Research on Traffic Flow Prediction Methods Based on Deep Learning

Junjin Chen[1], Jiatong Song[2,3,*]

[1]International College, Wenzhou Business of College, Wenzhou, China
[2]Oulu College, Nanjing Institute of Technology, Nanjing, China

[3]X00217230106@njit.edu.cn
*corresponding author

Abstract. In recent years, traffic flow prediction technology has been transformed from statistics based parametric methods and machine learning driven non-parametric methods to big data driven deep learning methods. This paper summarizes and summarizes the existing methods and improvement measures of long and short term traffic flow prediction based on deep learning. The time range of traffic flow forecast based on the model is divided into long-term and short term. The short-term traffic flow forecasting methods are subdivided into time series model, non-parametric forecasting model and probability forecasting model, and the advantages and disadvantages of each method and the feasibility of the specific methods are summarized. As for the long-term model, it is mainly based on the application of GCN model to other models, and then the specific methods of its hybrid model are outlined, systematically describing the value of deep learning in traffic flow prediction. Finally, the future research direction and development trend in this field are predicted and prospected.

Keywords: Traffic flow prediction, deep learning, long and short term traffic flow.

1. Introduction

With the acceleration of the urbanization process and the increase of the number of motor vehicles, the problem of traffic congestion is becoming more and more serious, which brings great challenges to urban traffic management [1-3]. Traditional traffic flow forecasting methods often rely on historical data and statistical analysis, and it is difficult to deal with complex nonlinear relations and time series dependence effectively. As an advanced machine learning technology, deep learning has made remarkable achievements in various fields in recent years, and its application in traffic flow prediction has gradually attracted attention. At present, deep learning has made remarkable progress in the field of traffic flow prediction [4,5], which is mainly reflected in model development, data source diversification, model performance evaluation, real-time prediction and intelligent application. Deep learning is also of great significance in the research of traffic flow prediction.It can optimize resource allocation and traffic management strategies more accurately, so as to improve traffic efficiency. In promoting the development of intelligent transportation system, deep learning helps to promote the intelligent transformation of urban traffic management, and provides strong support for real-time traffic flow predicition [6].

This paper reviews the research progress of traffic flow prediction based on deep learning, and analyzes the application of long term memory network (LSTM), convolutional neural network (CNN) and graph neural network (GNN) in this field. By systematically reviewing the existing literature, this paper discusses the effectiveness of these deep learning models in capturing the temporal characteristics and spatial relationships of traffic data, and shows their obvious advantages in improving the prediction accuracy and real-time performance. Finally, this paper summarizes the key contributions of deep learning in traffic flow prediction and looks forward to the direction of future research.

2. Short-term traffic flow prediction

The short-term traffic flow prediction problem can be expressed as: the data series composed of traffic flow observations in several periods of history are input into. In a prediction model, a prediction of the sequence of traffic flows over a number of future periods is output. Table 1 is the comparison of traffic flow prediction.

Table 1. Comparison of traffic flow prediction

Method classification	Representative algorithm	Merits and demerits
Statistical model based on time series	ARIMA (Auto-regressive Integrated Moving Average) model, Kalman filter model, etc	Advantages: easy to understand;Can be used to handle nonlines Sex, non-stationary time sequence Column. Incomplete points: Over-fitting problems may exist; Only univariate time series can be handled
Probabilistic graph model	Bayesian network, Markov chain, Markov random field, hidden Markov model, etc Combining the Stochastic User Equilibrium (SUE) principle, a Bayesian network model is established to predict the OD flow and road segment flow [2]	Advantages: The structure and parameters have clear physical meaning, which makes the model easier to understand and interpret. Incomplete Point: It is difficult to modeling; parameter optimization difficulty.
Nonparametric prediction model	(Back Propagation Neural Network, BPNN)、(KNearest Neighbor, KNN)etc. Mixed prediction model of traffic flow [3]	Advantages: non-parametric method often more accurate, they seek the best fitting data point;more flexible. Incomplete Point: At the cost of the need to carry out large amount of observation,which may lead to low efficiency; It may introduce overfitting.

