IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 13, 2019, accepted July 18, 2019, date of publication July 30, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931920

Short-Term Traffic Flow Prediction Considering Spatio-Temporal Correlation: A Hybrid Model Combing Type-2 Fuzzy C-Means and Artificial Neural Network

JINJUN TANG[1], LEXIAO LI[1], ZHENG HU[2], AND FANG LIU[3]

[1]Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[2]School of Intelligent transportation, Hunan Communication Engineering Polytechnic, Changsha 410132, China
[3]School of Transportation Engineering, Changsha University of Science and Technology, Changsha 410205, China

Corresponding author: Fang Liu (rcliufang@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71701215, in part by the Foundation of Central South University under Grant 502045002, in part by the Science and Innovation Foundation of the Transportation Department in Hunan Province under Grant 201725, and in part by the Postdoctoral Science Foundation of China under Grant 2018M630914 and Grant 2019T120716.

ABSTRACT Traffic flow prediction is a key step to the efficient operation in the intelligent transportation systems. This paper proposes a hybrid method combing clustering methods and spatiotemporal correlation to predict future traffic trends based on artificial neural network. First, for the traffic flow collected from different loop detectors, a spatio-temporal correlation of data samples is evaluated by considering time correlation and spatial equivalent distance. Second, in order to improve classifying performance and reliability to anomalous data samples, a type-2 fuzzy c-means (FCM) is adopted to make fuzzification of the membership function. Then, a hybrid prediction model combined classification algorithm and neural network is designed to predict various patterns or trends in traffic flow data. Furthermore, the results from the prediction model are modified according to quantized spatio-temporal correlation. Finally, traffic volume data collected from the highway is used to optimize the parameter in the prediction model combination. Several traditional models are used as candidates in comparison, and the higher prediction accuracy demonstrates the effectiveness and feasibility of the hybrid prediction model.

INDEX TERMS Traffic flow prediction, spatio-temporal correlation, type-2 fuzzy c-means, artificial neural network.

I. INTRODUCTION

With the development of social economy and the acceleration of urbanization, the range of travel in urban city has been greatly expanded. Although the urban transportation system has continuously developed, and its supply capability is also improved largely through the construction of transportation infrastructure, it still cannot meet the demand of travel which causes problems such as traffic congestion, accidents and pollution. Furthermore, these issues could result in negative impacts on the quality of city living environment. The emergence of Intelligent Transportation Systems (ITS) provides a new way to solve these problems. Traffic guidance and traffic control systems are two very important components of ITS, in which collecting real-time and accurate traffic flow information is the key factor to impact the performance of guidance and control, and high accurate traffic flow prediction is one of most important steps to achieve the application of ITS.

Historical data reflect the changing characteristics or patterns of traffic flow, which is an important basis for short-term traffic flow prediction model. A good prediction model has great advantages in historical data analysis and feature recognition. Recently, traffic prediction has been a direction concerned by scholars in the field of transportation [1], which can be generally divided into the following three parts: statistical models [2]–[6], artificial intelligence models [7]–[27], hybrid models [28]–[35].

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

(1) Because the statistical model (SM) is easy to capture the periodicity and repeatability of traffic flow, and the model performance is relatively stable, it has been widely used in the field of traffic flow prediction. Cetin and Comert [2] proposed an adaptive algorithm for the consideration of emergencies to update the intercept of the autoregressive comprehensive moving average model [3]. In 2015, Kumar and Vanajakshi [4] proposed a prediction method using seasonal ARIMA (SARIMA) model. The model changed the input time series into a stationary one through difference and used the maximum likelihood method to find the parameters of the model. Considering the effect of the cross correlation between the upstream and downstream [5] on the prediction of the target traffic flow data, Chandra and Al-Deed [6] applied the vector autoregressive model (VAR) to conduct traffic prediction in the case of highway and found that it had better prediction accuracy than the traditional ARIMA model. In recent years, some new methods have been proposed, such as generalized autoregressive conditional heteroscedasticity (GARCH) [2], which improves the accuracy of the model by optimizing the degree of consideration of uncertain factors.

(2) Different from SM model, artificial intelligence technology has attracted attention from the field of transportation [7]–[10] because of its generalization ability and efficient learning ability. (a) Neural network is one of the commonly used prediction methods. Recurrent Neural Network (RNN) introduces the concept of "memory", making the output dependent on input and memory storage, and this model is widely used in traffic prediction [11]–[13]. Long Short-Term Memory Neural Network (LSTM) [14]–[16] and state-space neural network model [17] also expressed better performance in traffic flow prediction comparing with some traditional methods. (b) Yan *et al.* [18] proposed a prediction model based on Fuzzy neural network [19], which divided traffic distribution patterns with similar characteristics, and utilized the training method of online circulation to improve the adaptive ability of the model. Dimitriou *et al.* [20] used genetic algorithm (GA) to optimize the membership function of adaptive fuzzy rules offline and online, which has a beneficial effect in traffic prediction of urban trunk roads. (c) Compared with neural networks, support vector machines (SVM) [21] can overcome the overfitting characteristics of neural networks. Its principle is to use the nonlinear mapping algorithm to transform the low-dimensional input space into the high-dimensional feature space, making it possible for the linear algorithm to perform linear analysis on the input space. Support for adjacent regression (SVR) [22], [23] is one kind of SVM, which performs well in time series prediction and analysis. Zhang and Liu [24] proposed a non-parametric least squares support vector machine (LS-SVM) prediction method, which has high prediction accuracy and excellent stability for traffic data with weak regularity. (d) Kalman filtering is an efficient algorithm that describes a series of recursive mathematical formulas to optimize system state estimation, which performs well in traffic forecasting [25], [26] and denoising. Guo *et al.* [27] proposed an adaptive kalman filtering method that can update the process variance, which has better performance in the case of unstable data.

(3) A single prediction model may not be applicable for all scenarios due to different application conditions. The hybrid model combines different prediction algorithms or models to make full use of advantages of individual models and obtain better prediction results. Voort *et al.* [28] combined the Kohonen self-organizing classifier with the ARIMA time series model to create a hybrid forecasting model. Chang and Tsai [29] introduced a hybrid SVMGM method combining grey prediction model (GM) and support vector machine (SVM), which reduced the overshoot effect in traffic flow prediction. The combination of genetic algorithm and time-delayed neural network [30], the combination of discrete wavelet transform and neural network [31], [32], these hybrid models are more conducive to traffic prediction than the single model [33]–[35].

