

# Short-Term Traffic Flow Prediction Method Based on Spatio-Temporal Characteristics of Complex Road Network

Huawei ZHU<sup>a</sup>, Yihao MAO<sup>b</sup>, Yiming ZHAO<sup>b</sup>, Liyuan SHI<sup>b</sup> and Yanli SHAO<sup>b,1</sup>

<sup>a</sup> *School of Information and Electrical Engineering, Zhejiang University City College, Hangzhou 310015, China*

<sup>b</sup> *Key Laboratory of Complex Systems Modeling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China*

**Abstract.** With the development of urbanization, the number of residents' motor vehicles has increased sharply, and traffic congestion problem has become increasingly serious. The construction of Intelligent Traffic System (ITS) has become the main means to alleviate traffic congestion. Short-term traffic flow prediction has guiding significance for residents' travel planning and intelligent management of transportation, and has become one of the research hotspots in intelligent transportation field. Therefore, A short-term traffic flow prediction method based on the spatio-temporal characteristics of complex road networks is proposed to further improve the prediction accuracy and reduce the prediction cost. Firstly, a graph convolutional network (GCN) capable of processing non-Euclidean data structures is used to extract the spatial characteristics of traffic flow data. Then, the long and short-term memory (LSTM) neural network is used to process the time characteristics. Finally, the two are combined to realize the effective processing of the spatio-temporal characteristics of traffic flow data. Experimental results on the real traffic flow dataset prove the feasibility and effectiveness of the proposed method, and can provide a basis for intelligent traffic control and smart city construction.

**Keywords.** Traffic Flow Prediction, Spatio-temporal, GCN, LSTM.

## 1. Introduction

With the development of urbanization, the number of residents' vehicles is increasing, and the traffic conditions of urban road network are becoming more and more complicated, which affects people's travel convenience and becomes one of the factors restricting my country's economic development. Many researchers have put forward many opinions to government departments, such as scientifically planning urban traffic routes, rationally planning urban layout, supporting public transportation, etc., and meanwhile adopting policies such as vehicle odd and even number restriction, motor vehicle restriction, license plate lottery, etc. However, the former costs a lot of time and money, while the latter treats the symptoms rather than the root cause. Recently, many

---

<sup>1</sup> Corresponding Author, Yanli SHAO, Key Laboratory of Complex Systems Modeling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China; Email: shaoyanli@hdu.edu.cn.

scholars have focused their attention on the development and improvement of intelligent transportation systems, aiming to accelerate the integrated innovation and application of modern information technologies such as big data and deep learning, and effectively improve the intelligent level of transportation systems. Efficient and accurate traffic flow prediction is an effective means to build intelligent transportation systems and alleviate traffic congestion. It can not only assist drivers in rational planning of routes, but also help transportation departments to manage scientifically, and can even be used as a basis for smart city construction.

Traditional traffic flow prediction technologies, such as statistical models and machine learning methods, have disadvantages such as high data scale and quality requirements, low prediction accuracy, and high prediction overhead. The rapid development of monitoring hardware facilities and deep learning technology has made the use of large-scale traffic data possible. The road traffic system is a complex nonlinear system with the characteristics of real-time change and human participation. The traffic flow is not only large in data scale, but also has strong temporal and spatial characteristics and uncertainty. In particular, the urban road network is a complex non-Euclidean spatial topology, which increases the difficulty of traffic flow prediction. How to deal with the complex urban road network structure and effectively extract the temporal and spatial characteristics of traffic flow data is still one of current challenges in the field of short-term traffic flow prediction.

Based on the above analysis, this study proposes a short-term traffic flow prediction algorithm based on the spatio-temporal characteristics of complex road networks. It uses Graph Convolutional Network (GCN) and Long Short-Term Memory (LSTM) to effectively process the spatio-temporal characteristics of traffic flow data under complex road networks to improve short-term traffic flow prediction accuracy. The organization structure is as follows: Section 2 reviews the related work of traffic flow prediction. Section 3 introduces the detailed implementation of the proposed method. Section 4 is related experiments and performance analysis. The conclusions and future work are given in the last section.