2.1. Statistical model based on time series

The time series approach is a statistical analysis method that relies on chronological order to forecast future trends and changes by examining a dataset organized in sequential order. This technique primarily depends on historical data and posits that past trends and patterns can persist into the future, thereby serving as a basis for future predictions [4]. Tan Chunman and colleagues utilize the ARIMA model to estimate the linear component of short-term traffic flow while using artificial neural networks to forecast its nonlinear residual component. Nonetheless, depending solely on a single model for predictions complicates efforts to enhance the accuracy of short-term traffic flow forecasts. Furthermore, the ARIMA model has specific drawbacks; it presumes that the data remains stable, indicating that the

statistical properties of the data do not fluctuate over time. In reality, however, factors such as accidents can cause traffic flow data to display unstable patterns. Moreover, the ARIMA model typically necessitates a substantial amount of historical data to function effectively, which is not always accessible. Consequently, to emphasize the predictive strengths of both stable and unstable segments, Yin5) and colleagues seek to optimize the genetic example group of the wavelet neural network during gray correlation analysis. They integrate the simplified input structure of the neural network with Genetic Particle Swarm Optimization Wavelet Neural Network (GPSOWNN) and include predicted outputs from the ARIMA model as part of their methodology. Kalman Filtering is a recursive filtering algorithm used to estimate the state of a dynamic system. The core idea of Kalman Filtering is to represent the state of the system as a state space model, in which the state variables change with time, and the observation variable is a linear combination of state variables. The filter estimates the current state of the system by constantly updating the state estimate and covariance matrix.

Based on the literature [6], in order to solve the problem of time lag in the Kalman Filter model to predict traffic flow, on the basis of the existing Kalman Filter short-time traffic flow prediction model, combined with the daily similarity features of urban road traffic flow. The existing Kalman Filter prediction model is improved, and the parameters in the improved model are determined through the fuzzy logic method, and the fuzzy Kalman Filter traffic flow prediction model is designed (Figure 1). So as to make accurate predictions of short-term traffic flow. Numerical analysis and comparison results show that compared with Kalman's short-term traffic flow prediction method, fuzzy Kalman's short-term traffic flow prediction method can improve the realtimeness of the prediction process, and reduce average relative error by 027%, the average absolute relative error by 7.26%, and the maximum absolute relative error by 32 .43, further improve the prediction accuracy.
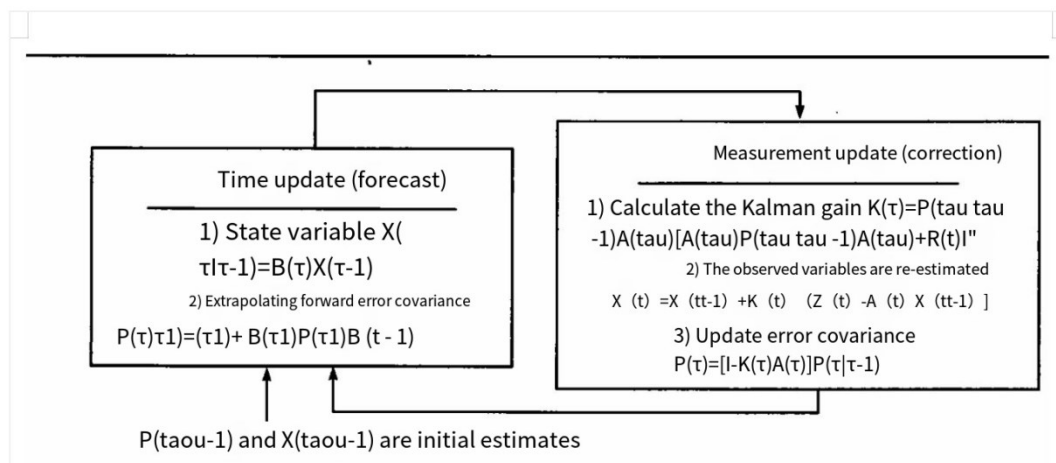

Figure 1. Working principle of Kalman Fillter traffic flow prediction model [6]

