ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 16, No. 1, 2021, *pp*.016-023

A short-time traffic flow prediction model based on TCN-LSTM with causal convolutional layer

Chenmin Ma
Nanjing University of Information Science and Technology, Nanjing-210044, China
(Received January 04, 2021, accepted February 18, 2021)

Abstract. Short term traffic flow prediction is of great significance for traffic management, early guidance and dispersion, avoiding congestion and improving safety. Traffic flow is affected by multiple factors and exists strong time dependency. Therefore, it is very meaningful to establish a prediction model with multiple features, such as time-day-week-biweekly, holiday and weather situation, etc. In this paper, a TCN- LSTM model with causal convolution block is proposed, which is composed of two subnets: LSTM subnetwork is used to extract feature from original traffic flow data sequence, three TCN+LSTM subnetworks are used to extract features from traffic flow data with day-week-biweekly, holiday and weather. TCN is embedded to maintain causation of the input traffic flow data. Finally, features extracted from the two sub networks are merged and imported into top-level full connection network. The prediction sequence of the future short-term traffic flow is obtained at the output layer of FCN. Experimental results show that the proposed TCN-LSTM model has high accuracy and stability in short-term traffic flow prediction.

Keywords: Short-term traffic flow forecast; Convolutional neural network; Long and short term memory neural network; Causal convolutional layer

1. Introduction

In recent years, with the continuous growth of various types of vehicles in urban and rural areas, highway traffic flow is also growing rapidly, traffic congestion is serious and major casualties occur frequently. Scientific and reasonable scheduling, especially during holidays, the traffic flow problem has been the call of the whole society. In order to relieve the traffic pressure, the construction of traffic infrastructure can only alleviate the traffic pressure in a period of time. While, improving the efficiency of traffic management, making rational use of the road network and dredging ahead of time are the most effective ways to reduce traffic congestion and improve traffic safety [1]. Short-term traffic forecasting is a process of forecasting the future short-term traffic flow condition directly according to the continuous feedback of traffic information. In order to solve the problem of short-term traffic prediction, people have done a lot of different modeling work.

Traditional analysis models based on time series include: autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model. ARIMA model can eliminate the short-term fluctuations in time series, so as to better capture the long-term characteristics of time series. However, ARIMA model needs to assume that the time series are linearly correlated, and the traffic flow data is a complex and changeable non simple linear series. When the traffic conditions change dramatically, the model has obvious shortcomings in forecasting delay. In order to solve the shortcomings of ARIMA, Okutani and Stephanedes applied Kalman filtering theory to the dynamic prediction of short-term traffic flow [3], Kalman filtering model (KFM) is a state space model based on the Kalman filter theory. The equations are updated by constantly adding new sample data to make real-time prediction. KFM model can deal with both stationary and non-stationary data, and has the advantages of more flexible prediction factors and high prediction accuracy [4]. In addition, k-nearest neighbor algorithm (k-NN), wavelet decomposition and reconstruction (WDR) are popular traffic flow prediction models. Due to the randomness of traffic flow and the highly nonlinear characteristics of short-term prediction, artificial intelligence technology as an alternative method of traffic flow prediction model has been widely concerned in recent years. Smith et al. used BP neural network to predict traffic flow earlier [5], Hu et al. input historical traffic flow data into BP neural network after difference, and then predict short-term traffic flow at the next moment [6], Chan et al. (2012) proposed a new neural network training method, which used mixed exponential smoothing method

(EXP) and Levenberg method Marquardt (LM) algorithm aims to improve the generalization ability of previous methods used to train NNs for short-term traffic flow prediction [7].

In recent years, deep learning (DL) has been proved to be an effective method to extract data features, and its performance has been confirmed in many fields such as image recognition, video processing and natural language processing [13,14]. Different from the traditional ML algorithm, DL model can accept the input data in the original format and automatically discover the required features step by step. This technology is called end-to-end learning. DL greatly simplifies the process of feature engineering and improves the quality of feature [15]. Long and short term memory (LSTM) is a special deep neural network (RNN), which is suitable for modeling the dynamic time dependence in time series. Therefore, it has been proved that the prediction accuracy of LSTM model is much higher than that of traditional prediction methods [16,17]. Tian et al. discussed the performance of LSTM recurrent neural network in predicting short-term traffic flow, and compared it with several other commonly used models [18]. Fu et al. used LSTM and GRU neural network models to predict short-term traffic flow [19]. Convolutional neural networks (CNNs) are specially designed for data domain with conventional grids. They can directly identify the spatial dependencies between meshes, utilize various localized filters or kernels, and automatically learn these shift invariant kernels from the data. Based on CNN, Zhang et al. used multi-channel deep convolution neural network to classify multivariate time series [20]. Wu et al. combined CNN and LSTM, proposed a hybrid CLTFP model that can capture spatiotemporal correlation [21].

