Z_t is obtained by adding the output h_t-1 of the previous neuron and the input X_t of the current neuron, multiplying by updating the gate weight W_z, and then using the sigmoid function. For the update gate Z_t, when the value is larger, the more information indicating the current neuron to be retained, and the less information to be retained by the previous neuron.

Then the Reset Gate model in the GRU neural network is shown in Figure 4. According to r_t in Fig. 4 and formula (5), a reset gate is shown, h_t-1 represents the output of the previous neuron, X_t represents the input of the current neuron, W_r represents the weight of the reset gate, and σ represents Sigmoid function. The reset gate r_t is obtained by adding the output h_t-1 of the previous neuron to the input X_t of the current neuron and multiplying by the reset gate weight W_r, and then using the sigmoid function. For the reset gate r_t, when the value of the equation is 0, it means that the information from the previous neuron is discarded.

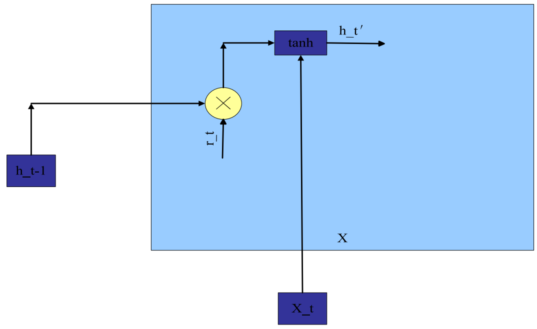


FIGURE 5. Pending output value model.

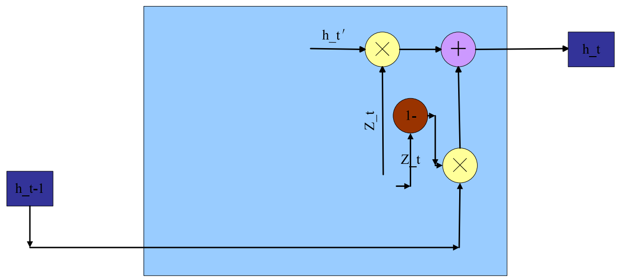


FIGURE 6. Output value model.

That is, as long as the input of the current neuron is input, this allows the current neuron to discard some of the useless information of the previous neuron.

Second, the pending output value model in the GRU neural network is shown in Fig. 5. According to Figure 5 and h_t' in Equation (6), the output value to be determined in this neuron is represented, and r_t represents the reset gate, h_t-1 represents the output of the previous neuron, X_t represents the input of the current neuron, W represents the weight of the updated gate, and tanh represents the hyperbolic tangent function. The pending output value h_t' is obtained by multiplying the output of the previous neuron, h_t-1, and the reset gate r_t, with the input X_t of the current neuron, multiplying by the weight, and finally using the hyperbolic tangent function.

The output value model in the final GRU neural network is shown in Figure 6. According to Fig. 6 and formula (7), h_t represents the output value of the current neuron, Z_t represents the update gate, h_t-1 represents the output of the previous neuron, and h_t' represents the pending output value in the current neuron. The output value h_t of this neuron is a value obtained by subtracting the update gate Z_t from 1 and multiplying the output value h_t-1 of one neuron. This value is then added to the value obtained by multiplying the update gate Z_t by the pending output value h_t' in this neuron.

IV. EXPERIMENTAL STEPS AND RESULTS

In this section, the detailed steps of the experimental implementation and the analysis of the experimental results are described. In order to evaluate the effectiveness of the proposed method, two other models are used for

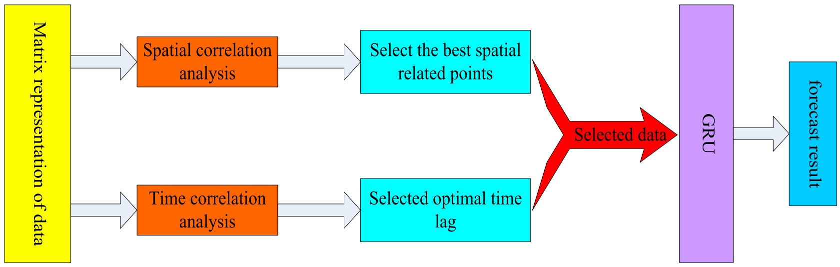


FIGURE 7. Forecast process block diagram.

comparative experiments [30]–[38], Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) models. The regression prediction results were evaluated using several indicators of MAE, MSE, RMSE, and MAPE [39]–[45].