2.2. Probability graphic model

Over the years, domestic and foreign scholars have conducted lots of research on traffic flow forecasting. The existing methods are based on different theories or different starting points, and large amounts of traffic historical data is used to train neural network models on neural network-based methods, which can be obtained from a mapping relationship between input and output, and then the corresponding traffic prediction results are obtained. However, these methods have many shortcomings in short-term traffic flow prediction, such as neural network-based attribute and the methods of data mining and fusion have a high degree of dependence on historical data. These methods mainly focus on the real-time traffic flow situation of the target road section, ignore the road network structure and the traffic flow situation of adjacent road sections and other influencing factors, and cannot effectively reflect the interdependence between the more complex traffic network and the intersection in reality. And the uncertainty of dependence does not better reflect the time and space dependence between intersections

in the traffic network, and the combination of multiple models makes the prediction method more complex. Bayes is an important probability model. Its good nature in the representation and reasoning of uncertainty knowledge makes it stand out in solving the real-time and efficiency of short-term traffic predictions. At present, the Bayesian linear regression model has made remarkable achievements in combining historical traffic data with weather, time and other factors, while Bayesian deep learning has an advantage in effectively capturing complex nonlinear relationships. Overall, the research results of the Bayesian method in the field of traffic flow prediction show that it has unique advantages in data processing, multi-source fusion, dynamic update, etc. With the deepening of research and the advancement of technology, the Bayesian method is expected to play a greater role in intelligent transportation systems in the future.

2.3. Nonparametric prediction model

Due to the randomness and complexity of the evolution of traffic flow in urban road traffic systems, and the uncertainty of distinguishing real-time traffic status, real-time and accurate short-term traffic flow prediction is a key factor in traffic control and induction. Using nonparametric regression to predict short-term traffic flow is a good prediction method, which can well reflect the nonlinearity, time variability and uncertainty of traffic flow. Nonparametric methods (such as Mechine Learning based methods) do not depend on welldefined equations or human intervention. The proximity nonparametric regression prediction algorithm is one of them. But it also has some problems, which limit its practical application. The main defect of the K nearneighbor nonparametric regression prediction algorithm is that there are many factors affecting the traffic flow. At the same time, the preprocessing of the original data is insufficient, resulting in the component of the state vector being too simple or too complex. At the same time, it also causes the unreasonable structure of the historical database, which eventually leads to too much time consuming time in the matching process. Because the original data is not classified and all flow modes adopt a single K value, the prediction error increases due to misjudgment of the current flow mode. Therefore, according to the defects of the kproximity method, Lin [7] proposed four improvement measures such as data preprocessing, rapid database creation, variable K value strategy and error feedback mechanism to improve algorithm accuracy.

These four improvement measures finally simulated the improved algorithm, and the results indicated that the time consumption and prediction error of the improved algorithm were obviously better than those of the improved algorithm. At the same time, artificial neural networks have also received continuous attention. At present, due to the adequacy of traffic big data acquisition and the relative maturity of deep learning theory, the adoption of deep neural networks as the core methodology of short-term traffic flow prediction has become a hot spot and mainstream of research. Automatic Encoder [8], Various Types Of Convolutional Neural Networks (1D Causal Convolution [9]), 2D Image Convolution [10]), 3D Image Convolution [11], Etc.), Circular Neural Networks (Long short-term Memory (LSTM) And Other Neural Networks The theory of the network field has been used in the research of short-term traffic flow prediction. It can be seen that the artificial neural network model is still a model with great potential in the field of short-term traffic flow forecasting. In order to predict short-term traffic flow more accurately, Sun [12]) proposed a combined model (SARIMAGRNN model) composed of seasonal autoregressive summation moving average model (SARIMA) and generalized regression neural network (GRNN) model (figure 2), which combines the advantages of time series prediction between time series model and neural network model. Empirical research results show that the combined model constructed by the scheme not only has good predictive performance but also a simple structure, which is very suitable for the real-time prediction of short-term traffic flow on urban roads.
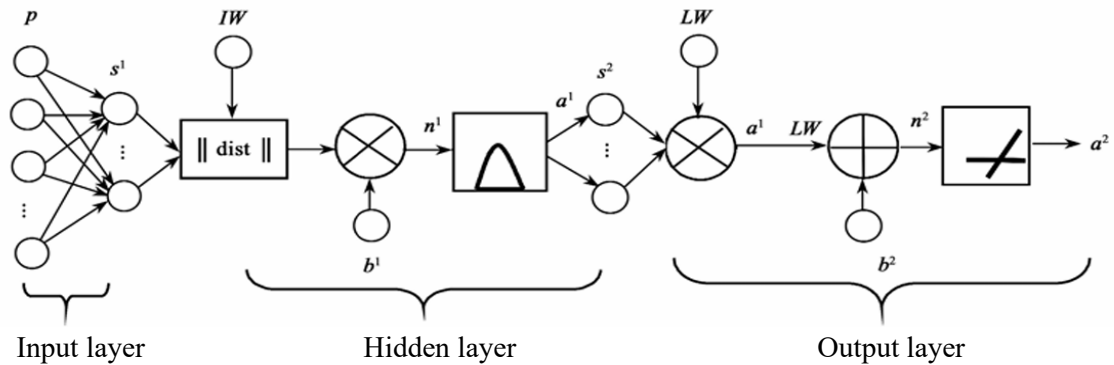
Figure 2. A structure diagram of the generalized regression neural network [12]