## **2. Related Work**

At present, there are many researchers at home and abroad studying traffic flow prediction technology, which can be roughly divided into two categories: model-driven methods and data-driven methods. Model-driven methods mainly explain the instantaneous and steady-state relationships between traffic volume, speed and density. Representative methods include queuing theory model [1], cell transmission model [2], traffic speed model [3], etc. In reality, traffic data is affected by many factors, and it is difficult to obtain an accurate prediction model to accurately describe changes in traffic data in a complex real environment. In addition, the construction of these models requires significant computing power and is susceptible to traffic interference and sampling point spacing.

Data-driven methods are the current mainstream methods, including: methods based on statistical theory, methods based on machine learning, and methods based on deep learning. Methods based on statistical theory are earlier data-driven methods used for traffic flow prediction, including Autoregressive Integrated Moving Average model (ARIMA)[4] [5], seasonal autoregressive integrated moving average (SARIMA)[6], etc. These methods have fast prediction speed but low accuracy, poor prediction performance

of non-linear traffic flow. The method based on machine learning uses machine learning algorithms to learn the nonlinear association between traffic flow and characteristics, which greatly improves the prediction performance, such as Bayesian Model [7], K-Nearest Neighbor (KNN) [8][9], etc. E.I.Vlahogianni et al. used neural networks for short-term traffic flow prediction, and used genetic algorithms for algorithm optimization [10]. Gao et al. combined KNN and SVM to deal with time series problems [11]. Jeong et al. proposed an online learning weighted SVM model for short-term traffic flow prediction, which considered the relative difference of time dependence between traffic flow data [12]. Although the above methods have improved the prediction accuracy, the current traffic flow data sets are becoming increasingly large, and it is difficult for machine learning algorithms to model large-scale data. After 2010, deep learning models have been frequently applied to traffic flow forecasting [13] [14]. RNN was first used for traffic forecasting due to its excellent time series processing capabilities, but this model has problems such as the disappearance of gradients when processing feature data with long-term time dependence. To avoid this, LSTM gradually replaced RNN to deal with time series problems [15]. Xue et al. proposed Multi-LSTM Models based on LSTM, which analyze and predict the distribution of traffic flow in different states, and can better deal with the complexity and diversity of the time characteristics of traffic flow data [16]. These models take into account the time characteristics of traffic flow data and improve the prediction accuracy to a certain extent. Later, more and more studies have shown that traffic flow data in adjacent areas will also have an impact on the traffic flow in the current area for a period of time. People began to consider introducing spatial features to further improve the prediction accuracy. Chen et al. proposed an FDCN model based on fuzzy theory and convolutional neural network [17], which can extract the spatial characteristics of traffic flow data and achieved good results. After that, An et al. used CNN to extract the spatial features in the state of traffic accidents, and used them to improve the prediction effect under special circumstances [18]. Yu et al. also tried to use 3DCNN to replace the traditional convolutional neural network to process traffic flow data with spatiotemporal feature fusion, and achieved good results [19]. However, the CNN model has its own shortcomings, such as not being able to effectively process non-Euclidean traffic flow data. Therefore, GCN models suitable for non-Euclidean spatial structures have begun to be applied in the field of traffic flow prediction to better describe the spatial characteristics of traffic flow [20].

The above research has laid the foundation for short-term traffic flow prediction. However, how to effectively fit the temporal and spatial characteristics of traffic flow data under complex road networks and improve the accuracy of short-term traffic flow prediction is still a challenging problem.

### **3. Traffic Flow Prediction Method**

#### *3.1. Method Overview*

Short-term traffic flow prediction is to predict the future traffic conditions of a certain area based on historical traffic flow data. However, the large scale of traffic flow data, complex factor characteristics, and strong temporal and spatial correlation often make it difficult to make accurate predictions. Therefore, this study proposes a short-term traffic flow prediction method based on the spatio-temporal characteristics of complex road networks to deeply explore the potential value of traffic big data to support the

construction of intelligent transportation systems. The overall method steps are as follows:

- (1) Data preprocessing. The traffic flow data information of each area of the city in the historical time period is combined into a two-dimensional matrix according to the latitude and longitude information, and the spatial information and time information are retained, so that the model can fit the spatial dependence and time dependence of the traffic flow data.
- (2) Spatial dependency modeling based on GCN model. Non-Euclidean spatial features are one of the most important features in traffic flow data. Here, GCN is used to effectively fit the spatial dependence of complex traffic network.
- (3) Time-dependent modeling based on LSTM model. Use LSTM to capture time series correlation to effectively deal with the periodic time characteristics and sudden characteristics of traffic flow.

