HIGH TECHNOLOGY LETTERS | Vol. 30 No. 3 | Sep. 2024 | pp. 221-230

doi:10. 3772/j. issn. 1006-6748. 2024. 03. 001

Adaptive spatial-temporal graph attention network for traffic speed prediction①

ZHANG Xijun (张玺君)②, ZHANG Baoqi, ZHANG Hong, NIE Shengyuan, ZHANG Xianli
(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, P. R. China)

Abstract

Considering the nonlinear structure and spatial-temporal correlation of traffic network, and the influence of potential correlation between nodes of traffic network on the spatial features, this paper proposes a traffic speed prediction model based on the combination of graph attention network with self-adaptive adjacency matrix (SAdpGAT) and bidirectional gated recurrent unit (BiGRU). Firstly, the model introduces graph attention network (GAT) to extract the spatial features of real road network and potential road network respectively in spatial dimension. Secondly, the spatial features are input into BiGRU to extract the time series features. Finally, the prediction results of the real road network and the potential road network are connected to generate the final prediction results of the model. The experimental results show that the prediction accuracy of the proposed model is improved obviously on METR-LA and PEMS-BAY datasets, which proves the advantages of the proposed spatial-temporal model in traffic speed prediction.

Key words: traffic speed prediction, spatial-temporal correlation, self-adaptive adjacency matrix, graph attention network (GAT), bidirectional gated recurrent unit (BiGRU)

0 Introduction

Intelligent transportation systems (ITS) play an important role in building smart city[1]. Traffic prediction provides favorable technical support for ITS, and its main method is to forecast the future traffic situation by combining the historical data collected by traffic sensors and the real road network structure.

In the early stage, traditional linear time series models, such as auto-regressive integrated moving average model (ARIMA) and its variants[2-3], Kalman filter[4], and other statistical methods were widely used. However, these models assume that the conditional variance of the time series is stationary, which does not satisfy the actual traffic network condition. Later, machine learning (ML) methods were introduced in the field of traffic prediction, such as Bayesian estimation model[5], K-nearest neighbor algorithm[6]. However, the prediction accuracy of these methods is not high, and the model performance is not good.

Deep learning (DL) technology is widely used in traffic forecasting by many scholars. For example, models based on convolution neural networks (CNNs) and recurrent neural networks (RNNs) are helpful to extract spatial-temporal dependent features[7-9]. However, when the traffic network is regarded as Euclidean space, its topological structure may be lost, which shows the limitations of CNN in extracting the spatial features of the traffic network.

In recent years, graph neural network (GNN) has become a cutting-edge technology in deep learning research. The network based on graph convolutional network (GCN) and graph attention network (GAT) are used to process non-Euclidean spatial graph data and improve the ability of the model to extract the spatial features of the traffic network. Ref. [10] proposed a combined model of GCN and gated recurrent unit (GRU) in the early stage, which extracted the spatial topological structure features and time series features of the traffic network respectively. Ref. [11] innovatively proposed a spatial-temporal graph convolutional network based on attention mechanism to capture spatial-temporal dynamic correlation features. Ref. [12] proposed an adaptive graph learning algorithm based on graph convolutional networks to capture the dependencies between nodes. Ref. [13] proposed a component

① Supported by the National Natural Science Foundation of China (No. 61461027, 61762059), the Provincial Science and Technology Program supported the Key Project of Natural Science Foundation of Gansu Province (No. 22JR5RA226)

② To whom correspondence should be addressed. E-mail: zhangxijun198079@sina.com.

Received on Jan. 4, 2024

of adaptive learning graph structure using spatial-temporal convolution network, which learns the optimal graph structure from macro and micro perspectives respectively. Among them, the short-term spatial-temporal dependence is learned from micro perspective, and the long-term spatial-temporal dependence is learned from macro perspective. Ref. [14] adopted the method of node embedding and utilized the GAT to generate spatial adjacent subgraphs adaptively at different time steps to improve the ability of capturing spatial features. Ref. [15] designed a trend graph attention network, which does not depend on the predefined graph structure at all, focusing on learning potential spatial and temporal dependency features to improve the generalization ability of the model. However, traffic conditions change dynamically, and the correlation between nodes is not only affected by the real distance between static network nodes, but also by the dynamic change of traffic flow. Therefore, considering only the static topology of the traffic network may lose some potential relationships between nodes, which is not conducive to capturing the spatial correlation between potential network nodes.

In view of the above problems, this paper proposes a traffic speed prediction model based on the combination of graph attention network with self-adaptive adjacency matrix (SAdpGAT) and bidirectional gated recurrent unit (BiGRU).

The main contributions are as follows.

(1) Constructing an adjacency matrix which adaptively learns the correlation degree between network nodes according to the traffic speed time series features, thus improving the ability of capturing the spatial information of potential road network and improving the generalization of the model.