However, by studying the existing short-term traffic flow prediction models, it was found that they seldom consider the external conditions, such as weather, holidays and so on. These factors are also the main that people consider before going out and naturally needed to be considered in the road traffic. In this paper, we propose a TCN-LSTM model embedded with causal convolution layer module. The main intention is: 1) TCN can better capture the causality of time series data, and better solve the defect that CNN can't maintain the time sequence after feature extraction; 2) Combined with the influence of multi-timelines, holiday, weather and other features on traffic flow, a multi-subnetwork is established to extract features. Then, the fused multiple features are used as the input the FCN to generate the traffic flow. Experiments show that the proposed TCN-LSTM model greatly improves the accuracy of the model.

2. Related work

2.1 Long and short term memory neural network (LSTM)

Long and short term memory model (LSTM) is a special RNN structure, was proposed by Hochreiter and Schmidhuber in 1997. The main innovation of LSTM is its storage cell, which is essentially a calculator for entering information. The cell is controlled by several automatic parameters to access, write and clear. When a new input x_t appears, if the input gate z^i is activated, its information will be accumulated into the cell. At the same time, if the forget gate z^f is opened, the past cell state c_{t-1} may be forgotten in the process. Whether the latest cell output $c^{'}$ will propagate to the output state needs to further controlled by the output gate z^o. One of advantages of the LSTM model using memory cells and gates to control information flow is to prevent the gradient disappearing too quickly. Figure 1 shows the internal structure of LSTM model. The main calculation formulas of LSTM model are as follows (1) - (5):

$$f_t = \sigma\left(W_{x_f}x_t + W_{h_f}h_{t-1} + W_{c_f} \cdot C_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_{x_i}x_t + W_{h_i}h_{t-1} + W_{c_i} \cdot C_{t-1} + b_i\right) \tag{2}$$

$$C^{'} = f_t \cdot C_{t-1} + i_t \cdot g_t \tag{3}$$

$$O_t = \sigma\left(W_{x_o}x_t + W_{h_o}h_{t-1} + W_{c_o} \cdot C_t + b_o\right) \tag{4}$$

$$h_t = O_t \cdot tanh(C^{'}) \tag{5}$$

Where $\sigma(\cdot)$ is the sigmoid activation function, W_{x_f}, W_{h_f} and W_{c_f} represents the gate weight of each unit.