### 3.2. Data Preprocessing

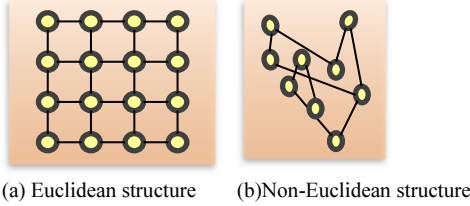
This study divides the entire city into  $a \times b$  grid map, which contains  $n$  areas ( $n = a \times b$ ), and uses  $\{1, 2, \dots, n\}$  to represent them. Divide the entire time period (for example, one month) into  $m$  continuous time intervals of equal length. The start/end traffic volume of an area is defined as the traffic volume departing/arriving in the area within a fixed time interval. In form,  $Y_i, T_s$  and  $Y_i, T_e$  represent the start/end flow of area  $i$  in the  $t$ -th time interval. The collection of individual trips forms a traffic flow, describing the time-enhanced movement between specific areas. Formally, the traffic flow starting from area  $i$  in the time interval is marked as  $f_{i,t,j}$  at the end of area  $j$  in the time interval  $t$ . Traffic flow reflects regional connectivity and the mobility of individual transmission.

Currently, the real and effective traffic set often has certain uncertainties and deficiencies, which need to be preprocessed. Due to subjective or objective reasons such as extreme weather, the failure of traffic flow monitoring hardware and software, etc., a considerable proportion of missing traffic flow data exists. If the model is directly trained with missing data, it will bias the model training and reduce the prediction accuracy. Missing value completion is very important. In this study, the types of missing values in the traffic flow data are missing at random. As mentioned earlier, traffic flow data has a strong time correlation and periodic characteristics. The data surrounding the missing value data and the data at the same position in the cycle have a very large theoretical correlation with the missing data. Therefore, the possible value interpolation method of missing values is used here, and the adjacent data is used for an average, and then the data at the same position in the adjacent period is used for a certain weighting process to comprehensively obtain the missing values of the data.

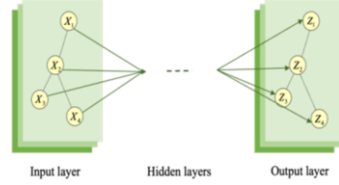
### 3.3. Spatial Dependency Modeling Based on Gcn Model

In the past, CNNs were mostly used for traffic flow prediction, mainly dealing with the Euclidean structure with neat data and a stable number of neighbor nodes. However, the topological structure of traffic flow is a complex non-Euclidean structure data. Different traffic nodes have different traffic connection methods, and the number of adjacent nodes of each node is uncertain, as shown in figure 1. GCN model can perform end-to-end learning of graph node features and graph structure, and is one of the best methods for processing and modeling non-Euclidean structure graph data. The GCN model can encode the characteristics of traffic topology by extracting the structural relationship

between the traffic node in the center and its neighbor nodes to effectively extract the spatial features of traffic flow data. Based on the above analysis, the GCN structure is shown in figure 2, and the steps of spatial dependency modeling based on GCN model are as follows: Firstly, transform the characteristic information of each graph node and input it into neighbor nodes; Secondly, aggregate the characteristic information of each neighbor node into that node; Finally, aggregate the previous information and make non-linear changes to strengthen the performance capabilities of the characteristics.



**Figure 1.** Comparison of different structure data



**Figure 2.** Graph Convolutional Network

Suppose that the traffic flow data has  $N$  traffic areas, that is, traffic nodes. The characteristics of each traffic node are different and have strong own characteristics. GCN turns the feature combination of traffic nodes into a feature matrix  $X$ , whose dimension is  $N \times D$ , and then the connection between each traffic flow node will also be combined into a matrix  $A$ , whose dimension is  $N \times N$  dimension. This is the adjacency matrix.  $A$  and  $X$  are the inputs of GCN. The propagation from layer to layer of the GCN neural network is as equation (1), where  $H^{(l)}$  refers to the input feature of the  $l$ -th layer, and  $H^{(l+1)}$  naturally refers to the output feature.  $W^{(l)}$  is the linear transformation matrix and  $\sigma(\cdot)$  is a nonlinear activation function.  $\tilde{A}$  is called adjacency matrix with self-connection,  $\tilde{D}$  is corresponding degree matrix.