Most of the previous studies were single-step predictions based on a small amount of data. One reason why the short-term traffic flow model did not perform as expected was that the prediction was regarded as a point process and ignored the dynamic mobility of traffic flow. Cheng *et al.* [36] explored the spatio-temporal autocorrelation of the traffic flow data on the road network. Furthermore, the interaction between dynamic traffic flows in adjacent road segment and time periods is obvious and understanding autocorrelation of traffic flow form time and space scope will be of great help to improve short-term prediction accuracy [37]. Therefore, in recent works, researchers consider to using spatio-temporal correlation among traffic flow data from target and surrounding detectors in the road network to improve prediction performance. Yue and Yeh [38] discussed the significance of spatio-temporal relationship for short-term traffic flow prediction and quantified the dependence of traffic flow. Min and Wynter [39] developed a new way to estimate spatio-temporal interaction of road traffic and demonstrated it on the test network. Haworth and Cheng [40] used the traffic flow condition of upstream and downstream for adjacent links to finish prediction using non-parametric spatio-temporal kernel regression model. Zheng *et al.* [41] proposed an online prediction method based on the feature selection of spatio-temporal traffic patterns of intelligent algorithms and the optimization of state vectors in the off-line process. In 2019, a non-linear Granger causality analysis was proposed to detect the spatiotemporal causality between various roads [42].

From the summaries of current works, researchers have taken into account the spatio-temporal correlation of traffic flow in the prediction. However, these works still did not clearly and effectively quantify the spatio-temporal correlation of traffic flow within the road network. At the same time, the traffic flow in the road network expresses a variety of distribution modes or patterns, and the traffic flow under different patterns behave different characteristics. So, firstly, classification method is applied to categorize original traffic flow data into different patterns. Liu *et al.* (2018) [43] proposed a preprocessing process for predictive applications, which uses heuristic feature selection to determine the best features of the input vector to improve the performance of traffic prediction algorithms. Then, according to clustering results, prediction model can be further modified to reach better prediction results. Some previous researches have introduced related works based on fuzzy c-means (FCM) [44], [45], k-means clustering [46] and other classifiers [28]. The prediction model in this study is designed to integrate the spatial-temporal distribution characteristics of traffic flow and consider using these characteristics under different modes to improve prediction accuracy.

Firstly, the spatio-temporal correlation between each detector is established by using historical data, and it is quantified by considering correlation between historical data and topology structure. Secondly, an improved FCM, type-2 FCM algorithm is applied to extract distribution characteristics of different traffic modes. Using the fixed membership function in traditional FCM [44], [45], when dataset include abnormal samples, classification results will express deviation. Unlike the traditional type 2 fuzzy set with interval membership [47], [48], the paper uses a simplified form of the type-2 FCM algorithm. By adding a degree of membership confirmation, the degree of influence of the anomaly point on the algorithm is reduced and the computational efficiency of the algorithm is improved. Type-2 FCM algorithm with degree of confirmation is used for traffic pattern classification, which can identify abnormal data in samples and improve the accuracy of classification.

Then, consider the actual impact of spatio-temporal correlation on traffic flow conditions, the paper uses the quantized spatio-temporal correlation to correct the preliminary results of the improved model using the weighted average method to obtain the final prediction results.

Finally, the parameters of the improved model are optimized by the traffic flow data of the highway, and the effectiveness of the improved model is verified. The improved model still maintains a high level of prediction accuracy compared to the FCM-NN and NN models.

This paper is organized as follows. Section 2 introduces methodology and structure of calculation process. Section 3 describes the data sources and model calibration. The prediction result in the case study is provided in Section 4. Section 5 is the conclusion of the study.

II. METHODOLOGY

A. SPATIO-TEMPORAL CORRELATION ANALYSIS

For most traffic flow prediction models, the historical variation of traffic flow time series is the main basis to achieve high prediction performance. Although there exists a strong randomness and uncertainty in traffic system, the traffic flow on the same road still expresses similar patterns to follow, which is greatly helpful for the prediction of traffic flow. Generally, the traffic condition of specific section on the road will often be affected by the conditions from upstream and downstream, and will also be impacted by the traffic states on adjacent lanes. Thus, the spatio-temporal variation characteristics of traffic flow in surrounding detectors will definitely provide valuable information to implement traffic flow prediction and modeling. In this study, the spatio-temporal correlation of traffic flow data in the road network is considered to implement the traffic flow prediction and modeling.

Firstly, for the spatial correlation, we define two indexes to evaluate it: (1) the direct physical distance d, it can simply represent the spatial position relationship between various detectors. The spatial distance matrix (SD) is a common method to express spatial features [49]. In general, the degree of data impact between different detectors gradually decreases as the physical distance increases. However, the physical distance d does not accurately reflect the spatial characteristics of the detector traffic flow in different lanes. (2) Therefore, we need further define the spatial correlation level l to estimate topological structure of detectors in the road network. For the same road, the characteristics of the detector traffic flow are more likely to spread to the upstream and downstream of the same lane. The traffic flow relationship between any two adjacent detectors on the same road can be divided into two cases:1) the two detectors are located from the upper and lower streams in the same lane, and the traffic flow data between the two detectors expresses direct influence. 2) the two detectors are located in parallel positions of different lanes, and the traffic flow data between them has indirect influence. Accordingly, the values of spatial correlation level can be estimated using following rules (shown in Fig.1): 1) for the target detector we focusing on predicting traffic flow patterns, the value of spatial correlation level is set as $l = 1$; 2) for the upstream and downstream detectors directly adjacent to the target detector in the same lane, the spatial correlation level is set as $l = 2$; 3) for the detectors directly be adjacent to the target detector in different lanes, the spatial correlation level is set as $l = 3$; 4) for the other detectors in the studying area, the spatial correlation level is set as $l = 4$;

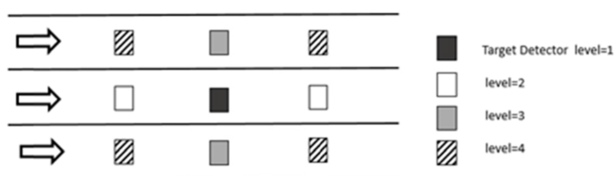


FIGURE 1. Spatial correlation level of Road detectors.