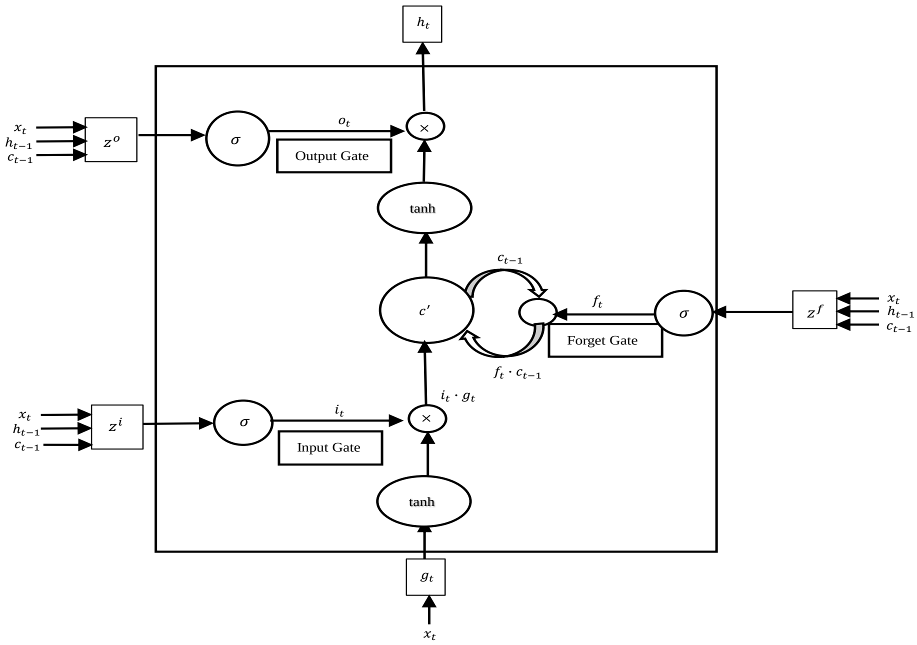


Figure.1 LSTM model flow diagram

2.2 Temporal convolution network (TCN)

Temporal convolution network (TCN), is a one-dimensional CNN [25]. Different from the general one-dimensional CNN, the advantage of TCN is that it can keep the order of the input data in the time axis direction. When using TCN module to deal with time series, the data sequence before the current time needs to be preprocessed properly to generate labeled time series pairs. Figure 2 shows the forward propagation process of a time series in the TCN module, in which the size of the TCN convolution kernel is 2, after a convolution operation, the size of the data sequence will be reduced by half. As shown in the figure, TCN strengthens the causal relationship between the current value and the input value at the previous time. The process of TCN processing data sequence can be regarded as a process of dilated causal convolution. By stacking more convolution layers, using larger dilation coefficient and increasing the size of filter, the size of perception field of TCN network can be changed, so as to better control the memory length of the model. Similar to CNN based image classification or object recognition model, padding method is also used in TCN module to keep the invariance of feature size of data sequence. Because TCN can better maintain the sequence relationship between historical data, compared with the disorder of CNN, the data sequence features in the extraction process, in the time axis direction to maintain the consistency of data feature time series, can greatly improve the accuracy of model prediction.

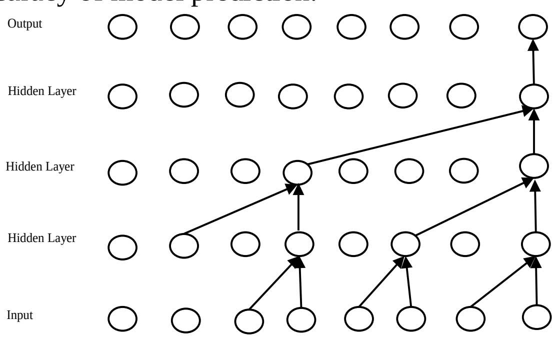


Figure.2 Temporal Convolution Network basic structure

In addition, TCN has the advantages of short network training and verification time, low memory overhead, large-scale data parallel processing and improving network operation speed; its backpropagation path and time direction of sequence are different, which can avoid the problem of gradient explosion or gradient disappearance [26]. Figure 3 shows the flow chart of TCN.

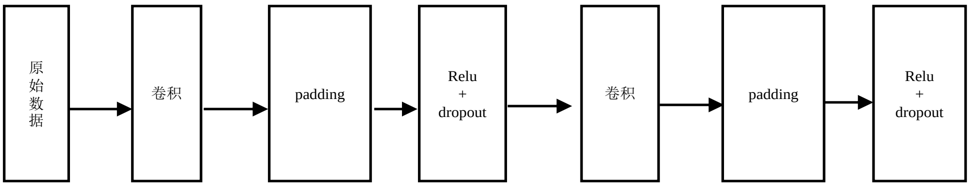


Figure.3 Temporal Convolution Network diagram

3. TCN-LSTM model

In this paper, a new TCN-LSTM model is constructed. The model aims to use the historical data set of traffic flow to obtain the internal and external characteristics of traffic flow through multiple timelines, holiday and weather characteristics, among which the external factors include temperature, holiday and weather conditions. The TCN-LSTM subnet is built by embedding TCN module, which can maintain the time sequence of traffic flow data, so that the output results of each time axis after convolution processing can ensure the consistency, thus ensuring the prediction accuracy of the TCN-LSTM model.

3.1 Model structure and parameter description

The model structure includes two subnets: (1) LSTM subnet: input time-traffic flow data, which is directly processed by LSTM model; (2) CRNN subnet: It is a combination of TCN and LSTM models, Traffic flow data were re-labeled and grouped according to three timelines and external features, Daily, Weekly and Biweekly, respectively, and input into three CRNN models to extract multiple features of traffic flow data. Finally, the feature fusion uses the full connection layer to output the prediction results. The structure of the proposed TCN-LSTM prediction model is shown in Figure 4.