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (1)$$

A 2-layer GCN model is used to learn spatial features from the traffic data, as shown in equation (2).

$$f(X, A) = \sigma(\hat{A} \text{Relu}(\hat{A} X W_0) W_1) \quad (2)$$

Where  $X$  represents feature matrix,  $A$  represents the adjacency matrix,  $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$  denotes the preprocessing step.  $W_0$  and  $W_1$  represent the weight matrix in the first and second layers, and  $\text{Relu}()$  represents an activation function.

### 3.4. Time-Dependent Modeling Based on Lstm Model

The LSTM model is an upgraded model on the RNN, which solves the problems of gradient explosion and gradient disappearance caused by the large time span of RNN. The LSTM model has four neural network layers that interact with each other. As shown in figure 3. The LSTM network has memory and forgetting properties. It can forget unimportant features in long-term traffic sequence training, highlight the required time features, and effectively extract the time features of traffic flow. Here, the LSTM model is combined with the GCN model, so that the temporal and spatial features of the traffic data can be better extracted. The specific calculation process of the new LSTM model is as follows.

$$i_t = i(t) = \sigma(W^{(i)} f(x^{(t)}) + U^{(i)} h^{(t-1)}) \quad (3)$$

$$f_t = f(t) = \sigma(W^{(f)} f(x^{(t)}) + U^{(f)} h^{(t-1)}) \quad (4)$$

$$o_t = o(t) = \sigma(W^{(o)}f(x^{(t)}) + U^{(o)}h^{(t-1)}) \quad (5)$$

$$\tilde{c}_t = \tilde{c}(t) = \tanh(W^c f(x^{(t)}) + U^c h^{(t-1)}) \quad (6)$$

$$c_t = c(t) = f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t-1)} \quad (7)$$

$$h_t = h(t) = o^{(t)} \circ \tanh(c^{(t-1)}) \quad (8)$$

Wherein  $i(t)$ ,  $f(t)$  and  $o(t)$  represents input Gate, Forget Gate and Output Gate respectively.  $\tilde{c}(t)$  and  $c(t)$  are cell state at the beginning and at the end.  $h(t)$  is the final output.

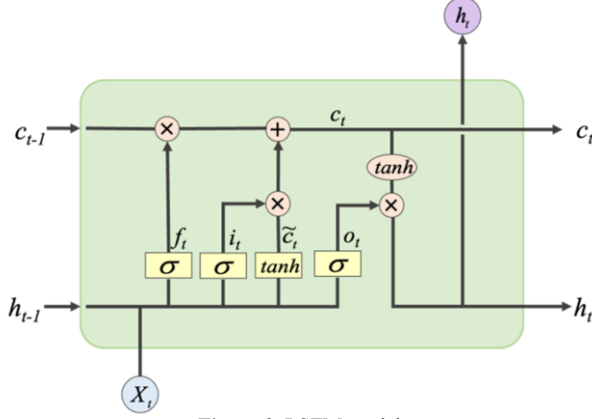


Figure 3. LSTM model

### 3.5. Model Training

This study uses Adam optimization algorithm for model training to build a traffic flow prediction model, which can dynamically adjust the learning rate and are better than other gradient descent algorithms. Other optimization algorithms may have forgotten problems, which may cause model performance degradation. Limited by the used hardware devices, some parameter settings are relatively simple. The number of hidden units is a very important parameter in the model, because different hidden units will greatly affect the prediction accuracy. After testing, the number of hidden units in the proposed model is set to 64. In addition, the learning rate is 0.001, the training batch size is 64, and the number of training epochs is 3000. The loss function used in the model training process is shown in equation (9). Among them,  $Y_t$  and  $Y_{bt}$  are used to represent the actual traffic flow and the predicted flow, respectively. The first item is used to reduce the error between the actual traffic flow and the predicted flow. The second term  $L_{reg}$  is L2 regularization, and the term  $\lambda$  is a hyperparameter that helps avoid overfitting problems.

$$loss = ||Y_t - Y_{bt}|| + \lambda L_{reg} \quad (9)$$

## 4. Experiments and Performance Analysis

### 4.1. Experiment Setting

We implemented our model in Python with PyCharm 21 library. The experiments are run on a Windows 10 platform with an Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz, Nvidia GeForce GTX 1060 GPU, 6GB and 8GB memory. The experimental dataset is