3. Long-term traffic flow forecasting

These methods can help reduce environmental impact, improve traffic safety, optimize public transportation services, urban planning and development, and improve residents' travel time. The following analysis is based on two methods.

3.1. GCN based method analysis

In [13], the authors introduced a temporal graph convolutional network (TGCN) by combining the advantages of GCN and GRU. The model utilizes graph convolution to capture the interdependence between different road segments and uses GRU to model the dynamics of Temporal data. This enables effective traffic flow prediction even when dealing with complex traffic networks and changing traffic conditions. The processing flow is shown in Figure 3, and the TGCN model can effectively cope with the problems in prediction of traffic flow, mainly due to its dual grasp of spatial and temporal factors. Specifically, the model uses GCN to deeply analyze the structure of urban traffic networks, thereby revealing the spatial relationships between different road segments. In addition, the GRU can be used to effectively monitor the fluctuation of traffic flow over time, thereby grasping the continuity and regularity of time. Through this dual processing, the TGCN model can accurately predict future traffic flow.
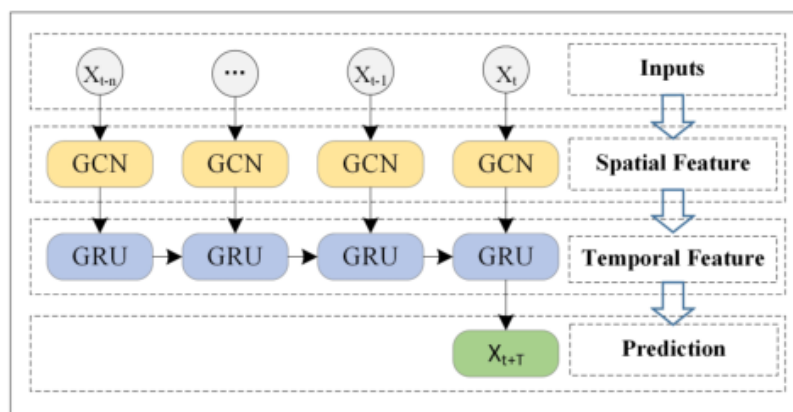

Figure 3. TGCN model [13]

TGCN As an innovative traffic prediction model, the spatial and temporal dependencies of traffic data are successfully captured by combining GCN and GRU.STGCN further extends this concept by introducing spatio-temporal convolutional blocks. In [14] Based on STConv, the team proposed spatio-temporal graph convolutional networks (STGCAN).The processing flow is shown in Figure 4.
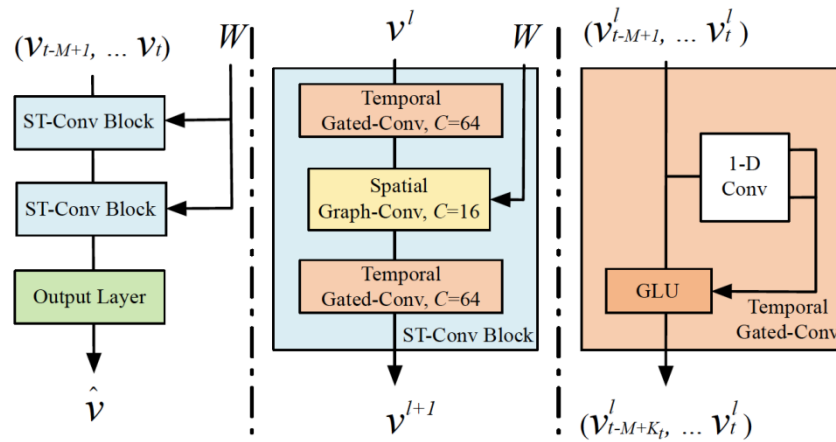
Figure 4. STGCAN [14]