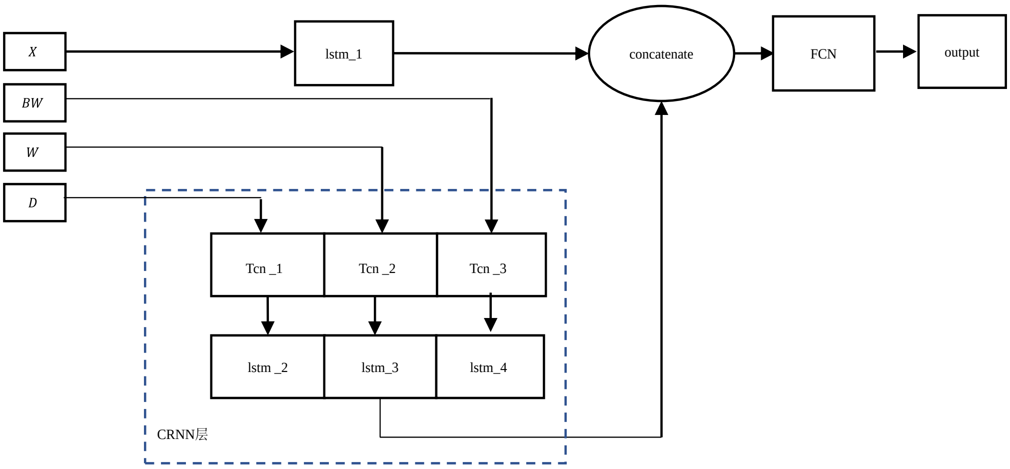


Figure.4 TCN-LSTM model structure chart

The parameters of each layer of the TCN-LSTM model are shown in Table 1:

Table 1 model parameter description

Layer		feature	step size
LSTM_1: LSTM		(None,13622)	1
CRNN_1	Tcn_1	(None,13622)	1
	LSTM_2	(None,13622)	2
CRNN_2	Tcn_2	(None,7,1946)	1
	LSTM_3	(None,7,1946)	3
CRNN_3	Tcn_3	(None,14,973)	1
	LSTM_4	(None,14,973)	4

3.2 Model processing flow

Figure 5 shows the processing flow of TCN-LSTM model, which is described as follows:

(1) Collect traffic flow data, preprocess traffic flow data, divide training set and test set at the same time;

(2) Extract the characteristics of the original traffic flow data: input the original time-traffic flow data into the LSTM subnet, input the test set in the test and forecast stage and evaluate the forecast accuracy of the test set;

(3) Feature extraction of multi-time traffic flow data: input "D-W-BW" data training set in the training stage to train the parameters of CRNN subnet; input other external factor data test set in the test and forecast stage to evaluate the forecast accuracy;

(4) Fusion of multi-time and external feature: the features from LSTM and CRNN subnet are fused, and the fused feature data is input to the FCN. After the nonlinear transformation of FCN with two hidden layers, the prediction sequence is obtained at its output end.