(2) Combining GAT and BiGRU model to form spatial-temporal blocks, the residual connection of multiple spatial-temporal blocks can preserve more important spatial-temporal features information. Meanwhile, it can effectively alleviate the gradient disappearance problem caused by model training, speed up model training and improve the accuracy of model prediction.

(3) The results of multi-step prediction are analyzed and compared with the benchmark model of the same data sets. The performance of the proposed model is better than other models, which provides theoretical support for the field of traffic prediction.

1 Prepared work

In this section, the traffic network is defined, and then the specific problems of traffic speed prediction are clarified.

1.1 Definition of traffic network

The traffic network is defined as a directed graph $G = (V, E, \boldsymbol{A})$, where V represents the set of network nodes, E represents the edges between nodes, and $\boldsymbol{A} \in R^{N \times N}$ represents the weighted adjacency matrix. At time step t, the traffic speed observed by node N (sensor) is feature F, represented by $\boldsymbol{X}_t \in R^{N \times F}$. The same node and its domain node show different correlations at different time steps t, which shows the spatial-temporal correlation of traffic prediction, as shown in Fig. 1.

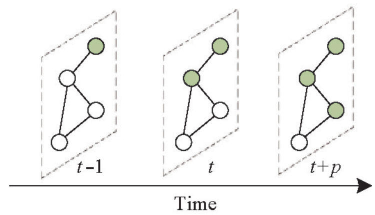


Fig. 1 Node correlation at different time steps t

1.2 Definition of problem

In the problem of traffic speed time series prediction, the h historical steps time series $\boldsymbol{H}_t = \{\boldsymbol{X}_{t-h+1}, \boldsymbol{X}_{t-h+2}, \cdots\cdots, \boldsymbol{X}_t\}$ is given to predict the p steps time series $\boldsymbol{P}_t = \{\boldsymbol{X}_{t+1}, \boldsymbol{X}_{t+2}, \cdots\cdots, \boldsymbol{X}_{t+p}\}$ in the future. Each time step t will produce an adjacency matrix with dynamic changes in node correlation. According to the data of the past 1 h, this paper predicts the traffic speed in the next 15 min, 30 min and 60 min.

2 Methods

In this paper, a combined model SAdpGAT-BiGRU is designed to extract spatial-temporal dependency features. As shown in Fig. 2, the proposed model is composed of the spatial and temporal module, and the residual mechanism is used to design multi-layer spatial-temporal block stacking. In the spatial module, GAT is used to capture the spatial dependence features of the real road network, and SAdpGAT is used to capture the spatial dependence features of potential road network. Temporal dependence features can be captured by inputting the spatial dependent features into BiGRU. Then, the output features of the spatial-temporal blocks are fully connected, and the prediction results of the real road network and the potential road network are connected to get the final prediction results of the model.