The team directly performs graph convolution on graph structured data through an improved graph CNN model to extract spatial features in transportation networks, then use gated CNN to extract temporal features, the convolutional layer controls the input state by activating the gate, enables the model to effectively reveal the structure and dynamic changes of time series, Finally, two STConv modules are stacked to achieve fast spatial state propagation in graph convolution.

In general, STGCN is capable of modeling and predicting traffic networks, and it can also be applied to spatio-temporal sequence learning tasks. Secondly, by combining spatio-temporal convolution blocks with graph convolution and gated temporal convolution, it can acquire features in spatial and temporal, effectively capturing the spatial and temporal connections of traffic flow. At the same time, due to its convolutional structure, STGCN requires fewer parameters, which makes the model more efficient and easier to train.

3.2. LSTM based method analysis

LSTM [15] Compared with RNN, there are 3 more control gates, These components enable LSTM to effectively solve the gradient vanishing problem and capture long-term dependencies, making it more powerful and stable than ordinary RNN when processing long sequence data. Based on the traditional method of processing sequence data in only one direction, a bidirectional LSTM (BiLSTM) is introduced. The processing flow is shown in Figure 5.
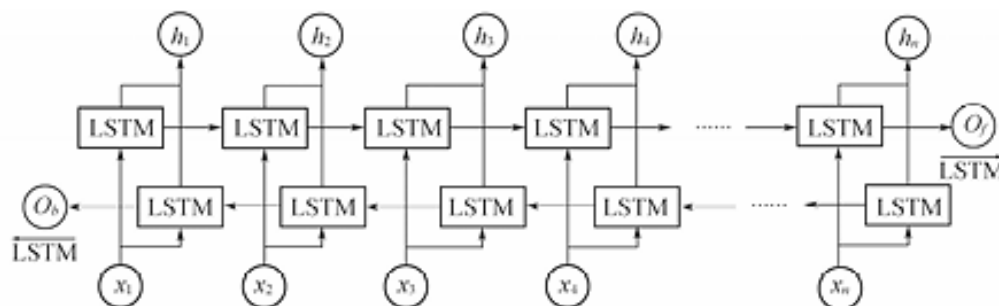

Figure 5. The bidirectional LSTM [15]