Then, for the temporal correlation, the correlation coefficient (R) is used to estimate temporal characteristics of traffic flow as following., For the traffic flow data collected from two detectors: $\{x^1, x^2, x^3, \ldots x^n\}$ and $\{y^1, y^2, y^3, \ldots y^n\}$, and N represents the length of the data samples.

$$R = \frac{Cov(X, Y)}{\sqrt{DX} \cdot \sqrt{DY}} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}} \quad (1)$$

Finally, based on the spatial and temporal relation between detectors aforementioned, a combined indicator to quantify the traffic flow correlation as following:

$$\lambda = (\frac{1}{d \cdot l})^{R} \tag{2}$$

where d is the actual physical distance between the surrounding detector and the target detector, l represents the spatial correlation level, and R indicates the temporal correlation coefficient.

B. TYPE-2 FCM CLUSTERING ALGORITHM

In order to deeply exploring characteristics of traffic flow, a clustering method based on Fuzzy C-Means (FCM) is utilized to categorize the changing patterns implying in the historical data. Its idea is to maximize the similarity between objects divided into the same cluster and minimize the similarity between different clusters. The FCM algorithm is generalized from Hard C-Means clustering algorithm (HCM). The main difference of FCM to HCM is that the values of given data point belong to each cluster estimated by fuzzy membership degree between 0 and 1. The membership degree of each data sample to all cluster centers is obtained by optimizing the objective function.

1) ORIGINAL FCM ALGORITHM

Firstly, in HCM algorithm, the objective function Q is defined to minimize the total distance as follows:

$$\min Q = \sum_{i=1}^{c} \sum_{k=1}^{N} d_{k,i}^{2} \tag{3}$$

Accordingly, the objective function is improved by adding membership degree in the FCM is defined as follows:

$$\min Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{k,i}^{m} d_{k,i}^{2} \tag{4}$$

where $d_{k,i}^{2}$ denotes the distance between the data sample x_k to the cluster center V_i, for the ith cluster, and $u_{k,i}$ denotes the degree of membership of x_k in the ith cluster, m represents the fuzzy weighted index, and its value will affect the performance of the algorithm, and $m \geq 1$, c and N respectively mean the total number of clusters and data samples. After setting initial values for u and v, the optimal values of these two key variables can be achieved by following iterative calculation procedure.

$$u_{ik} = \frac{1}{\sum_{s=1}^{c} \left(\frac{d_{k,i}}{d_{k,s}}\right)^{-\frac{2}{m-1}}} \tag{5}$$

$$v_i = \frac{\sum_{k}^{N} u_{k,i}^{m} x_k}{\sum_{k=1}^{N} u_{k,i}^{m}} \tag{6}$$

2) TYPE-2 FCM CLUSTERING ALGORITHM

Type-2 fuzzy set is an extension of traditional fuzzy membership definition, and it has stronger uncertainty expression ability. In this study, a new parameter, ρ ranges in [0,1], is defined as the degree of membership degree and adopted in Type-2 fuzzy clustering algorithm.

Furthermore, the object function Q can be further modified according to Type-2 fuzzy set as follows:

$$\text{mimQ} = \sum_{i=1}^{c} \sum_{k=1}^{N} \left(\rho_k^2 u_{k,i}^{m} d_{k,i}^{2} - 2\beta\rho_k\right) \tag{7}$$

and

$$\sum_{i=1}^{c} u_{k,i} = 1$$

where the ρ_k is applied to estimate uncertainty of membership degree, and β is the design parameter, which affects the effectiveness of the algorithm in identifying abnormal points. We can see that if $\rho_k = 1$, then Equation (7) will be the same as traditional FCM.

The objective function of Equation (7) is divided into two parts. The former part can make all ρ_k values as small as possible, especially for data points that deviate from the cluster center ($\sum_i^C u_{k,i}^m d_{k,i}^2$ value is relatively large). The latter part makes all ρ_k value data points as large as possible, especially the data points very close to the cluster center ($\sum_i^C u_{k,i}^m d_{k,i}^2$ value is relatively small).By the action of these two parts, the ρ value of the normal data can be close to 1, and the ρ value of the abnormal point is close to zero.

Then, constructing the Lagrange function for the a given data sample x_k

$$Q_k = \sum_{i=1}^{c} \left(\rho_k^2 u_{k,i}^{m} d_{k,i}^{2} - 2\beta\rho_k\right) - \lambda \left(\sum_{i=1}^{c} u_{k,i} - 1\right) \tag{8}$$

Make $\frac{\partial Q_k}{\partial u_{k,i}} = 0$, obtain $\frac{\partial Q_k}{\partial u_{k,i}} = m u_{k,i}^{m-1} \rho_k^2 d_{k,i}^2 - \lambda = 0$,

So, $u_{k,i} = \left(\frac{\lambda}{m\rho_k^2 d_{k,i}^2}\right)^{\frac{1}{m-1}}$, consider the condition $\sum_{i=1}^{c} u_{k,i} = 1$,

$$\sum_{i=1}^{c} u_{k,i} = \sum_{i=1}^{c} \left(\frac{\lambda}{m\rho_k^2 d_{k,i}^2}\right)^{\frac{1}{m-1}} = \lambda^{\frac{1}{m-1}} \sum_{i=1}^{c} \frac{1}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}} = 1$$

So, $\lambda = \frac{1}{\left(\sum_{i=1}^{c} \frac{1}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}}\right)^{m-1}}$,

Then, we can obtain the iterative value of $u_{k,i}$

$$u_{k,i} = \left(\frac{\lambda}{m\rho_k^2 d_{k,i}^2}\right)^{\frac{1}{m-1}} = \frac{\sum_{i=1}^{c} \frac{1}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}}}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}}$$

$$= \frac{\frac{1}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}}}{\sum_{i=1}^{c} \frac{1}{\sqrt[m-1]{m\rho_k^2 d_{k,i}^2}}} = \frac{\frac{1}{\sqrt[m-1]{d_{k,i}^2}}}{\sum_{i=1}^{c} \frac{1}{\sqrt[m-1]{d_{k,i}^2}}} \tag{9}$$