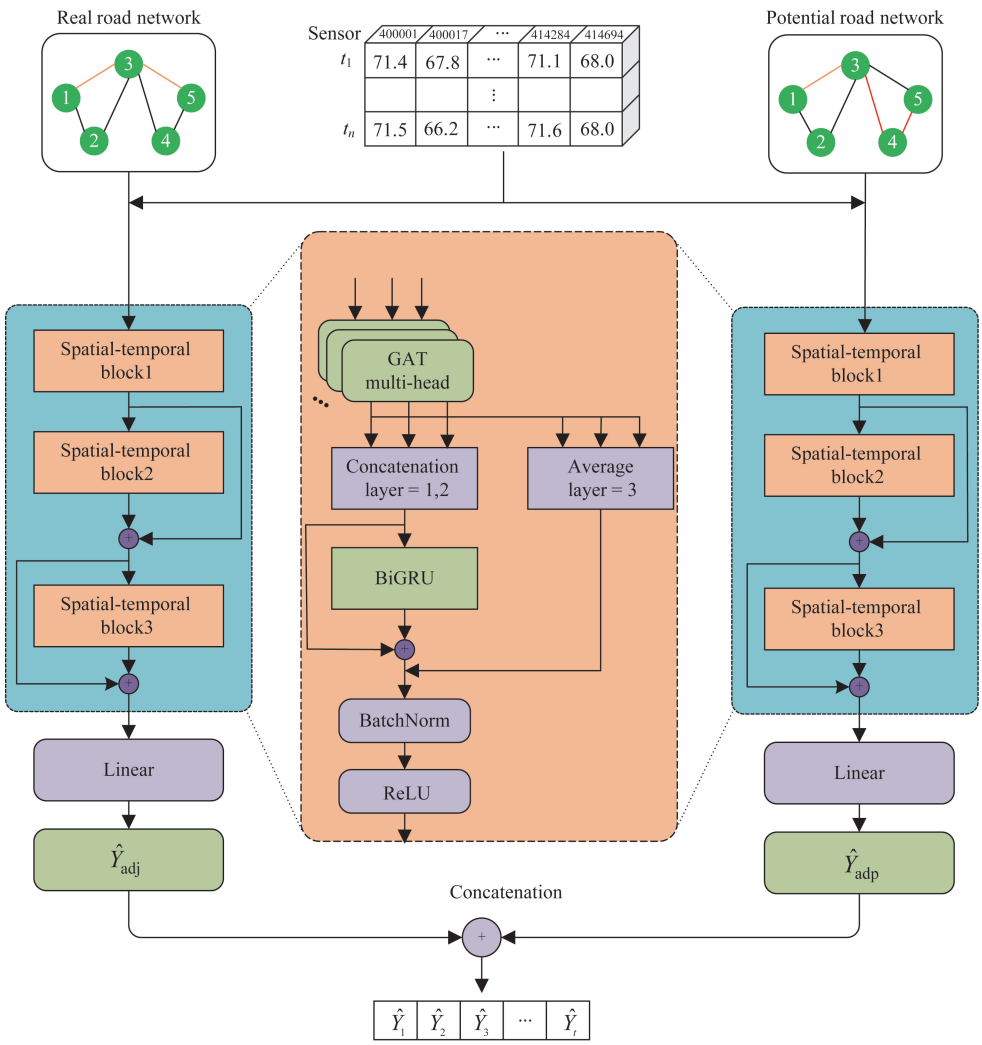


Fig. 2 SAdpGAT-BiGRU model structure

2.1 Spatial feature extraction based on adaptive graph attention

As shown in Fig. 2, the left half of this model extracts the spatial features of the real road network, and uses Gaussian kernel function to calculate the correlation between road network nodes to obtain the correlation adjacency matrix. The right half of the model extracts the spatial features of the potential road network, and constructs an adaptive correlation adjacency matrix based on the topological structure of the real road network, which assigns the same initial correlation degree to all nodes, that is, $\boldsymbol{A}'[i][j] = C$, $0 \leqslant i, j \leqslant N-1$, where $\boldsymbol{A}'$ represents the newly constructed adaptive correlation adjacency matrix, i and j represent the row and column indexes of the matrix, C is a constant between $(0,1)$. Then, the initial time series is used as a condition, which is linearly transformed with the adjacency matrix of real road network and the adaptive correlation adjacency matrix of potential road network, and then input into GAT network. Finally, the attention score function is used to calculate the correlation coefficient between nodes.

GAT[16] is a network based on an attention mechanism proposed in the field of graph neural networks, which does not depend on the overall structure of the graph and only focuses on the correlation between nodes. Therefore, the spatial dependent features can be better extracted and the generalization performance of the model can be improved. The implementation of GAT is as follows: node features of road network adjacency matrix are used as input, $\boldsymbol{h} = \{\vec{\boldsymbol{h}}_1, \vec{\boldsymbol{h}}_2, \cdots\cdots, \vec{\boldsymbol{h}}_N\}$, $\vec{\boldsymbol{h}}_i \in R^F$, where N represents the number of nodes, F represents the feature of each node, and the node features output by the model is $\boldsymbol{h}' = \{\vec{\boldsymbol{h}}'_1, \vec{\boldsymbol{h}}'_2, \cdots\cdots, \vec{\boldsymbol{h}}'_N\}$, $\vec{\boldsymbol{h}}'_i \in R^{F'}$.

As shown in Fig. 3, the self-attention mechanism is used to calculate the attention coefficient e_{ij} from node i to node j, as shown in Eq. (1).

$$e_{ij} = \mathrm{LeakyReLU}(\vec{a}^{\mathrm{T}}[X\vec{h}_i \parallel X\vec{h}_j]) \quad (1)$$

where $\vec{a} \in R^{2F'}$ represents a learnable weight parameter, $\parallel$ represents a connection operation, and $X \in R^{F' \times F}$ represents a shared weight matrix, that is, traffic speed time series feature. The linear transformation is normalized by using the LeakyReLU activation function, which solves the problem that the gradient of the ReLU function disappears when the independent variables are negative.

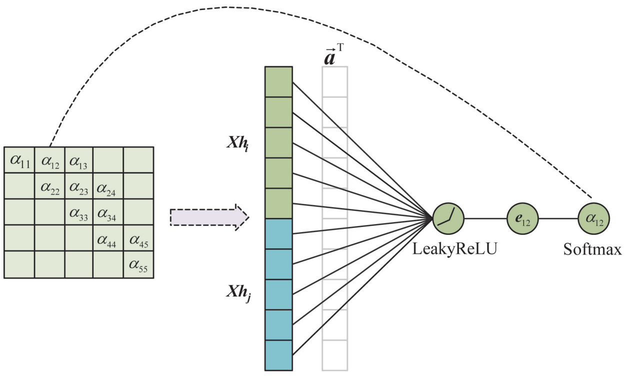


Fig. 3 The GAT network calculates correlations between nodes in the road network

The attention coefficient between nodes is calculated using softmax function, as shown in Eq. (2).

$$\alpha_{ij} = \mathrm{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (2)$$

where, N_i represents the domain node set of node i.