This frame can run two LSTMs, one forward and one reverse, while processing sequence data. Therefore, in time series, BILSTM can take into account time series information in both directions and provide more accurate prediction results. By combining BILSTM in time series with CNN model for extracting local space, a CNN BiLSTM traffic fusion prediction model is proposed. Figure 6 is a CNN BiLSTM model.
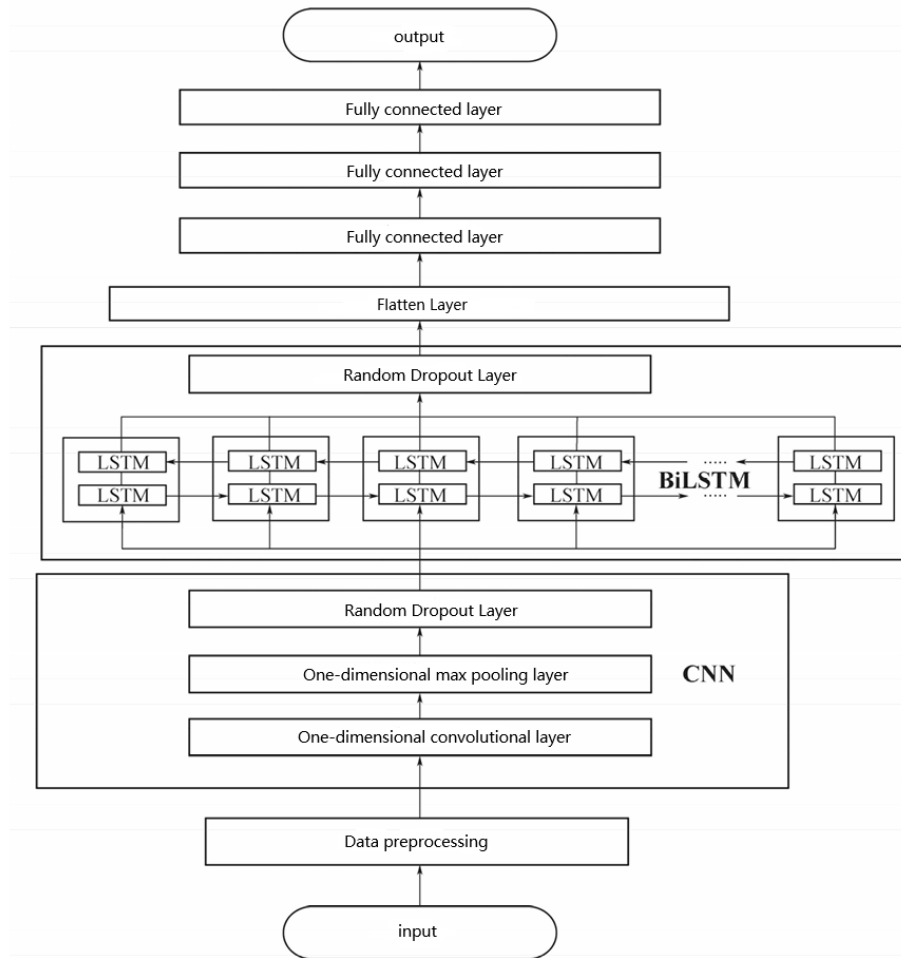
Figure 6. CNN BiLSTM model [16]

To avoid overfitting, the team [16] added random dropout layers after the CNN and the BiLSTM to randomly drop nodes. The normalized and missing value processed historical traffic flow data was first fed into the CNN and random dropout layers to extract local spatial features. These feature vectors were then fed into the BiLSTM layer, another random dropout layer, a flattening layer, and three fully connected layers in sequence to finally generate the prediction results.

Overall, the model effectively extracts spatial and temporal features by integrating the advantages of CNN and BiLSTM, and prevents overfitting through the Dropout layer, thereby achieving high precision prediction of complex time series data and predicting changes in traffic flow at different geographical locations.

4. Conclusion

This paper systematically classifies and sorts out the traffic flow prediction method based on deep learning and its research status. According to the different forecasting time range, this method can be divided into short term traffic flow prediction and long term traffic flow prediction. Short term traffic flow prediction methods can be subdivided into three categories: prediction methods based on time series, non-parametric model and probability map model. Time series based methods can identify trends, periodic and nonlinear time series, but there may be problems such as fitting problems and can only handle single-variate time series; non-parametric models are known for their flexibility and accuracy, but require more data and computing resources; the probability graph model can capture through the

graph structure. Capture spacetime characteristics, but the disadvantage is that the modeling is difficult and the cost requirements are too high. Long-term traffic flow forecasting is divided into two types, one is GCN based traffic flow prediction, and the other is LSTM based traffic flow prediction. GCN based forecasts are suitable for revealing the spatial distribution characteristics of traffic flow, while LSTM based predictions perform better in terms of time dynamics. From these basic models themselves, GCN may require a lot of computing resources, and LSTM may face problems such as gradient disappearance or explosion.