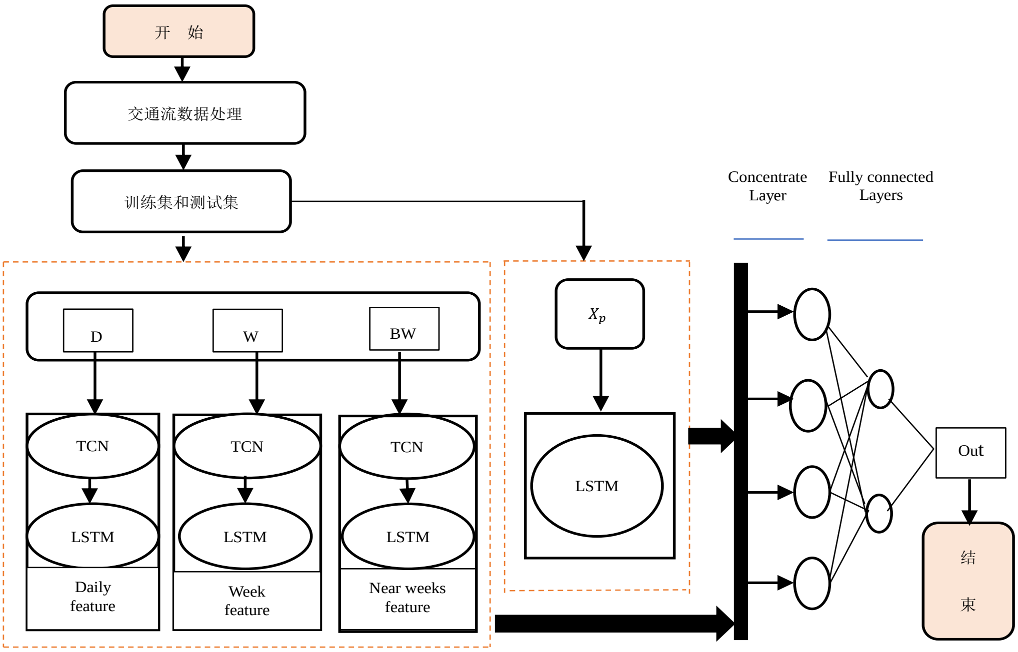


Figure.5 TCN-LSTM Model flow diagram

4. Experimental evaluation and analysis

4.1 Data set preprocessing

The traffic flow prediction problem in this paper can be described as follows: Let the traffic flow sequence before the known current time t is $\{y_{t-N}, y_{t-(N-1)}, \cdots, y_{t-1}, y_t\}$, and forecast the traffic flow at a certain time in the future is $\{y_{t+1}, y_{t+2}, \cdots, y_{t+M-1}, y_{t+M}\}$. The model in this paper uses data of 13,622 moments as the current input. The traffic flow data set is preprocessed on time, daily, weekly and biweekly, as shown in Equations (6) and (7):

$$x_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{13622} \end{bmatrix} \tag{6}$$

$$D = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{13622} \end{bmatrix} \quad W = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,1946} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,1946} \\ \vdots & \vdots & \ddots & \vdots \\ x_{7,1} & x_{7,2} & \cdots & x_{7,1946} \end{bmatrix} \quad BW = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,973} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,973} \\ \vdots & \vdots & \ddots & \vdots \\ x_{14,1} & x_{14,2} & \cdots & x_{14,973} \end{bmatrix} \tag{7}$$

Where, D represents the daily traffic flow, W represents the traffic flow in the latest week, and BW represents the biweekly traffic flow. The process of dividing the data into multiple timelines aims to take advantage of the periodicity of traffic flow. By dividing the data into short time series, the gradient can be effectively transferred in the deep time dimension, and the training time of the model can be greatly reduced.

4.2 Experimental evaluation index

In order to better analyze the prediction effect of the model, this paper adopts three common evaluation indexes [29]: absolute mean error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE).

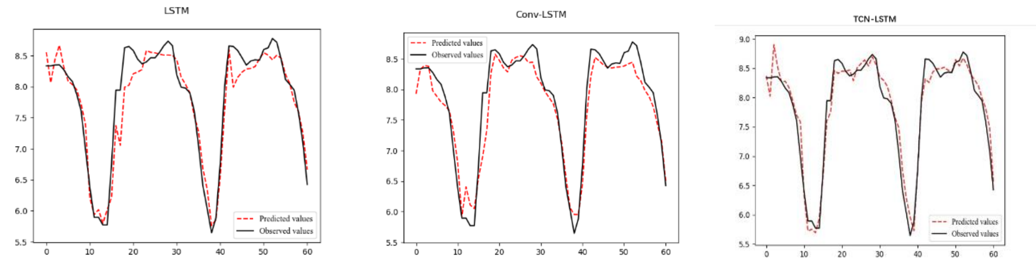
$$MAE = \frac{1}{T}\sum_{i=1}^{T}|y_i - \hat{y}_i| \tag{8}$$

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\frac{|y_i-\hat{y}_i|}{y_i} \times 100\% \tag{9}$$

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

4.3 Result analysis and comparison

The accuracy of the proposed TCN-LSTM model is verified by comparing the predicted results of traffic flow with the original data, and comparing the performance of different model methods. In order to verify the validity of the TCN-LSTM model, 13,622 data were used for model training, 10,000 consecutive data were used for cross-validation of the model, and the model results were used to predict the traffic flow in the next 60 moments. The prediction results of the model are shown in Figure 6.