A. EXPERIMENTAL STEPS

This study firstly analyzes the correlation of input traffic data, initializes the relevant time lag and the number of spatial segments used for prediction according to the selection rules, and constructs a neural network prediction model using the initialized network data. After the prediction model is established, the time lag of the best input data and the number of spatial segments are determined by STFSA. In terms of predictive model selection, this paper selects GRU as the predictive model and processes the matrix data containing relevant traffic information. Figure 7 is a traffic flow forecasting framework.

The experimental steps are as follows:

Step1: Select experimental data;

Step2: Analyze the temporal and spatial correlation of traffic data;

Step3: Introduce a spatiotemporal feature selection algorithm (STFSA);

Step4: The extracted traffic flow data are processed according to the time dimension and the spatial dimension to construct a space-time correlation matrix with traffic flow information;

Step 5: Input the constructed space-time traffic data matrix into the GRU prediction model;

Step 6: The traffic flow at the predicted time of the predicted point is used as the output of the GRU prediction model.

B. EVALUATION FUNCTION

This study used Mean Absolute Error (MAE), Root Mean Squard Error (RMSE), and Mean Absolute Percentage Error (MAPE) to evaluate regression prediction results [46]–[50].

MAPE is the average of the absolute values of relative percentage errors. It can be used to evaluate the prediction results of a model. The calculation formula is as follows:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y*-y}{y}\right| \tag{8}$$

FIGURE 8. Segments selected for experimental data.

MAE is the average absolute error, which is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y*-y| \tag{9}$$

RMSE is the root mean square error. The root mean square error is also called the standard error. It is the square root of the ratio of the square of the deviation between the observed value and the true value and the number of observations. The root mean square error is used to measure the deviation between the observed value and the true value. Standard error is very sensitive to very large or very small errors in a set of measurements. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y*-y)^2} \tag{10}$$

In these four formulas, n is the sample size, y^* is the predicted value, and y is the actual value.

C. DATA SOURCE

The data selected for this test are the traffic flow data of the Xiangjiang Middle Road in Changsha City except the weekend of May 20th, 2018, for a total of 40 days. Changsha City is affiliated to Hunan Province, the provincial capital of Hunan Province, referred to as Chang. It is located in the north of Hunan Province, north of the Xiangjiang River and