According to the normalized attention coefficient, the node features are calculated as shown in Eq. (3).

$$\vec{h}_i' = \sigma\left(\sum_{j \in N_i} \alpha_{ij} X \vec{h}_j\right) \quad (3)$$

where, σ represents the nonlinear activation function, X represents the shared weight matrix, and $\vec{h}_i'$, $\vec{h}_j$ represent the new features of nodes.

In order to better capture the spatial correlation between nodes, the multi-head attention mechanism is used to calculate the nodes features, and K represents the number of attention heads. In this model, the features are connected at the first $n-1$ layer, and the calculation is shown in Eq. (4). In the n layer, the features are averaged, and the calculation is shown in Eq. (5).

$$\vec{h}_i' = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k X^k \vec{h}_j\right) \quad (4)$$

$$\vec{h}_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} \alpha_{ij}^k X^k \vec{h}_j\right) \quad (5)$$

where $\parallel$ represents a connection operation.

2.2 Temporal feature extraction based on BiGRU

GRU[17] is a simplified variant of long short-term memory (LSTM)[18], which can achieve almost the same effect as LSTM. Moreover, it can improve the training efficiency of the model to a great extent and solve the problem of gradient vanishing and gradient explosion. In the GRU network, the reset gate is helpful to capture short-term dependent features. The update gate is helpful to capture long-term dependent features.

BiGRU is composed of a forward hidden layer and a backward hidden layer. It solves the problem that GRU can only propagate forward and can better extract time-dependent features. The feature h extracted from the spatial GAT model is input into the BiGRU model for extracting time series features. The network structure of BiGRU is shown in Fig. 4, and the calculation is shown in Eq. (6).

$$\boldsymbol{O}_t = \boldsymbol{W}_t[\vec{\boldsymbol{X}}_t, \overleftarrow{\boldsymbol{X}}_t] \quad (6)$$

where $\boldsymbol{W}_t$ represents the output at t time step, $\vec{\boldsymbol{X}}_t$ represents the forward hidden layer output at t time step obtained by inputting the output feature h of spatial dimension into GRU network. Similarly, $\overleftarrow{\boldsymbol{X}}_t$ represents the backward hidden layer output at t time step.

BatchNorm2d is used to normalize the output results of each layer of spatial-temporal modules, and then the ReLU activation function is used to accelerate the convergence of the model. The calculation is shown in Eq. (7).

$$O_{ST} = \mathrm{ReLU}(\mathrm{BatchNorm2d}(O_t)) \quad (7)$$

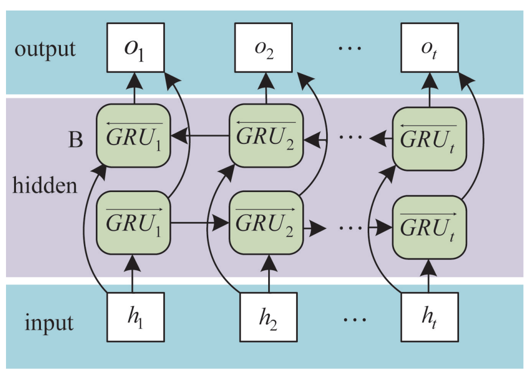


Fig. 4 BiGRU network structure

2.3 Connecting the output results of two types of road networks

The output results of the spatial-temporal modules of the real road network and the potential road network are input into the fully connected layer for dimension transformation. Then, the prediction results of the two paths are weighted and summed to obtain the final output results of the model, as shown in Eq. (8).

$$Output = W_1 \hat{Y}_{\mathrm{adj}} + W_2 \hat{Y}_{\mathrm{adp}} \tag{8}$$

where, W_1 and W_2 are the weight coefficient, $\hat{Y}_{\mathrm{adj}}$ is the output result of the real road network, and $\hat{Y}_{\mathrm{adp}}$ is the output result of the potential road network.

3 Experimental analysis

3.1 Datasets and data processing

Two datasets of real road networks are used in this paper.

(1) METR-LA, the Los Angeles highway dataset, was collected from March 1 to June 30, 2012.

(2) PEMS-BAY, a highway dataset in the Bay Area of California, USA, was collected from January 1 to May 31, 2017.

The specific parameters of the datasets are shown in Table 1.

Table 1 METR-LA and PEMS-BAY datasets

Dataset	Nodes	Time interval/min	Edges
METR-LA	207	5	1 515
PEMS-BAY	325	5	2 369

The datasets are divided into three parts: 70% training set, 10% validation set and 20% test set.

Since the traffic network is a topological structure reflected on the non-Euclidean distance, this paper constructs the adjacency matrix of the traffic network according to its features, and calculates the correlation between the nodes of the adjacency matrix by using the Gaussian kernel function according to the real distance between the nodes of the road network and generates the adjacency matrix. The calculation is as shown in Eq. (9).

$$W_{ij} = \begin{cases} \exp\left(-\dfrac{d^2(v_i, v_j)}{\sigma^2}\right) & \dfrac{d^2(v_i, v_j)}{\sigma^2} \geqslant k \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where σ represents variance and k represents a threshold, which is generally set as 0.1.