Next, in order to calculate the clustering center $V_i = (v_i^{(1)}, \cdots, v_i^{(m)})$, the partial derivative tth component $v_i^{(t)}$ in v_i and be set to zero as follows:

$$\frac{\partial Q}{\partial v_i^{(t)}} = \sum_{k=1}^{N} \rho_k^2 u_{k,i}^m 2\left(v_i^{(t)} - x_k^{(t)}\right) = 0,$$

Get

$$v_i^{(t)} = \left(\frac{\sum_{k=1}^{N} \rho_k^2 u_{k,i}^m x_k^{(t)}}{\sum_{k=1}^{N} (\rho_k^2 u_{k,i}^m)}\right).$$

Let

$$\omega_{k,i} = \rho_k^2 u_{k,i}^m \tag{10}$$

Then

$$V_i = \frac{\sum_{k=1}^{N} \omega_{k,i} X_k}{\sum_{k=1}^{N} \omega_{k,i}} \tag{11}$$

By deriving the Lagrange function of the data x_k, the iteration Equation (11) of the cluster center can be obtained. Different from the traditional FCM derivation in Equation (6), the derivation formula of type-2 FCM includes the effect of ρ. The formula $v_i^{(t)} = \left(\frac{\sum_{k=1}^{N} \rho_k^2 u_{k,i}^m x_k^{(t)}}{\sum_{k=1}^{N} (\rho_k^2 u_{k,i}^m)}\right)$ can weaken the effect of the data samples from all cluster centers (or abnormal data samples) by the influence of ρ_k^2.

In the updating procedure of ρ_k, we can similarly consider the partial derivation in Equation (7), and then obtain expression as $2\rho_k^2 u_{k,i}^m d_{k,i}^2 - 2\beta = 0$, $\rho_k \in [0, 1]$, it can be represented as:

$$\rho_k = min\left(\frac{\beta}{\sum_{i=1}^{c} u_{k,i}^m d_{k,i}^2}, 1\right) \tag{12}$$

The Equation (12) indicates that the size of β should be close to the normal value of $u_{k,i}^m d_{k,i}^2$, so that ρ can play a role in screening abnormal data and normal data.

C. ARTIFICIAL NEURAL NETWORK

After implementing Type-2 FCM clustering algorithm, we obtain different clusters or variation patterns in original traffic flow data. For each cluster or pattern, a learning mechanism based on Artificial Neural network is designed to achieve high prediction performance. The BP neural network is adopting in this study, which is a multi-layer feedforward neural network trained by error inverse propagation algorithm. Because of its simple structure and strong learning ability, BP neural network has been widely used in the fields of function approximation, pattern recognition, information classification and data compression. The learning rule is that the steepest descent method is adopted to adjust the weights and thresholds of the neural network continuously through reverse retransmission so as to minimize the square error of the network. Its network model topology includes input layer, hidden layer and output layer. The activation function is sigmoid function, whose derivative is related to itself, that is, $f'(x) = f(x)(1 - f(x))$.

The process of BP neural network error is calculated as follows

(1) The generalized error d_t^k of each unit of the output layer is defined as difference between target $T_k = \left(y_1^k, y_2^k, \ldots, y_q^k\right)$ and the output C_t of the network:

$$d_t^k = \left(y_t^k - C_t\right) \cdot \mathrm{f} = \left(y_t^k - C_t\right) \cdot C_t\,(1 - C_t)\,, \quad \mathrm{t} = 1, 2, \ldots, \mathrm{q}. \tag{13}$$

where q is the dimension of the target vector.

(2) The generalized error e_j^k of each unit of the intermediate layer is calculated by using the connection weight v_{jt}, the generalized error d_t of the output layer, and the output b_j of the intermediate layer.

$$e_j^k = \left[\sum_{t=1}^{q} d_t . v_{jt}\right] b_j\left(1 - b_j\right) \tag{14}$$

(3) The connection weight v_{jt} and the threshold γ_t are adjusted by the error d_t^k and the output b_j of each unit of the intermediate layer.

$$v_{jt}\,(N+1) = v_{jt}\,(N) + \alpha . d_t^k . b_j \tag{15}$$
$$\gamma_t\,(N+1) = \gamma_t\,(N) + \alpha . d_t^k \tag{16}$$
$$\mathrm{t} = 1, 2, \ldots, \mathrm{q}; \quad \mathrm{j} = 1, 2, \ldots, \mathrm{p}; \; 0 < \alpha < 1$$

(4) The connection weight w_{ij} and the threshold θ_j are corrected by using the error e_j^k, the input $P_k = (a_1, a_2, \ldots a_n)$ of each unit of the input layer.

$$(N+1) = w_{ij}\,(N) + \varepsilon e_j^k a_i^k \tag{17}$$
$$\theta_j\,(N+1) = \theta_j\,(N) + \varepsilon e_j^k \tag{18}$$
$$\mathrm{i} = 1, 2, \ldots, \mathrm{n}; \quad \mathrm{j} = 1, 2, \ldots, \mathrm{g}; \; 0 < \varepsilon < 1$$

where g is the dimension of the input vector of the middle layer and n is the dimension of the network input.

(5) Randomly select a set of input and target samples from learning dataset and repeat above calculation process until the global error of the network is less than a preset threshold, then the calculation procedure will end.

After obtaining prediction results from each sub-neural network based on Type-2 FCM clustering algorithm, the spatio-temporal characteristics between traffic flow data collected from different detectors are considered to adjust initial prediction results. Finally, a weighted average is adopted to fuse spatio-temporal correlation and prediction results based on clustering learning method as follows:

$$Q_t = \frac{\sum_{i=1}^{C} Q_i \bar{\lambda}_i}{\sum_{i=1}^{C} \bar{\lambda}_i} \tag{19}$$

where Q_t represents the final prediction result, Q_i is the prediction result of each sub neural network, C is the number of clusters, and $\bar{\lambda}_i$ is the average of the spatio-temporal correlation coefficient for the data samples in each cluster.