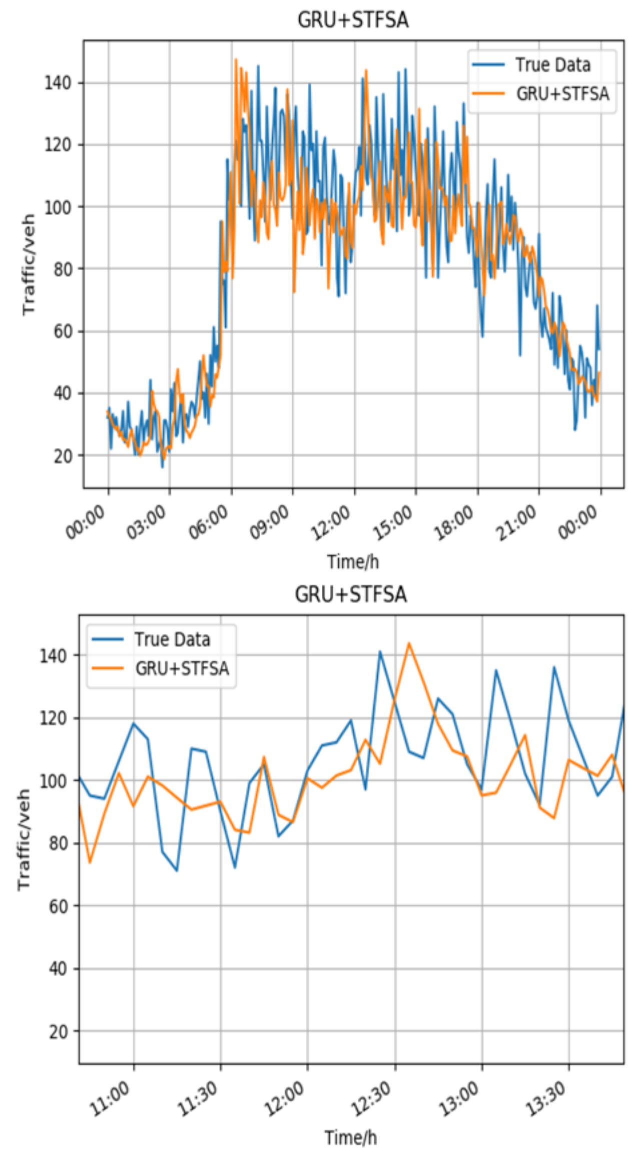


FIGURE 9. GRU+STFSA model prediction results.

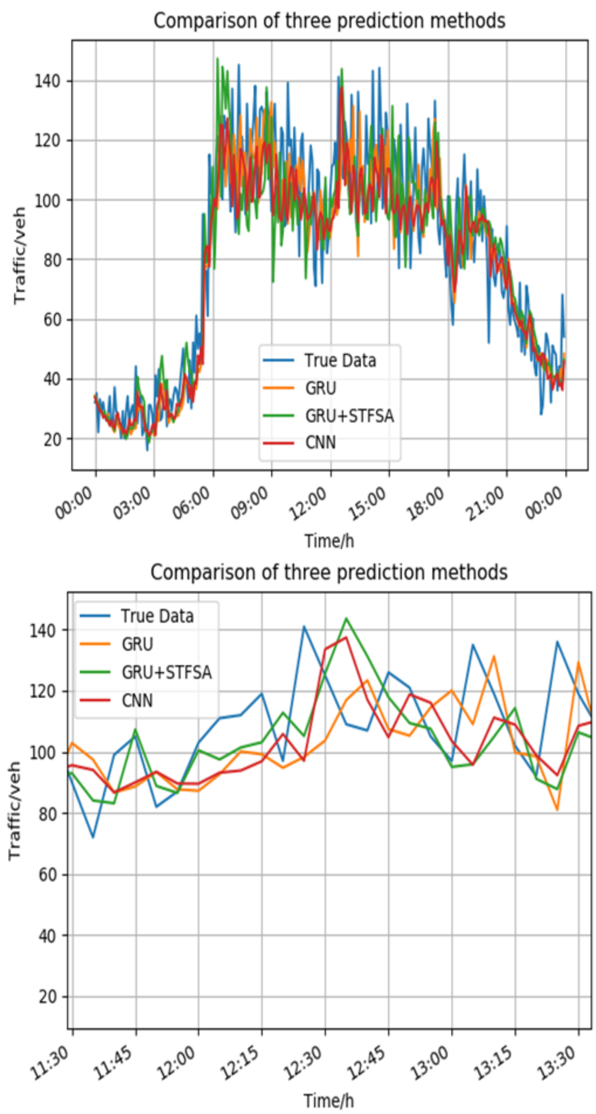


FIGURE 10. Comparison of prediction results from three methods.

the western edge of the Xiangliu Basin. The data sampling interval is 5 min, and there are 288 samples per day. During the experiment, the traffic data of the first 35 days of data collection was used as the training data of the GRU network. The last 5 days were used as test samples, and there were 10080 training samples containing the verification data and 1,440 test samples.

D. ANALYSIS OF SINGLE-STEP PREDICTION RESULTS

This article uses the tensorflow framework in Python and the Keras library to implement the GRU+STFSA method. The actual value and predicted value of the traffic flow on the road section on June 20th are shown in the figure. The upper part of Figure 9 is the comparison of the true values of the traffic flow at various times throughout the day on June 20th and the prediction results obtained using the prediction method proposed in this study. The lower part is the comparison of the true value of the traffic flow around 12 noon on June 20th and the prediction results obtained using the prediction method proposed by the Institute. As can be seen from Fig.9, the traffic flow began to climb from 6 am to 7 pm and peaked at around 8 am. Traffic flow remained relatively stable between 8 am and 12 noon, and traffic flow began to fall after 6 pm. The predicted value is in good agreement with the actual flow, and it can be seen that the algorithm is effective and feasible.