The time series of traffic speed is standardized, and the calculation is shown in Eq. (10).

$$x' = \frac{x - \mu}{\sigma} \tag{10}$$

where μ is the mean, and σ is the variance.

3.2 Parameter setting

In this model, three layers of spatial-temporal blocks are stacked. GAT is introduced in the spatial dimension, and the multi-head attention mechanism is used. The first two layers adopt concatenation multi-head attention, and the last layer adopts average multi-head attention; BiGRU is introduced in the temporal dimension. The specific parameters of the model are shown in Table 2. Adam optimizer is used to optimize the model parameters. To prevent gradient explosion, the clip_grad_norm_ method is used to carry out gradient clipping.

Table 2 Model parameters

Model	Model parameters	Value
Spatial dimension	Multi-head(concatenation head K)	4
	Multi-head (average head)	$K+2$
Temporal dimension	Hidden layer	2
	Hidden unit	200
Hyper-parameter	Learning rate	0.005
	Batch_size	128

This paper takes the PEMS-BAY dataset as an example to do a comparative experiment on the number of spatial-temporal block stacking layers. It can be seen from Table 3 that when predicting the traffic speed in the next 60 min, the residual block is 3 layers, the prediction performance of the model is the best.

Table 3 Comparison of stacking layers of spatial-temporal blocks

Stacked layers	*MAE*	*RMSE*	*MAPE/%*
2	1.89	4.35	4.52
3	**1.85**	**4.28**	**4.41**
4	1.87	4.33	4.48
5	1.89	4.36	4.51

Due to the multi-head attention mechanism adopted in this paper, a comparative experiment is conducted on the selection of K, and the most appropriate number of attention heads is selected according to the spatial features of the adjacency matrix. As can be seen from Table 4, when the number of heads of attention is 4, the prediction performance of the model is the best.

Table 4 Comparison of the number of multi-head attention

Number of head	*MAE*	*RMSE*	*MAPE/%*
2	1.88	4.33	4.43
4	**1.85**	**4.28**	**4.41**
6	1.87	4.32	4.42
8	1.89	4.33	4.52

3.3 Evaluation index

The loss function is used to evaluate the error between the predicted value of the model and the true value of the data set. The loss functions selected in this paper include mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The calculation of the three loss functions is shown in Eqs (11), (12) and (13).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i|^2} \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Y_i - \hat{Y}_i}{\hat{Y}_i}\right| \tag{13}$$

where, n represents the number of predicted samples, $\hat{Y}_i$ represents the predicted value of the model, and Y_i represents the true value of the dataset.

3.4 Results

3.4.1 Base lines

In order to evaluate the overall performance of this model, the following models are used as baseline model for comparison.

(1) ARIMA[19]: auto-regressive integrated moving average model with Kalman filter is widely used in time series prediction.

(2) FC-LSTM[20]: a recurrent neural network with fully connected LSTM hidden units is used.

(3) STGCN[21]: a complete convolution framework is constructed to solve the problem of time series prediction in the field of transportation.

(4) DCRNN[22]: a bidirectional random walk on the graph is used to capture spatial dependencies and an encoder-decoder architecture with planned sampling is used to capture temporal dependencies.

(5) TransGAT[23]: a node embedding algorithm based on dynamic graph attention mechanism is proposed to capture spatially dependent information. The temporal features captured by temporal convolutional network (TCN) are added by jumping connection to build a spatial-temporal module.

The prediction performance of the proposed model in the next 15 min, 30 min and 60 min is analyzed and compared with other benchmark models with the same dataset. The experimental results are shown in Table 5. It can be seen from the table that the proposed model in this paper has the lowest index value of MAE, RMSE and MAPE, and the prediction accuracy has been significantly improved.

(1) For traditional time series models (ARIMA[19], FC-LSTM[20]), only temporal dimension features were taken into account, but spatial dimension features of traffic network were not taken into account. The prediction performance of the model is inferior to that of the model combined with spatial and temporal features. For the prediction of the next 60 min, on METR-LA and PEMS-BAY datasets, compared with ARIMA and FC-LSTM, the MAE accuracy of this model is improved by 49.71%, 20.59% and 45.27%, 21.94% respectively, the RMSE accuracy of this model is improved by 46.49%, 18.53% and 34.15%, 13.71% respectively, the MAPE accuracy of this model is improved by 43.79%, 25.91% and 47.23%, 23.16% respectively. It can be seen that the predictive performance of the proposed model is significantly better than that of these time series models.

(2) For the spatial-temporal combined baseline models (STGCN[21], DCRNN[22]), the traffic speed in the next 60 min is predicted on METR-LA and PEMS-BAY datasets. Compared with STGCN and DCRNN, the MAE accuracy of this model is improved by 24.40%, 3.61% and 25.70%, 10.63% respectively, the RMSE accuracy of this model is improved by 24.68%, 6.84% and 24.78%, 9.70% respectively, the MAPE accuracy of this model is improved by 22.99%, 6.86% and 24.35%, 10.61% respectively. Therefore, the overall performance of the proposed model is better than other benchmark models in the extraction of spatial-temporal features, and has achieved prominent results in traffic flow prediction.