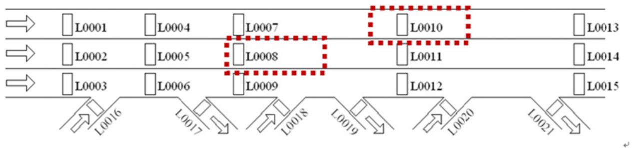


FIGURE 2. Detector position distribution.

III. CASESTUDY

A. DATA SOURCE DESCRIPTION

As shown in Figure 2, the data used in this study was derived from the Minnesota Department of Transportation (Mn/DOT) and the Traffic Data Research Laboratory (TDRL) at Duluth University, Minnesota [50]. It provides traffic flow data detected from loop detectors in the Minnesota Expressway network. The collection time is from January 1 to January 31, 2011, with an interval of five minutes. Each data sample includes detector tags, recording time, and traffic flow volume, and the data collecting time lasts 30 days, In the model evaluation, all data samples are divided into three parts, 1/3 of all data samples are set as training sets, 1/3 of the sample data is set as the verification data set, and the last 1/3 of the data is used as the testing set.

And we used the Data of the first 12 time steps as input to make predictions. In order to test the reliability and validity of the model, two detectors are selected in the cases study, shown in Fig.2, the loop detector L0010 is located at the edge of the road section and the loop detector L0008 is located at the center of the road network.

B. PARAMETERS OPTIMIZATION

In this study, two traditional indicators, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used to evaluate the prediction performance of the mode, which are shown in Equation (20) and (21). MAPE not only considers the error between the predicted value and the true value, but also the ratio between the absolute error and the true value, and the RMSE is very sensitive to the very large or small error in the results, that is, these two indicators have good effect to reflect the prediction precision of the model.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\tilde{Q}_i - Q_i\right|}{Q_i} \tag{20}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Q}_i - Q_i\right)^2} \tag{21}$$

where N is the number of data samples, $\tilde{Q}_i$ represents the predicted traffic volume from prediction models, and Q_i is the actual data collected from detectors.

In the calibration, three key parameters need to be determined in Type-2 FCM algorithm before model validation: fuzzy weighting index m, cluster number c and β. On the one hand, the m affects the convexity of the objective function; on the other hand, it controls the degree of fuzzy clustering results, that is, the degree of sample sharing in fuzzy clustering [51]. The number of clustering clusters directly affects the data recognition accuracy of clustering algorithm. However, the design parameter β affects the accuracy of the algorithm in discriminating between normal data and abnormal data, and its value affects the combination degree of the algorithm clustering. For these three parameters, over- or under-values may lead to inaccurate predictions, so we performed several tests to determine their range based on changes in prediction accuracy. The value of β ranges from 20 to 40, and the range of c is from 5 to 15, m will be selected in [1.1, 2.0] in increments of 0.1. Using RMSE as the evaluation criteria, then the parameter optimization can be performed by the following steps:

Step.1: According to the value of β ranges from 20 to 40, 21 sub-models $T_i(i = 20, 22, \ldots, 40)$ with different β are established. The corresponding values of m and c in 21 sub-models are then analyzed in optimization.

Step.2: Search the optimal values of m and c for each sub-model according to the minimal RMSE. The number of occurrences of the optimal m and c for each sub-model is counted, and the combination with the highest frequency of occurrence is labeled Um and Cu as the optimal m and c values of the total model.

Step.3: According to the combination of optimal values obtained in step (2), the corresponding RMSE value is found in the sub-model, and the optimal value of parameter β can be estimated according to the smallest RMSE.

By implementing the above parameters searching process, the optimal vales of three parameters are determined according to RMSE for two selected detectors under different prediction step ahead, shown in the Table 1.

TABLE 1. Parameter optimizing results.

No. of detector	Prediction steps	RMSE	Optimal values of three parameters		
			β	c	m
L0008	1 step	6.998	30	7	1.5
	3 step	7.630	32	7	1.5
	5 step	8.003	30	7	1.4
	10 step	8.558	24	9	1.4
L00010	1 step	6.899	27	6	1.5
	3 step	7.560	23	6	1.4
	5 step	7.995	30	6	1.5
	10 step	8.363	31	6	1.6

C. MODEL PERFORMANCE COMPARISON

In this study, we compared the prediction performance of the proposed Type-2FCMNN model to two traditional models: FCMNN and original ANN model. The FCMNN [42] is a combination of fuzzy c-means and neural networks.

Fig.3 shows prediction results of three prediction models compared with actual collecting traffic flow data. It can be found that all three candidate models can produce good prediction performance and follow the variation of observed data. We further use error distribution percentage shown in Fig.4 to compare prediction accuracy of three models. The x-axis represents the distribution of the error range and the

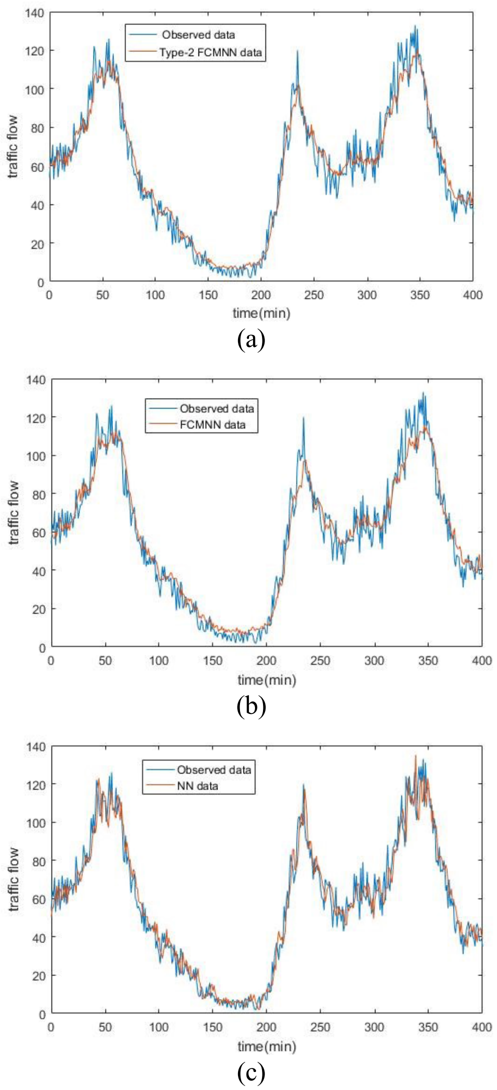


FIGURE 3. Comparison of observed data and model predictions (Time Scale: 5min). (a) Type-2 FCMNN. (b) FCMNN. (c) NN.