(a) LSTM model prediction results diagram (b) Conv-LSTM model prediction results diagram (c) TCN-LSTM model prediction results diagram

Figure.6 model prediction result diagram

It can be seen from Figure 6 that the prediction performance and data generalization ability of the TCN-LSTM model are better than that of the Conv-LSTM model and the LSTM model alone. It can be seen from Fig. 6(c) that the error between the prediction results and the actual traffic flow data is kept within a small range of the TCN-LSTM model, the model can replace approximately the trend of the real traffic flow to a certain extent, and the predicted results are close to the original data. From the figure 6 (a) LSTM model itself has very good accuracy in time series prediction, but in some complex period, the result of the model is volatile. so in complex traffic flow data to predict the separate use of a model to predict the results of the accuracy and precision is not enough. It can be seen from Fig. 6(b) that the Conv-LSTM model has better prediction result than the simple LSTM model. When the result of the LSTM model fluctuated greatly, it can be seen that the result of the Conv-LSTM model at the same time are closer to the real value, and the fitting effect is better than that of the simple LSTM model. In figure 6(c), in the same complex traffic flow data, because the TCN - LSTM add a causal convolution layer, increasing the time sequence consistency, prediction results in the graph to keep the smaller error between predicted values and real values.

Table 2 The evaluation index analysis of different models

Model	MAE	MAPE	RMSE	R^2
TCN-LSTM	0.934	2.10	0.20	0.95
Conv-LSTM	0.951	2.63	0.28	0.91
LSTM	0.949	3.27	0.54	0.86
BP	0.931	4.75	0.65	0.77

It can be seen intuitively from Table 2. The MAE, MAPE, RMSE and determination coefficient $R^2$of TCN-LSTM model are 0.934,2.10,0.20,0.95, respectively. The TCN-LSTM model has smaller MAE, MAPE, RMSE and larger R^2, and the evaluation indexes of LSTM model and Conv-LSTM model have little difference, but the prediction results of Conv-LSTM model are better than that of LSTM model.

Through the comparison of the above experimental results, it can be concluded that the TCN-LSTM model proposed in this paper shows good performance in terms of fitting with real data and the results of evaluation index calculation, which proves that the proposed TCN-LSTM model can be used as an effective short-term traffic flow prediction model with high prediction accuracy.

5.Conclusion

Traffic flow data is complex and uncertain, and it is often not enough to adopt a certain model to forecast the short-term traffic flow. In this paper, a model framework for short-term traffic flow time series analysis and prediction based on the TCN-LSTM model is proposed. Through Python, from data processing to model implementation is completed, and the prediction result of the TCN-LSTM model for short-term traffic flow is realized. By comparing with different forecasting models, it is proved that TCN-LSTM model has high accuracy and stability in short-term traffic flow forecasting. Since there are many problems in the original traffic flow data collected, there is still a lot of room for improvement in data processing. On the other hand, this paper only studies and forecasts the traffic volume of a single intersection, and then studies the traffic connections between different intersections.