In order to analyze the prediction effect of the proposed method, this study also compares the three methods of CNN, GRU, GRU+STFSA, and the comparison results are shown in Figure 10. The upper part of Fig. 10 is a comparison of the true values of traffic flow at various times throughout the day on June 20th and the prediction results obtained using the three prediction methods. The lower part is a comparison chart of the true value of the traffic flow around 12 noon on June 20th and the prediction results obtained by the three prediction methods. The advantage of feature extraction using

TABLE 1. Comparison of 5min traffic flow forecast.

model	MAE	RMSE	MAPE
CNN	7.21	9.90	19.58%
GRU	7.20	9.96	20.78%
GRU+STFSA	7.05	9.59	17.80%

CNN is that the user does not care at all about the specific features, that is to say, the encapsulation of feature extraction is implemented. And CNN can share the convolution kernel and has no pressure on high-dimensional data processing. However, CNN is too dependent on the GPU when training the network, and the efficiency on the CPU is much lower than that of the GPU. When the data is insufficient, the predicted result is quite different from the actual result. The existence of the pooling layer leads to the loss of many very valuable information, and also ignores the relationship between the whole and the part. The GRU network is simple, and highly efficient, which automatically learns features and effectively models long-distance dependent information as well as solving the gradient disappearance problem of the RNN by using an update gate and a reset gate. However, the GRU does not perform well in processing the timing series. Therefore, this paper proposes to use GRU+STFSA to predict the traffic flow with spatio-temporal characteristics.

Table 1 shows the results of a 5 min traffic flow prediction using different algorithms. It can be seen that the MAPE value of CNN is 19.58%, which is lower than 20.78% of GRU, and its MAE value is 7.21, which is relatively lower than 7.20 of GRU. And the CNN has an RMSE value of 9.90, which is lower than the GRU. It can be concluded that CNN performs better than GRU in short-term traffic flow prediction. From the table, it can be also found that the GRU with the STFSA performs significantly better than the CNN and GRU in short-term traffic flow prediction. The MAPE value of GRU+STFSA is 17.80%, which is significantly lower than 19.75% of CNN and 20.78% of GRU. The MAE value of GRU+STFSA is also only 7.05, which is lower than CNN and GRU. Its RMSE value is 9.59, which is also lower than the first two. Through the comparative analysis of CNN, GRU and GRU+STFSA models, it is found that the addition of STFSA can effectively improve the prediction performance, and its RMSE, MAPE and MAE values are small.

E. ANALYSIS OF MULTI-STEP PREDICTION RESULTS

We further discuss the prediction performance of the proposed model and other models at different time periods, and predict the traffic flows of the next 10 min, 15 min and 20 min with different models, and compare their prediction performance. Table 2, Table 3, and Table 4 compare the effects of different network models on traffic flow prediction at different time intervals.

The following conclusions can be drawn from Table 2, Table 3 and Table 4. First, as the prediction time increases, the prediction performance of the prediction model decreases.

TABLE 2. Comparison of 10min traffic flow forecast.

model	MAE	RMSE	MAPE
CNN	8.19	11.06	21.36%
GRU	8.33	11.39	22.93%
GRU+STFSA	7.81	10.84	19.52%

TABLE 3. Comparison of 15min traffic flow forecast.

model	MAE	RMSE	MAPE
CNN	10.97	13.76	23.02%
GRU	11.71	14.88	24.62%
GRU+STFSA	9.27	12.57	21.33%

TABLE 4. Comparison of 20min traffic flow forecast.

model	MAE	RMSE	MAPE
CNN	11.43	14.29	24.13%
GRU	12.59	16.36	26.84%
GRU+STFSA	10.03	13.28	21.99%

In the same prediction model, when the time interval is increased from 10min to 20min, the MAE of the CNN algorithm increases from 8.19 to 11.43, and the RMSE increases from 11.06 to 14.29. The MAE of the GRU algorithm increased from 8.33 to 12.59, and the RMSE increased from 11.39 to 16.36. The MAE of the GRU+STFSA model increased from 7.86 to 10.03, and the RMSE increased from 10.84 to 13.28. Second, the accuracy of predictions between different models varies with time intervals. When the prediction interval was extended from 5 min to 20 min, the MAE and RMSE of the GRU+STFSA model increased by 2.98 and 3.69, respectively. The MAE and RMSE of the CNN model increased by 4.22 and 4.39, respectively; the MAE and RMSE of the GRU model increased by 5.39 and 6.4, respectively. From the above data, it can be found that the GRU+STFSA model proposed in this paper has the smallest error, and the error growth rate is the slowest with the expansion of the prediction time. According to the discussion above, GRU+STFSA has good prediction accuracy and is stable on different time interval prediction problems.