As shown in Fig. 5, this paper plot 15 min-ahead predicted values vs real values of SAdpGAT-BiGRU and classical model DCRNN on a snapshot of the test data. It can be seen that the curve fluctuation of the two models is consistent on the whole. However, the curve of DCRNN model (at about 11:00 am, 5:00 pm) produces a sharp fluctuation, which drastically deviated from the true values. Therefore, the proposed model has better fitting effect and higher robustness.

3.4.2 Computation time

As shown in Table 6, this paper compares the average training time and inference time of SAdpGAT-BiGRU and DCRNN on the same GPU. It can be seen that the time of each round of SAdpGAT-BiGRU is obviously faster than that of DCRNN, which shows that the proposed model has achieved better prediction effect on a simpler structure.

Table 5 Comparison of prediction results between the proposed model and the benchmark model

Dataset	Baseline	15 min			30 min			60 min		
		MAE	*RMSE*	*MAPE/%*	*MAE*	*RMSE*	*MAPE/%*	*MAE*	*RMSE*	*MAPE/%*
METR-LA	ARIMA	3.99	8.21	9.60	5.15	10.45	12.70	6.90	13.23	17.40
	FC-LSTM	3.44	6.30	9.60	3.77	7.23	10.90	4.37	8.69	13.20
	ST-GCN	2.88	5.74	7.62	3.47	7.24	9.57	4.59	9.40	12.70
	DCRNN	2.77	5.38	7.30	3.15	6.45	8.80	3.60	7.60	10.50
	TransGAT	2.71	5.20	7.69	3.15	6.16	8.70	3.48	7.21	9.80
	SAdpGAT-BiGRU	**2.68**	**5.09**	**6.89**	**3.06**	**6.06**	**8.30**	**3.47**	**7.08**	**9.78**
PEMS-BAY	ARIMA	1.62	3.30	3.50	2.33	4.76	5.40	3.38	6.50	8.30
	FC-LSTM	2.05	4.19	4.80	2.20	4.55	5.20	2.37	4.96	5.70
	ST-GCN	1.36	2.96	2.90	1.81	4.27	4.17	2.49	5.69	5.79
	DCRNN	1.38	2.95	2.90	1.74	3.97	3.90	2.07	4.74	4.90
	TransGAT	1.31	2.78	2.74	1.62	3.69	3.66	1.89	4.38	4.52
	SAdpGAT-BiGRU	**1.29**	**2.69**	**2.67**	**1.59**	**3.60**	**3.58**	**1.85**	**4.28**	**4.41**

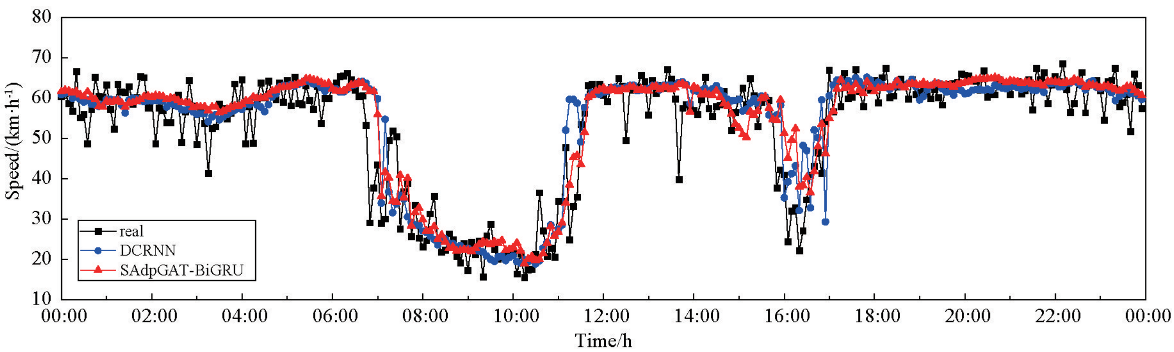


Fig. 5 Curve fitting of 15 min on METR-LA

Table 6 The computation time on the METR-LA dataset

Model	Training time/($s \cdot epoch^{-1}$)	Inference time/s
DCRNN	325.72	23.49
SAdpGAT-BiGRU	42.85	1.79

3.4.3 Multi-step predictive analysis

In order to further study the accuracy of the proposed model at different prediction steps, the nodes in METR-LA and PEMS-BAY datasets are selected, and the curve fitting graphs of the real values and predicted values of the nodes in the two datasets at different prediction steps are drawn respectively, as shown in Fig. 6 and Fig. 7. With the increase of prediction time step, the fitting degree of the speed curve between the predicted value and the real value of the node decreases slightly, but the fluctuation range of the curve is consistent as a whole, which shows that the model in this paper is effective in traffic speed prediction. Meanwhile, it can reflect the real trend of vehicle speed change, which is helpful to predict the changing traffic conditions.