NYC-taxi dataset which contains 2,349,490 taxi trip records in New York City from January 1, 2015 to March 1, 2015 (60 days). The data from January 1, 2015 to February 8, 2015 (40 days) are used as training data, and the rest are test data. We divide the entire city into 20×20 regions, each with an area of about 1km×1km, and use the minimum and maximum normalization to convert the traffic volume and flow to the [0,1] scale. The processed experimental data includes: one is a 20\*20 adjacency matrix to describe the spatial relationship between roads, each row represents a road, and the values in the matrix represent the connectivity between roads. The other is the feature matrix to describe the change of speed on each road over time. Each row represents a road, and each column is the traffic speed on the road at different times.

Two commonly used evaluation indicators are used to evaluate the prediction performance of each model on traffic flow prediction task. Suppose that  $n$  is the total number of test samples,  $y_i$  and  $\hat{y}_i$  are the observed value and the predictive value of the  $i$ -th test sample, respectively.

(1) Root mean square error (RMSE):

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

(2) Mean absolute percentage error (MAPE):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{11}$$

4.2. Performance Analysis

4.2.1. Overall Performance Analysis.

The first group is the overall performance verification. This experiment compares the prediction performance of this method with other mainstream methods to verify the feasibility and superiority of the proposed method. Comparison models include: (1) ARIMA [21] (2) Logistic Regression (LR) [22] (3) eXtreme Gradient Boosting (XGBoost) [23] (4) Multi-Layer Perceptron (MLP) [24] (5) Convolution LSTM (ConvLSTM) [25] (6) Deep Spatio-Temporal Residual Networks (ST-ResNet) [26]. The experimental results are given in table 1. According to the above experimental results, the following conclusions can be drawn:

Table 1. Overall performance comparison experiment results

Traffic flow prediction Models	Start		End	
	RMSE	MAPE	RMSE	MAPE
ARIMA [21]	36.53	22.21%	27.25	20.91%
LR [22]	28.51	19.94%	24.38	20.07%
XGBoost [23]	26.07	19.35%	21.72	18.70%
MLP [24]	26.676	18.43%	22.08	18.31%
ConvLSTM [25]	28.13	20.50%	23.67	20.70%
ST-ResNet [26]	26.23	21.13%	21.63	21.09%
OurModel	25.12	17.41%	20.05	<b>17.31%</b>

(1) ARIMA, the traditional time series prediction method, performs poorly since it pays more attention to the time characteristics and ignores spatial and other contextual characteristics. Compared with ARIMA algorithm, regression-based methods including LR and XGBoost algorithms have achieved better performance, i.e, the RMSE and MAPE of XGBoost algorithm reduce  $20.3\% = (27.25-21.72)/27.25$  and  $2.21\% = 20.91\% - 18.70\%$ , respectively, However, they fail to capture the complex nonlinear time correlation and dynamic spatial relationships.

(2) For NN-based methods, because MLP does not explicitly model spatial-temporal dependence, many traffic flow data features are not fully utilized, resulting in relatively low traffic flow prediction accuracy. The RMSE value of MLP is  $9.2\% = (22.08-20.05)/22.08$  lower than our method. ConvLSTM obtains space and time information by integrating the convolution operation into the LSTM unit, and designs two decoding branches. This processing method enables the model to process time information as well as LSTM, and it can also process part of spatial information. However, its ability to process spatial information is relatively limited, and it is not as good as the proposed model proposed that can make better use of all spatial information.

(3) ST-ResNet uses a residual network based on convolution to reduce the impact of network depth. It inputs a time segment into three ResNet branches to extract time characteristics, and all nodes are spatially modeled with each other. This structure is conducive to optimizing the spatial dependence relationship, but ignores the time series dependence. The RMSE and MAPE values of our model are 20.05 and 17.31%, respectively, which is also better than the ST-ResNet model.

(4) The proposed model is lower than the above models in RMSE and MAPE indicators, and the overall performance is the best. The reason is that it effectively combines GCN and LSTM, so that the temporal and spatial features of traffic flow data are fully utilized to further improve prediction performance.