References

1. Zhao Na, YUAN Jiabin, Xu Han. Overview of intelligent transportation systems [J]. Computer science,2014,41(11):7-11.
2. Kumar K, Parida M, Katiyar VK. Short term traffic flow prediction for a non urban highway using artificial neural network. Procedia-Social and Behavioral Sciences, 2013,104: 755–764.
3. Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory[J]. Transportation Research Part B, 1984, 18(1):1-11.
4. Xie YC, Zhang YL. A wavelet network model for short-term traffic volume forecasting. Journal of Intelligent Transportation Systems, 2006, 10(3): 141–150.
5. Smith B L, Demetsky M J. Traffic Flow Forecasting: Comparison of Modeling Approaches [J]. Journal of Transportation Engineering, 1997, 123(4):261-266.
6. Hu W, Liu Y, Li L, et al. The short-term traffic flow prediction based on neural network[C]// International Conference on Future Computer and Communication. IEEE, 2010:V1-293-V1-296.
7. Kit, Yan, Chan. Neural-Network-Based Models for Short-Term Traffic Flow Forecasting Using a Hybrid Exponential Smoothing and Levenberg-Marquardt Algorithm[J]. Intelligent Transportation Systems IEEE Transactions on, 2012.
8. Mikolov T, Deoras A, Povey D, et al. Strategies for training large scale neural network language models[C]// Automatic Speech Recognition and Understanding. IEEE, 2012:196-201.
9. Lee S, Fambro D. Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting[J]. Transportation Research Record Journal of the Transportation Research Board,1999, 1678(1):179-188.
10. Kumar K, Parida M, Katiyar V K. Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network[J]. Procedia Social & Behavioral Sciences, 2013, 104:755-764.
11. Zhu Zhengyu, LIU Lin, CUI Ming. A Short-time traffic flow prediction model combining SVM and Kalman Filtering [J]. Computer Science, 2013, 40(10):248-251.
12. Cheng Yun, Cheng Xiaogang, On Miao Miao, et al. Traffic flow prediction based on ARIMA and wavelet neural network combination model [J]. Computer Technology and Development, 2017, 27(1):169-172.
13. Lv Y, Duan Y, Kang W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2):865-873.
14. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
15. Zhao Z, Chen WH, Wu XM, et al. LSTM network: A deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 2017, 11(2): 68–75.
16. Liu BY, Cheng JR, Liu Q, et al. A long short-term traffic flow prediction method optimized by cluster computing. Electrical & Electronic Engineering, 2018.
17. Yang H F, Dillon T S, Chen Y P. Optimized Structure of the Traffic Flow Forecasting Model With a Deep Learning Approach.[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, PP(99):1-11.
18. Tian Y, Pan L. Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network[C]// IEEE International Conference on Smart City/socialcom/sustaincom. IEEE, 2015:153-158.
19. Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction[C]// Chinese Association of Automation. IEEE,2017:324-328.
20. Zhang Y, Liu Q, Chen E, et al. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification[J]. Frontiers of Computer Science, 2016, 10(1):1-17.

21. Wu Y, Tan H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework[J]. 2016.
22. Ma XL, Tao ZM, Wang YH, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 2015, 54: 187–197.
23. Yu D, Liu Y, Yu X. A Data Grouping CNN Algorithm for Short-Term Traffic Flow Forecasting[M]// Web Technologies and Applications. Springer International Publishing, 2016:92-103.
24. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
25. Yang H F, Dillon T S, Chen Y P. Optimized Structure of the Traffic Flow Forecasting Model With a Deep Learning Approach. [J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, PP(99):1-11.
26. Bai S, Kolter J Z, Koltun V. Trellis Networks for Sequence Modeling[J]. 2018.
27. Changfeng, Kuang Wenzhen, Haifeng, etc. Simulation Research on Accurate Prediction of Road Traffic Flow [J]. Computer Simulation, 2016, 33(6):111-116.
28. Williams B, Durvasula P, Brown D. Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models[J]. Transportation Research Record, 1998,1644(1):132-141.
29. Bai S , Kolter J Z , Koltun V . Trellis Networks for Sequence Modeling[J]. 2018.
30. He Guoguang, MA Shoufeng, LI Yu.A short-time traffic flow prediction method based on wavelet decomposition and Reconstruction [J]. Systems Engineering Theory and Practice, 2002, 22(9):101-106.
31. Wang J, Deng W, Guo YT. New Bayesian combination method for short-term traffic flow forecasting. Transportation Research Part C: Emerging Technologies, 2014, 43: 79–94.
32. Song Di. Research on expressway Short-time Traffic Flow Prediction Algorithm [D]. Zhejiang University of Technology, 2016.
33. Shi Yuqin. Overview of key Technologies of Intelligent Transportation System in smart City Construction [J]. Internet of Things Technology, 2017, 7(2):54-55.
34. Li Mingming, LEI Juyang, Zhao Congjian.Short-term traffic flow prediction based on LSTM-BP combined model. Computer System Applications, 19,28(10): 152-156.
35. Williams B M, Hoel L A. Modeling and Forecasting Vehicular Traffic Flow asa Seasonal ARIMA Process: Theoretical Basis and Empirical Results[J]. Journal of Transportation Engineering, 2003, 129(6):664-672.
36. K. Greff, R. K. Srivastava, J. Koutn´ık, B. R. Steunebrink, and J.Schmidhuber. LSTM: A search space odyssey. CoRR, abs/1503.04069.