V. CONCLUSION AND PROSPECT

This paper proposes a short-term traffic flow prediction model that combines spatio-temporal analysis with a GRU deep learning framework. Firstly, the spatio-temporal correlation analysis of the traffic flow data of this experiment is carried out, and the spatio-temporal feature selection algorithm is used to define the optimal input time interval and spatial data volume. Then the extracted traffic flow data are developed according to the time dimension and the spatial dimension to construct a space-time correlation matrix with traffic flow information. A five-layer GRU network is constructed to process the spatiotemporal feature information of the internal traffic flow in the matrix. Finally, the prediction results are compared with the results of CNN and GRU prediction. The results show that the proposed model is superior to CNN model and GRU model in accuracy and stability.

There exists some weakness about this paper, other factors (e.g. weather conditions,) of traffic flow are not considered, and only the traffic flow of a single road segment is predicted. Next, how to predict the entire roadway network is what we plan to continue in the future.

REFERENCES

[1] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 1688–1711, 2019.

[2] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[3] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong, "Traffic speed prediction and congestion source exploration: A deep learning method," in *Proc. IEEE 16th Int. Conf. Data Mining*, Barcelona, Spain, Dec. 2016, pp. 499–508.

[4] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing*, vol. 137, no. 15, pp. 47–56, Aug. 2014.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[6] D. Yu, Y. Liu, and X. Yu, "A data grouping CNN algorithm for short-term traffic flow forecasting," in *Web Technologies and Applications*. Suzhou, China: Springer, 2016, pp. 92–103.

[7] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification," *Frontiers Comput. Sci.*, vol. 10, no. 1, pp. 96–112, Feb. 2016.

[8] H. Lu, N. Zhang, J. X. Xia, and W. Huang, "Research progress of short term traffic flow prediction methods," *J. Transp. Eng. Inf.*, vol. 7, no. 4, pp. 84–91, 2009.

[9] H. Cheng, K. Xie, G. Song, and T. Wu, "Hybrid process neural network based on spatio-temporal similarities for short-term traffic flow prediction," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 253–258.

[10] X. Min, J. Hu, and Z. Zuo, "Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 1535–1540.

[11] W. Gao, B. Lu, T. Yun, and W. Tan, "Short-term traffic flow forecasting based on spatiotemporal characteristics of traffic flow and RBF neural network," *J. Traffic Inf. Saf.*, vol. 29, no. 1, pp. 16–24, 2011.

[12] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[13] L. Cui, "Research on methods of traffic flow predicting of urban road network based on multi cross-section information," Ph.D. dissertation, Dalian Maritime Univ., Dalian, China, 2012.

[14] C. Dong, C. Shao, C. Zhuge, and M. Meng, "A short-term state-space model for free flow prediction based on spatio-temporal characteristics," *China Civil Eng. J.*, vol. 46, no. 8, pp. 111–118, 2013.

[15] S. Qiu, B. Lu, Q. Ma, W. Zhou, and Q. Zhang, "Traffic forecasting based on spatial-temporal characteristics analysis and data fusion," *J. Wuhan Univ. Technol. (Inf. Manage. Eng. Ed.)*, vol. 30, no. 2, pp. 156–160, 2015.

[16] L. Li, S. He, and J. Zhang, "Online short-term traffic flow prediction considering the impact of temporal-spatial features," *J. Transp. Syst. Eng. Inf.*, vol. 16, no. 5, pp. 165–171, 2016.

[17] J. Kolen and S. Kremer, *Gradient Flow in Recurrent Nets: The Difficulty of Learning Long Term Dependencies*. Hoboken, NJ, USA: Wiley, 2009.

[18] 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," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[19] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom*, Dec. 2015, pp. 153–158.