Based on the existing research results, the field of traffic flow forecasting faces several key challenges and development directions. First of all, enhancing the interpret ability of deep learning models is an important research direction, because these models are often regarded as "black boxes" and lack transparency, which affects the understanding and trust of traffic managers in the predicted results. Secondly, real-time data processing is also an urgent problem to be solved. It is necessary to develop methods to efficiently and accurately process massive real time data from sensors, traffic cameras, etc., and provide forecast results in a timely manner. In addition, in order to improve the accuracy of prediction, it is necessary to study how to effectively integrate historical traffic data with a variety of heterogeneous data sources such as meteorology, events and social media. Finally, it is equally important to optimize the algorithm to improve the efficiency of model training and prediction, and to ensure scalability on large-scale data. By responding to these challenges, we can further promote the development of traffic flow prediction technology

Authors contribution
All the authors contributed equally and their names were listed in alphabetical order.

References
[1]   Cui, J. X., Yaojia, & Zhao, B. Y. (2024). A review of short-term traffic flow forecasting methods based on deep learning. Acta Photonica Sinica, 24(2), 5064. doi: 10.19818/j.cnki.16711637. 2024.02.003

[2]   Zhu, S. L., Cheng, L., & Chu, Z. M. (2013). Bayesian network model for traffic flow estimation using prior link flows. Journal of Southeast University (English Edition), 29(3), 322327.

[3]   Liu, Z., Du, W., Yan, D. M., et al. (2017). Short term traffic flow forecast based on K nearest neighbor algorithm and support vector regression. Journal of Highway and Transportation Research and Development, 34(5), 122128.

[4]   Tan, M. C., Feng, L. B., & Xu, J. M. (2007). Traffic flow prediction based on hybrid ARIMA and ANN model. China Journal of Highway and Transport, 20(4), 118121. doi: 10.3321/j.issn: 10017372.2007.04.021

[5]   Yin, L. S., Tang, S. Q., Li, S., & He, Y. G. (2019). Traffic flow prediction based on hybrid model of autoregressive integrated moving average and genetic particle swarm optimization wavelet neural network. Journal of Electronics & Information Technology, 41(9), 22732279. doi: 10. 11999/JEIT181073

[6]   Guo, H. F., Fang, L. J., & Yu, L. (2013). short-term traffic flow prediction based on fuzzy kalman filter method. Journal of Zhejiang University of Technology, 41(2), 4. DOI: 10.3969/j.issn. 10064303.2013.02.022

[7]   Lin, C. (2015). Based on K nearest neighbor nonparametric regression of short-term traffic flow prediction algorithm research. University of Electronic Science and Technology. The DOI: 10. 7666/d.D.664651

[8]   Lyu, Y. S., Duany, Y. J., Kang, W. W., et al. (2015). Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 865873.

[9]   Ge, L., Li, S. Y., Wang, Y. Q., et al. (2020). Global spatialtemporal graph convolutional network for urban traffic speed prediction. Applied Sciences, 10(4), 1509. doi: 10.3390/app10041509

[10]   Lin, H. X., Jia, W. J., Sun, Y. P., et al. (2019). Spatialtemporal selfattention network for flow prediction. arXiv preprint arXiv:1912.07663.

[11]   Guo, S. N., Lin, Y. F., Li, S. J., et al. (2019). Deep spatialtemporal 3D convolutional neural networks for traffic data forecasting. IEEE Transactions on Intelligent Transportation Systems, 20(10), 39133926.

[12]   Sun, X. H., & Liu, T. Q. (2008). short-term traffic flow forecasting based on a hybrid neural network model and SARIMA model. Journal of Transportation Systems Engineering and Information Technology, 8(5), 3237.

[13]   Yu, B., Yin, H., & Zhu, Z. (2017). spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

[14]   Liu, Y., & Gu, Y. (2022). Prediction of spatio-temporal characteristics of highway traffic flow based on CNNBiLSTM. Transportation Science and Technology and Economy, 24(1), 918.

[15]   Wan, S. X., Lan, Y. Y., Guo, J. F., et al. (2017). Localized bidirectional long short-term memory for text classification. Journal of Chinese Information Science, 31(3), 7.

[16]   Huang, W., Song, G., Hong, H., & Xie, K. (2018). Deep Architecture for Traffic Flow Prediction: Combining Convolutional Neural Networks and Bidirectional Long Short-Term Memory Networks. IEEE Transactions on Intelligent Transportation Systems, 19(5), 1447-1457. https://doi.org/10.1109/TITS.2017.2766864