TABLE 2. Multi-step prediction error comparison for No. L0008 detector.

Prediction steps	RMSE			
Models	1	3	5	10
Type-2FCMNN	6.998	7.630	8.003	8.558
FCMNN	7.636	8.846	9.579	11.373
NN	7.953	9.070	9.924	12.302
RNN	7.529	8.353	9.224	10.927
LSTM	7.337	8.265	8.976	10.631
Prediction steps	MAPE			
Models	1	3	5	10
Type-2FCMNN	0.248	0.261	0.266	0.292
FCMNN	0.251	0.289	0.309	0.370
NN	0.258	0.300	0.329	0.394
RNN	0.251	0.278	0.284	0.321
LSTM	0.249	0.276	0.287	0.332

y-axis is the percentage of each range. It can be seen that the Type-2FCMNN model has better prediction performances, with the percentage of 79.5% in the range of (−6, 6) errors, higher than other two models with 71.5% for FCMNN and

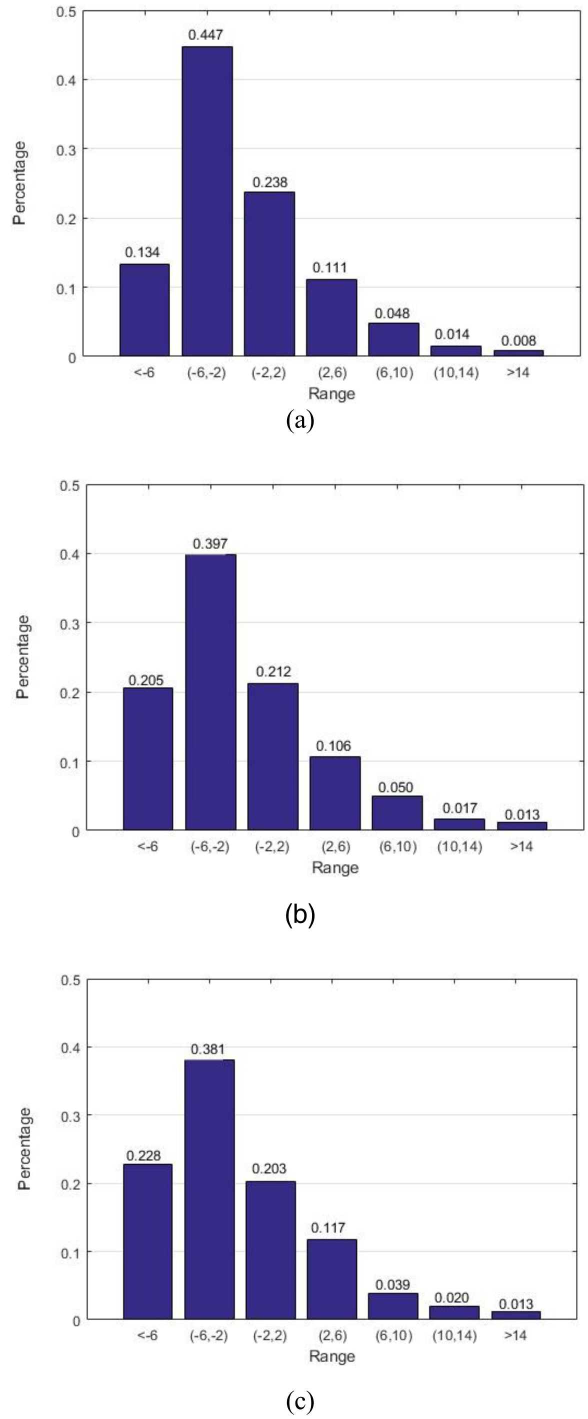


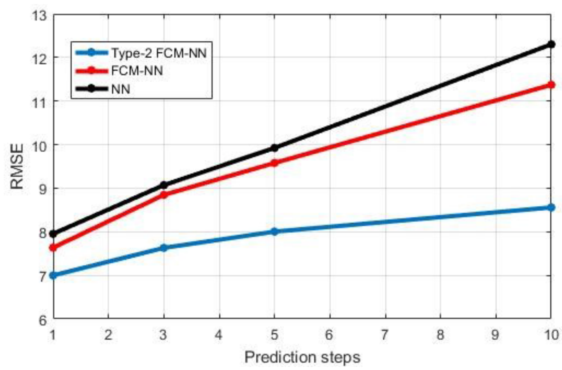
FIGURE 4. Percentage distribution of RMSE. (a) Type-2FCMNN. (b) FCMNN. (c) NN.

70.1% for NN. For the smaller error range of (−2, 2), Type-2FCMNN still has lower error distribution percentage compared with other two models. Furthermore, we can also find that the prediction accuracy of models combing with classification learning algorithm, Type-2FCMNN and FCMNN, is higher than that of individual model, NN.

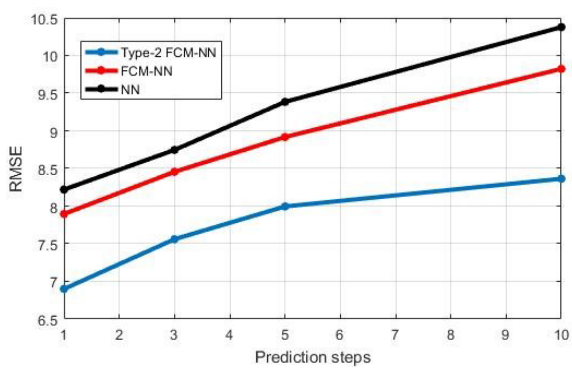
Table 2 shows the prediction performance evaluated by RMSE and MAPE of the three models under single and

TABLE 3. Multi-step prediction error comparison for No. L0010 detector.