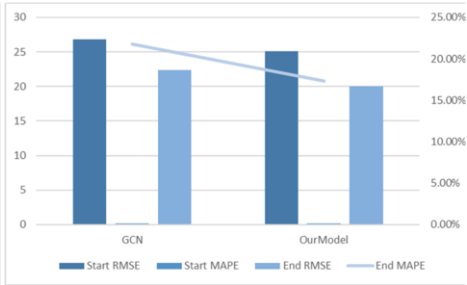
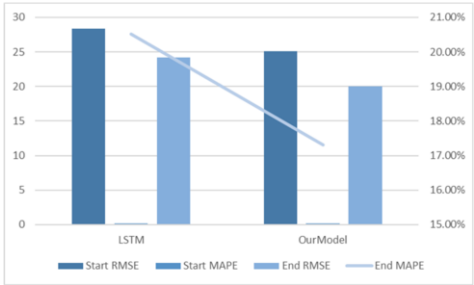
#### 4.2.2. Performance Improvement Comparison Analysis.

We performed two groups performance improvement verification experiments, and the results are shown in the table 2, figure 4 and figure 5.

**Table 2.** Comparison results of spatial dependence improvement experiments.

Method	Start		End	
	RMSE	MAPE	RMSE	MAPE
LSTM	28.31	21.06%	24.14	20.51%
GCN	26.81	21.56%	22.36	21.77%
OurModel	25.12	17.41%	20.05	17.31%





**Figure 4.** Spatial dependency improvement results. **Figure 5.** Temporal dependency improvement results.

Wherein LSTM is the model remove the spatial dependence model GCN, while GCN is the model remove the temporal dependence model LSTM. It can be seen that after adding and improving the spatial dependency modeling GCN network model, the model's ability to utilize spatial features is greatly improved, and RMSE and MAPE have reduced  $16.94\% = (24.14-20.05)/24.14$  and  $3.2\% = 20.51\%-17.31\%$ , respectively. Therefore, the GCN model is more suitable for non-Euclidean data structures such as traffic flow data under complex road networks, and can better model the spatial dependence to improve prediction accuracy. Similarly, it can be seen that after adding the LSTM model, the RMSE and MAPE reduce  $10.33\% = (22.36-20.05)/22.36$  and  $4.46\% = 21.77\%-17.31\%$  respectively. The reason is that the LSTM model can not only effectively model the time characteristics, but also alleviate gradient explosion and gradient disappearance problem, thereby improving the prediction performance.

5. Conclusions and Future Work

In order to support the construction of intelligent transportation systems and smart city planning, this study proposes a short-term traffic flow prediction algorithm based on complex road network traffic characteristics. A new combined model structure is proposed to use the advantages of both GCN and LSTM models to further improve the prediction accuracy. On the one hand, the GCN model, which is more suitable for processing complex road network non-Euclidean data structures, is used to better process the spatial characteristics of traffic flow data. On the other hand, the LSTM algorithm is applied to better fit traffic flow data with a long time span. The two models are combined to effectively process the temporal and spatial characteristics of traffic data, and further improve the prediction performance under complex road network conditions. The experimental results on the real experimental data set NYC verify the feasibility and superiority of the method in this study. However, in traffic flow prediction task, there are not only simple temporal and spatial effects, but also regional differences and social events. At present, researchers tend to focus on the easily available data such as date, time, speed, etc., and there is less research on factors such as holidays, traffic control policies, and public events that have a significant impact. How to consider the impact of these factors is a problem that we should tackle in the future.

## Acknowledgments

The authors appreciate the support from the Zhejiang Provincial Natural Science Foundation of China (LY20F020015, LY21F020015) and the National Science Foundation of China (61902345 and 61972121) and the Defense Industrial Technology Development Program (No. JCKY2019415C001), and the Open Project Program of the State Key Lab of CAD&CG (Grant No.2109), Zhejiang University.