[20] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Automat.*, Nov. 2016, pp. 324–328.

[21] H. Liu and R. Setiono, "Feature selection and classification-a probabilistic wrapper approach," in *Proc. 9th Int. Conf. Ind. Eng. Appl. AI ES*, 1996, pp. 419–424.

[22] Y. Song, R. N. Rao, and J. Shi, "Relation classification in knowledge graph based on natural language text," in *Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci.*, Nov. 2018, pp. 1104–1107.

[23] C. Ma, R. He, and W. Zhang, "Path optimization of taxi carpooling," *PLOS ONE*, vol. 13, no. 8, 2018, Art. no. e0203221.

[24] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment," *IEEE Access*, vol. 6, pp. 52471–52478, 2018.

[25] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver injury severity at highway-rail grade crossings accidents," *J. Adv. Transp.*, vol. 2018, 2018, Art. no. 9841498.

[26] Y. Zou, X. Zhong, J. Tang, X. Ye, L. Wu, M. Ijaz, and Y. Wang, "A copula-based approach for accommodating the underreporting effect in wildlife–vehicle crash analysis," *Sustainability*, vol. 11, no. 2, p. 418, 2019.

[27] R. Cheng and Y. Wang, "An extended lattice hydrodynamic model considering the delayed feedback control on a curved road," *Phys. A, Stat. Mech. Appl.*, vol. 513, pp. 510–517, Jan. 2019.

[28] C. Jiang, H. Ge, and R. Cheng, "Mean-field flow difference model with consideration of on-ramp and off-ramp," *Phys. A, Stat. Mech. Appl.*, vol. 513, pp. 465–476, Jan. 2019.

[29] R. Cheng, H. Ge, and J. Wang, "An extended continuum model accounting for the driver's timid and aggressive attributions," *Phys. Lett. A*, vol. 381, no. 15, pp. 1302–1312, 2017.

[30] Y. Sun, H. Ge, and R. Cheng, "An extended car-following model considering driver's memory and average speed of preceding vehicles with control strategy," *Phys. A, Stat. Mech. Appl.*, vol. 521, pp. 752–761, May 2019.

[31] H. Niu, X. Tian, and X. Zhou, "Demand-driven train schedule synchronization for high-speed rail lines," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2642–2652, Oct. 2015.

[32] H. Niu, X. Zhou, and X. Tian, "Coordinating assignment and routing decisions in transit vehicle schedules: A variable-splitting Lagrangian decomposition approach for solution symmetry breaking," *Transp. Res. B, Methodol.*, vol. 107, pp. 70–101, Jan. 2018.

[33] C. Ma, C. Ma, Q. Ye, R. He, and J. Song, "An improved genetic algorithm for the large-scale rural highway network layout," *Math. Problems Eng.*, vol. 2014, 2014, Art. no. 267851.

[34] J. Tang, J. Liang, C. Han, Z. Li, and H. Huang, "Crash injury severity analysis using a two-layer Stacking framework," *Accident Anal. Prevention*, vol. 122, pp. 226–238, Jan. 2019.

[35] J. Tang, J. Liang, S. Zhang, H. Huang, and F. Liu, "Inferring driving trajectories based on probabilistic model from large scale taxi GPS data," *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 566–577, Sep. 2018.

[36] Z. He, W. Guan, and S. Ma, "A traffic-condition-based route guidance strategy for a single destination road network," *Transp. Res. C, Emerg. Technol.*, vol. 32, pp. 89–102, Jul. 2013.

[37] C. Ma, D. Yang, J. Zhou, Z. Feng, and Q. Yuan, "Risk riding behaviors of urban e-bikes: A literature review," *Int. J. Environ. Res. Public Health*, vol. 16, no. 13, p. 2308, 2019.

[38] C. X. Ma, Y. Z. Li, R. C. He, B. Qi, F. Wu, L. Sun, and A. X. Diao, "Bus-priority intersection signal control system based on wireless sensor network and improved particle swarm optimization algorithm," *Sensor Lett.*, vol. 10, no. 8, pp. 1823–1829, 2012.

[39] C. Ma, Y. Li, R. He, F. Wu, B. Qi, and Q. Ye, "Route optimisation models and algorithms for hazardous materials transportation under different environments," *Int. J. Bio-Inspired Comput.*, vol. 5, no. 4, pp. 252–265, 2013.