PREDICTION STEPS / MODELS	RMSE 1	3	5	10
TYPE-2FCMNN	6.899	7.560	7.995	8.363
FCMNN	7.895	8.455	8.917	9.823
NN	8.219	8.746	9.384	10.377
RNN	7.325	8.023	8.885	9.698
LSTM	7.266	7.965	8.599	9.443
PREDICTION STEPS / MODELS	MAPE 1	3	5	10
TYPE-2FCMNN	0.302	0.318	0.336	0.361
FCMNN	0.317	0.338	0.364	0.405
NN	0.326	0.355	0.371	0.442
RNN	0.315	0.329	0.362	0.389
LSTM	0.319	0.325	0.359	0.384



(a)



(b)

FIGURE 5. Multi-step prediction error growth rate (a) No. L0008 detector (b) No. L0010 detector.

multi-step prediction ahead for the No. L0008 detector located at center lane. For single-step prediction, RMSE of FCMNN model is close to that of Type-2FCMNN model, and their performance are both better than the original NN model. When the prediction step increases, RMSE of NN model increases rapidly and reaches a maximum of 12.302 when the number of steps is 10. RMSE of Type-2FCMNN only grows to 8.558 comparing to 11.373 in the FCMNN model when prediction step is 10. Therefore, as the number of prediction step increases, the prediction errors of proposed model in this study are lower and grows slower than that of two other models, which demonstrates the superiority of Type-2FCMNN for its strong classification learning ability. For the MAPE, we can find similar comparing results. When the prediction step increases, the improved Type-2FCMNN model can still maintain higher accuracy than the other two models. The MAPE of the proposed model was 25% lower than that of NN model. Table 3 shows the prediction errors of three models for the No. L0010 detector under single and multi-step prediction ahead, also shown in the Fig.5.

IV. CONCLUSIONS

This paper introduces an improved traffic flow prediction model. Firstly, the spatial-temporal correlation is established by using the spatial position and historical traffic flow data collected in detectors on road network. Then, an improved Type-2FCMNN model is established to reasonably classify the historical data into different traffic flow patterns, and the clustering results were put into the neural network for model training. The clustering method adopt in this study is a simplified type-2 FCM algorithm including the fuzzification of membership function. Finally, the final prediction result is obtained by weighted average calculation of the preliminary prediction results of the model by the quantized spatiotemporal correlation coefficient. In the case study, traffic flow data were collected from 21 detectors on the highway, the MAPE and RMSE were introduced to evaluate the parameter optimization and prediction performance, and the proposed model was compared with the FCMNN and NN models. The following conclusions are drawn:

(1) In the road network, the data between the detectors is highly correlated. By taking the historical data of the detectors associated with the target detector as part of the input in prediction model and extracting their spatiotemporal features into the model correction, higher prediction accurate can be obtained.

(2) When using clustering method to analyze different modes of traffic flow, the perdition model combing type-2 FCM algorithm produce higher prediction accuracy than that combing original FCM clustering algorithm.

(3) Due to the ability of the type-2 FCM algorithm to distinguish outliers and categorize traffic flow state under different patterns, the proposed model express better prediction performance and a lower error growth rate than the other models in single-step and multi-step prediction.

REFERENCES

[1] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[2] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec.*, vol. 1965, no. 1, pp. 23–31, Apr. 2006.

[3] N. Huynh, H. S. Mahmassani, and H. Tavana, "Adaptive speed estimation using transfer function models for real-time dynamic traffic assignment operation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1783, no. 1, pp. 55–65, 2002.

[4] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, p. 21, 2015.

[5] A. Stathopoulos and G. M. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 11, no. 2, pp. 121–135, 2003.

[6] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, 2009.

[7] S. A. Zargari, S. Z. Siabil, A. H. Alavi, and A. H. Gandomi, "A computational intelligence-based approach for short-term traffic flow prediction," *Expert Syst.*, vol. 29, no. 2, pp. 124–142, 2012.

[8] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Phys. A, Stat. Mech. Appl.*, to be published. doi: 10.1016/j.physa.2019.03.007.

[9] D. Wang, J. Xiong, Z. Xiao, and X. Li, "Short-term traffic flow prediction based on ensemble real-time sequential extreme learning machine under non-stationary condition," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

[10] F. Zong, Y. D. Tian, Y. He, J. Tang, and J. Lv, "Trip destination prediction based on multi-day GPS data," *Phys. A, Stat. Mech. Appl.*, vol. 515, pp. 258–269, Feb. 2019.

[11] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS ONE*, vol. 10, no. 3, 2015, Art. no. e0119044.

[12] X. Jiang and H. H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *J. Transp. Eng.*, vol. 131, no. 10, pp. 771–779, Oct. 2005.

[13] M. I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Appl. Sci.*, vol. 7, no. 6, p. 476, 2017.

[14] Z. Zhao, W. H. Chen, X. M. Wu, P. C. Y. Chen, and J. M. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, 2017.

[15] 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.

[16] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. IEEE 31st Youth Acad. Annu. Conf. Chinese Assoc. Automat. (YAC)*, Nov. 2017, pp. 324–328.

[17] S. Ghosh-Dastidar and H. Adeli, "Wavelet-clustering-neural network model for freeway incident detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 18, no. 5, pp. 325–338, 2003.

[18] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Phys. A, Stat. Mech. Appl.*, vol. 477, pp. 149–160, Jul. 2017.

[19] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.

[20] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-Based system approach for modeling and predicting urban traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 16, pp. 554–573, Oct. 2008.

[21] B. Yao, P. Hu, M. Zhang, and M. Jin, "A support vector machine with the tabu search algorithm for freeway incident detection," *Int. J. Appl. Math. Comput. Sci.*, vol. 24, no. 2, pp. 397–404, 2014.

[22] C.-H. Wu, C.-C. Wei, D.-C. Su, J.-M. Ho, and M.-H. Chang, "Travel time prediction with support vector regression," in *Proc. IEEE Intell. Transp. Syst.*, vol. 2, Oct. 2003, pp. 1438–1442.

[23] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.

[24] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.

[25] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 326–334, 2010.

[26] C. Antoniou, M. Ben-Akiva, and H. N. Koutsopoulos, "Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 661–670, Dec. 2007.

[27] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, , no. 1, pp. 50–64, Jun. 2014.