3.4.4 Ablation experiment

In order to verify the effectiveness of each submodule of the proposed model, the experimental results of four model variants are analyzed and compared with SAdpGAT-BiGRU model.

(1) Remove_Muti-heads: remove the multi-head attention mechanism for extracting spatial features.

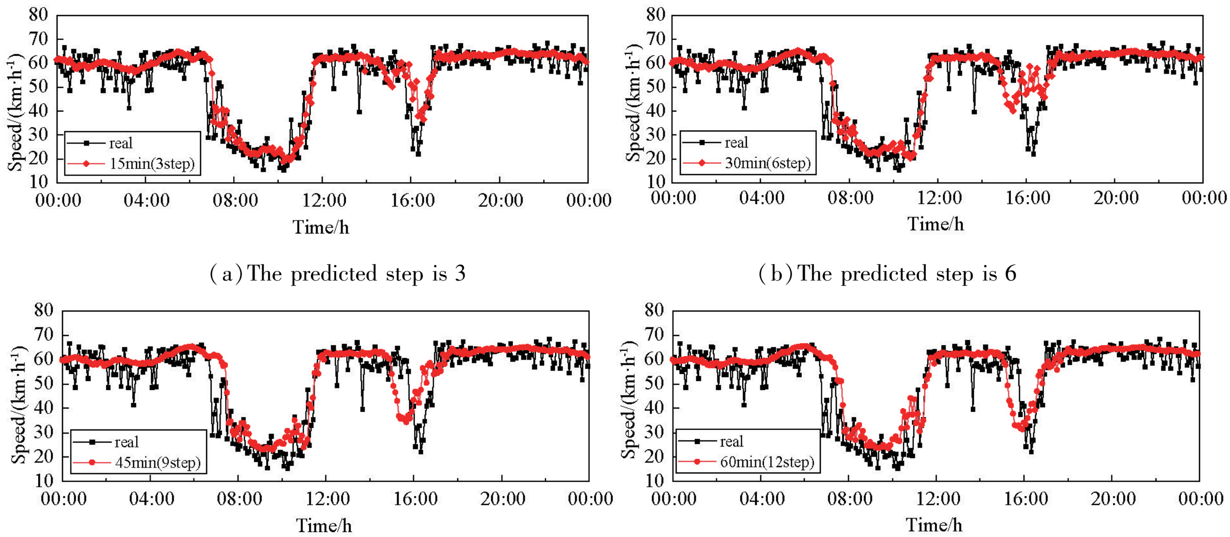
(2) Remove_Residual: remove residual mechanism, including spatial-temporal inter-block residual and intra-block residual.

(3) Potential_road: remove the part of extracting real road network features.

(4) Real_road: remove the part of extracting po-

tential road network features.

Fig. 8 shows the prediction results of the model and its variants in the future 60 min, using MAE, RMSE and MAPE evaluation functions respectively. From the curve fluctuation, it can be seen that each sub-module of the proposed model is effective for the whole model. In addition, the prediction results of both the real road network and the potential road network have declined, but the effective combination of the two extracted traffic network information can obtain better prediction results. Due to the spatial-temporal correlation and dynamic uncertainty of the traffic network, this paper not only considers the effect of the real network structure on the traffic speed prediction, but also considers the potential interaction between the nodes of the traffic network. The prediction results also prove the role of constructing adaptive adjacency matrix in the potential network.