[40] W. Hao, C. Ma, B. Moghimi, Y. Fan, and Z. Gao, "Robust optimization of signal control parameters for unsaturated intersection based on tabu search-artificial bee colony algorithm," *IEEE Access*, vol. 6, pp. 32015–32022, 2018.

[41] H. Pu, Y. Li, C. Ma, and H. Mu, "Analysis of the projective synchronization of the urban public transportation super network," *Adv. Mech. Eng.*, vol. 9, no. 6, pp. 1–8, 2017.

[42] C. Ma, "Network optimisation design of hazmat based on multi-objective genetic algorithm under the uncertain environment," *Int. J. Bio-Inspired Comput.*, vol. 12, no. 4, pp. 236–244, 2018.

[43] C. Ma and R. He, "Green wave traffic control system optimization based on adaptive genetic-artificial fish swarm algorithm," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2073–2083, Jul. 2019.

[44] C. Ma, W. Hao, R. He, X. Jia, F. Pan, J. Fan, and R. Xiong, "Distribution path robust optimization of electric vehicle with multiple distribution centers," *PLoS ONE*, vol. 13, no. 3, 2018, Art. no. e0193789.

[45] C. Ma, W. Hao, F. Pan, and W. Xiang, "Road screening and distribution route multi-objective robust optimization for hazardous materials based on neural network and genetic algorithm," *PLoS ONE*, vol. 13, no. 6, 2018, Art. no. e0198931.

[46] C. Ma, W. Hao, R. He, and B. Moghimi, "A multiobjective route robust optimization model and algorithm for hazmat transportation," *Discrete Dyn. Nature Soc.*, vol. 2018, 2018, Art. no. 2916391.

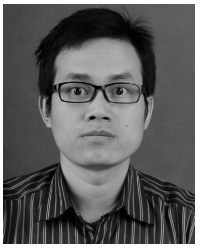
[47] J. Weng, G. Du, D. Li, and Y. Yu, "Time-varying mixed logit model for vehicle merging behavior in work zone merging areas," *Accident Anal. Prevention*, vol. 117, pp. 328–339, Aug. 2018.

[48] F. Chen, M. Song, X. Ma, and X. Zhu, "Assess the impacts of different autonomous trucks' lateral control modes on asphalt pavement performance," *Transp. Res. C, Emerg. Technol.*, vol. 103, pp. 17–29, Jun. 2019.

[49] X. Xu, Ž. Šarić, F. Zhu, and D. Babić, "Accident severity levels and traffic signs interactions in state roads: A seemingly unrelated regression model in unbalanced panel data approach," *Accident Anal. Prevention*, vol. 120, pp. 122–129, Nov. 2018.

[50] W. Wu, R. Liu, W. Jin, and C. Ma, "Stochastic bus schedule coordination considering demand assignment and rerouting of passengers," *Transp. Res. B, Methodol.*, vol. 121, pp. 275–303, Mar. 2019.

GUOWEN DAI received the B.E. degree from Lanzhou Jiaotong University, China, in 2017, where he is currently pursuing the master's degree in transportation planning and management with the School of Traffic and Transportation. His research interests include short-term traffic flow forecast and data mining.



CHANGXI MA received the B.S. degree in traffic engineering from the Huazhong University of Science and Technology, in 2002, and the Ph.D. degree in transportation planning and management from Lanzhou Jiaotong University, in 2013, where he is currently a Professor. He is the author of three books and more than 100 articles. His research interests include ITS, traffic safety, and hazardous materials transportation.

XUECAI (DANIEL) XU received the B.S. degree from the Xi'an University of Technology, Xi'an, China, the M.E. degree from Southwest Jiaotong University, Chengdu, China, and the Ph.D. degree from the University of Nevada, Las Vegas, NV, USA. He was a Senior Research Fellow with the School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, and a Senior Research Assistant with the Department of Civil Engineering, University of Hong Kong, Hong Kong. He is currently an Assistant Professor with the School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China. He has published more than 20 articles indexed by SCI/SSCI/EI. His research interests include transportation safety, intelligent transportation systems, and system engineering.