[28] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[29] B. R. Chang and H. F. Tsai, "Forecast approach using neural network adaptation to support vector regression grey model and generalized auto-regressive conditional heteroscedasticity," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 925–934, 2008.

[30] B. Abdulhai, W. Recker, and H. Porwal, "Short-term traffic flow prediction using neuro-genetic algorithms," *J. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 3–41, 2002.

[31] J. Gao, Z. Leng, B. Zhang, X. Liu, and G. Cai, "Traffic flow forecasting based on wavelet neural network optimized by GA," in *Proc. IEEE Control Conf.*, Jul. 2003, pp. 8708–8712.

[32] Y. Xie and Y. Zhang, "A wavelet network model for short-term traffic volume forecasting," *J. Intell. Transp. Syst., Technol., Planning, Oper.*, vol. 10 no. 3, pp. 141–150, 2006.

[33] W.-C. Hong, Y. Dong, F. Zheng, and C.-Y. Lai, "Forecasting urban traffic flow by SVR with continuous ACO," *Appl. Math. Model.*, vol. 35, no. 3, pp. 1282–1291, 2011.

[34] W.-C. Hong, Y. Dong, F. Zheng, and S. Y. Wei, "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Appl. Math. Comput.*, vol. 217, no. 15, pp. 6733–6747, 2011.

[35] J. Tang, S. Yu, F. Liu, X. Chen, and H. Huang, "A hierarchical prediction model for lane-changes based on combination of fuzzy C-means and adaptive neural network," *Expert Syst. Appl.*, vol. 130, pp. 265–275, Sep. 2019.

[36] T. Cheng, J. Haworth, and J. Wang, "Spatio-temporal autocorrelation of road network data," *J. Geograph. Syst.*, vol. 14, no. 4, pp. 389–413, Oct. 2012.

[37] J. K. Hackney, M. Bernard, S. Bindra, and K. W. Axhausen, "Predicting road system speeds using spatial structure variables and network characteristics," *J. Geograph. Syst.*, vol. 9, no. 4, pp. 397–417, 2007.

[38] Y. Yue and A. G.-O. Yeh, "Spatiotemporal traffic-flow dependency and short-term traffic forecasting," *Environ. Planning B, Planning Des.*, vol. 35, no. 5, pp. 762–771, Oct. 2008.

[39] 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.

[40] J. Haworth and T. Cheng, "Non-parametric regression for space–time forecasting under missing data," *Comput., Environ. Urban Syst.*, vol. 36, no. 6, pp. 538–550, 2012.

[41] L. Zheng, C. Zhu, N. Zhu, T. He, N. Dong, and H. Huang, "Feature selection-based approach for urban short-term travel speed prediction," *IET Intell. Transp. Syst.*, vol. 12, pp. 474–484, Aug. 2018.

[42] K. Zhang, L. Zheng, Z. Liu, and N. Jia, "A deep learning based multitask model for network-wide traffic speed predication," *Neurocomputing*, to be published. doi: 10.1016/j.neucom.2018.10.097.

[43] L. Liu, N. Jia, L. Lin, and Z. He, "A cohesion-based heuristic feature selection for short-term traffic forecasting," *IEEE Access*, vol. 7, pp. 3383–3389, 2018.

[44] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transp. Res. C, Emerg. Technol.*, vol. 51, pp. 29–40, Feb. 2018.

[45] H. Huang, Q. Tang, and Z. Liu, "Adaptive correction forecasting approach for urban traffic flow based on fuzzy *C*-mean clustering and advanced neural network," *J. Appl. Math.*, vol. 2013, no. 1, 2013, Art. no. 195824.

[46] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.

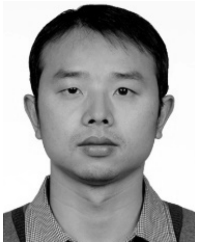
[47] C. Qiu, J. Xiao, L. Yu, L. Han, and M. N. Iqbal, "A modified interval type-2 fuzzy C-means algorithm with application in MR image segmentation," *Pattern Recognit. Lett.*, vol. 34, no. 12, pp. 1329–1338, 2013.

[48] P. Melin, O. Mendoza, and O. Castillo, "An improved method for edge detection based on interval type-2 fuzzy logic," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8527–8535, 2010.

[49] L. Lei, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.

[50] *Transportation Data Research Laboratory (TDRL) in University of Minnesota Duluth*. Accessed: 2019. [Online]. Available: http://www.d.umn.edu/tdrl/

[51] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," *Adv. Appl. Pattern Recognit.*, vol. 22, no. 1171, pp. 203–239, 1981.



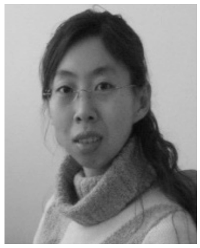
JINJUN TANG received the Ph.D. degree in transportation engineering from the Harbin Institute of Technology, Harbin, China, in 2016. He is currently an Associate Professor with the School of Traffic and Transportation Engineering, Central South University, Changsha, China. From 2014 to 2016, he was a Visiting Scholar with the Smart Transportation Applications and Research Laboratory (STAR Lab), University of Washington, Seattle, WA, USA. He published more than 65 technical papers in the journal and conference proceeding as the author or coauthor. His research interests include traffic accident analysis, traffic flow prediction, data mining in transportation systems, intelligent transportation systems, and transportation modeling.

LEXIAO LI was born in Yueyang, Hunan, China, in 1993. He received the B.S. degree in traffic engineering from Henan Polytechnic University, China, in 2015. He is currently pursuing the M.S. degree with the Transportation College, Central South University, China.

His research interests include traffic flow prediction and intelligent transportation systems.

ZHENG HU graduated from Chang'an University, in 2003. He received the master's degree with the Business Administration College, Hunan University, in 2011.

He is currently with Hunan Communication Engineering Polytechnic. His main research interests include intelligent transportation and the Internet of Things.



FANG LIU received the B.S. and M.S. degrees in traffic engineering from the Shandong University of Technology, Zibo, China, in 2006 and 2009, respectively. She is currently pursuing the Ph.D. degree with the School of Transportation Engineering, Changsha University of Science and Technology, China. Her research interests include traffic safety, and transportation modeling and simulation.