## References

- [1] Xu X Y, Liu J, Li H Y, Hu J Q 2014 Analysis of subway station capacity with the use of queueing theory *Transportation Research Part C* **38**.
- [2] Wei P, Cao Y, D Sun 2013 Total unimodularity and decomposition method for large-scale air traffic cell transmission model *Transportation Research Part B Methodological* **53** 1–16.
- [3] Qi W, Li L, Hu J et al 2011 Traffic velocity distributions for different spacings *Qinghua Daxue Xuebao/Journal of Tsinghua University* **51**(3) 309-12.
- [4] Hamed Mohammad M, Al-Masaed Hashem R, Bani Said Zahi M 1995 Short-Term Prediction of Traffic Volume in Urban Arterials *Journal of Transportation Engineering* **121**(3) 249-54
- [5] Voort M V D, Dougherty M and Watson S 1996 Combining kohonen maps with arima time series models to forecast traffic flow *Transportation Research Part C* **4**(5) 307-18.
- [6] Williams B M and Hoel L A 2003 Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results *Journal of Transportation Engineering* **129**(6) 664-72.
- [7] Sun S, Zhang C and Yu G 2006 A bayesian network approach to traffic flow forecasting *IEEE Transactions on Intelligent Transportation Systems* **7**(1):124-32.
- [8] Zhang X L, He G G, Lu H P 2009 Short-term traffic flow forecasting based on K-nearest neighbors non-parametric regression *Journal of Systems Engineering*
- [9] Chang H, Lee Y, Yoon B et al 2012 Dynamic near-term traffic flow prediction: systemoriented approach based on past experiences *Intelligent Transport Systems Iet* **6**(3) 292-305.
- [10] Vlahogianni E I, Karlaftis M G and Golias J C 2005 Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach *Transportation research. Part C, Emerging Technologies* **13**(3) 211-34.
- [11] Gao H H, Kuang L, Yin Y Y, Guo B, Dou K 2020 Mining Consuming Behaviors with Temporal Evolution for Personalized Recommendation in Mobile Marketing Apps *ACM/Springer Mobile Networks and Applications (MONET)* **25**(4) 1233–48.
- [12] Jeong Y S, Byon Y J, Manoel Mendonca C N and Said M E 2013 Supervised weighting-online learning algorithm for short-term traffic flow prediction *IEEE Transactions on Intelligent Transportation Systems* **14** 1700-7.
- [13] Lv Y, Duan Y, Kang W, Li Z and Wang F Y 2015 Traffic flow prediction with big data: a deep learning approach *IEEE Transactions on Intelligent Transportation Systems* **16**(2) 865-73.
- [14] Chen Y, Lei S and Lei W 2017 Traffic Flow Prediction with Big Data: A Deep Learning based Time Series Model IEEE INFOCOM 2017 -IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE
- [15] Rui F, Zuo, Z and Li L 2016 Using LSTM and GRU neural network methods for traffic flow prediction 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE.
- [16] Xue Z and Xue Y 2018 Multi long-short term memory models for short term traffic flow prediction *IEICE Transactions on Information and Systems* E101.D(12) 3272-5.
- [17] Chen W, An J, Li R, Fu L, Xie G, Bhuiyan M Z A and Li K 2018 A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features *Future Generation Computer Systems*
- [18] An J, Fu L, Hu M, Chen W and Zhan J 2019 A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information *IEEE Access*, 1-1.
- [19] Yu F, Wei D, Zhang S, Shao Y 2019 3D CNN-based Accurate Prediction for Large-scale Traffic Flow *4th International Conference on Intelligent Transportation Engineering (ICITE 2019)*.
- [20] Youngjoo S, Michael D, Pierre V et al 2016 Structured Sequence Modeling with Graph Convolutional Recurrent Networks[OL]. arXiv:1612.07659, <https://arxiv.org/abs/1612.07659>.
- [21] Wang Y 2017 Traffic Flow Prediction Based on Combined Model of ARIMA and RBF Neural Network *International Conference on Machinery*.

- [22] Li C and Xu P 2020 Application on traffic flow prediction of machine learning in intelligent transportation *Neural Computing and Applications*(11).
- [23] Chen T and Guestrin C 2016 XGBoost: A Scalable Tree Boosting System. *The 22nd ACM SIGKDD International Conference. ACM.*
- [24] Abdi, Javad, Moshiri and Behzad 2015 Application of temporal difference learning rules in short-term traffic flow prediction. *Expert systems: The international journal of knowledge engineering.*
- [25] Zhao Y B 2020 Short-term traffic flow prediction method based on Conv-LSTM *The 15th China Intelligent Transportation Annual Conference.*
- [26] Zhang J, Zheng Y, Qi D 2017 Deep spatio-temporal residual networks for citywide crowd flows prediction.