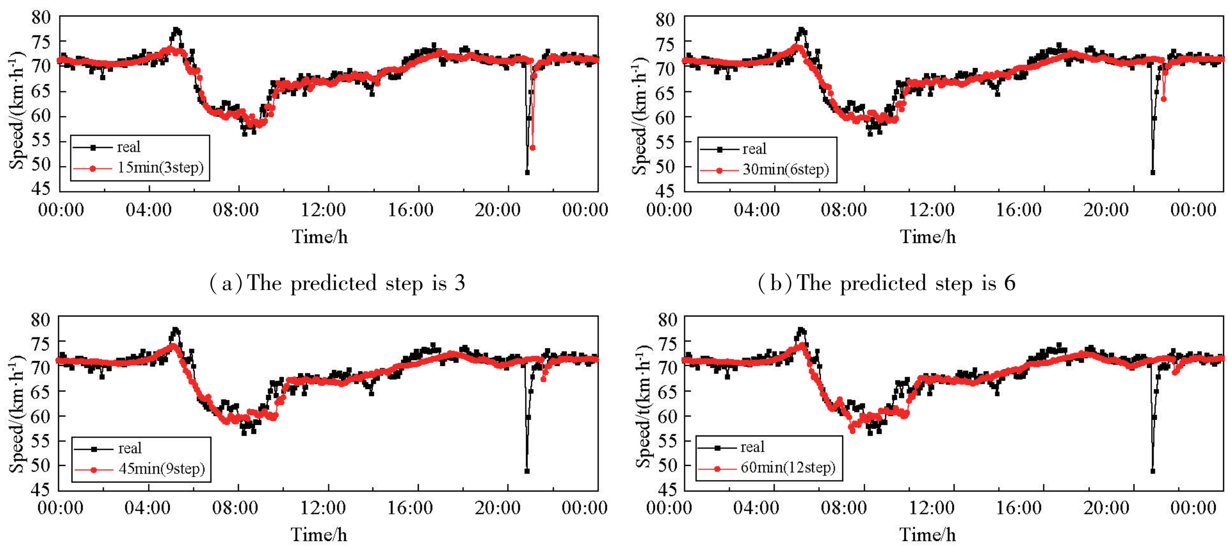
(a) The predicted step is 3

(b) The predicted step is 6

(c) The predicted step is 9

(d) The predicted step is 12

Fig. 6 Fitting curves of real and predicted values of nodes in METR-LA data set at different prediction steps



(a) The predicted step is 3

(b) The predicted step is 6

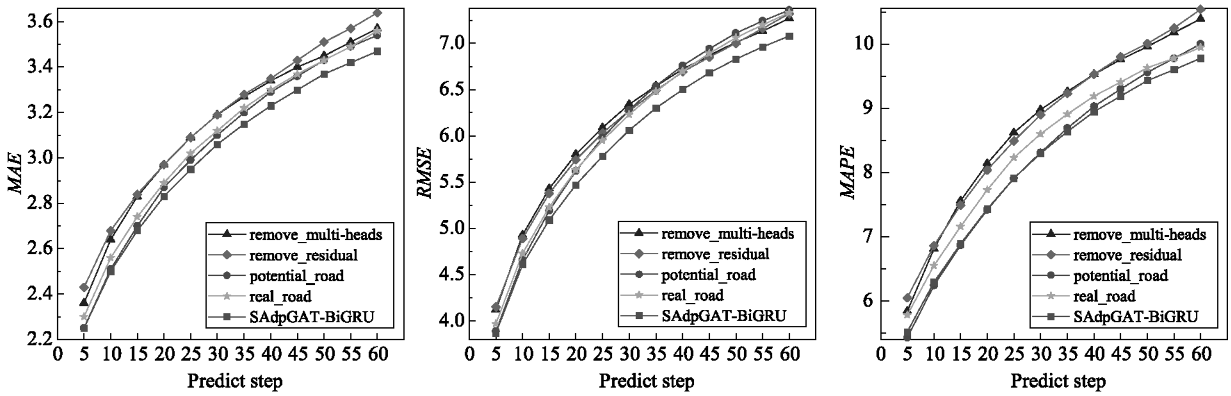
(c) The predicted step is 9

(d) The predicted step is 12

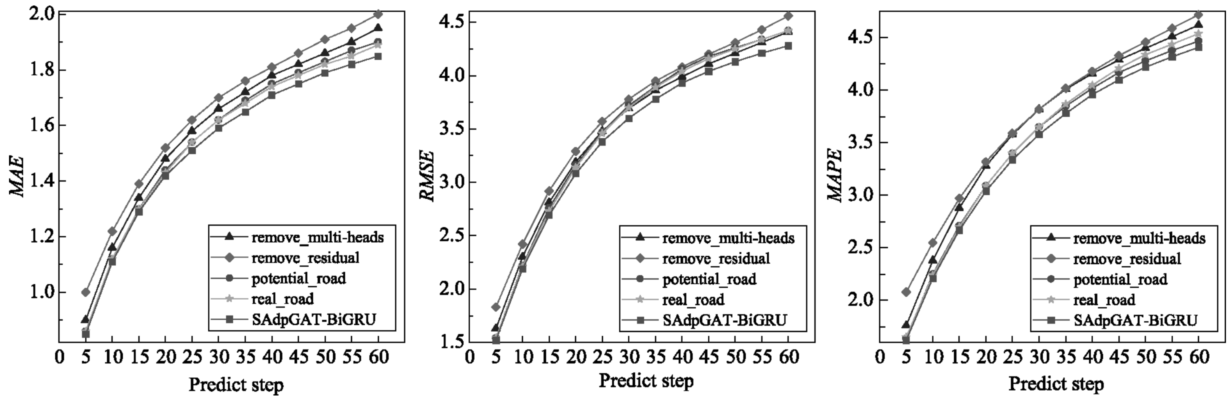
Fig. 7 Fitting curves of real and predicted values of nodes in PEMS-BAY data set at different prediction steps

4 Conclusion

This paper constructs adaptive adjacency matrix and uses GAT network to extract the spatial features of real road network and potential road network effectively, so that the model pays more attention to the correlation between nodes and improves the generalization ability of the model. BiGRU network is introduced to capture the time series features of the traffic network forward and backward to improve the prediction accuracy of the model. The experimental results show that the proposed model has good predictive performance on two real datasets. The next study will be to consider external factors such as weather.



(a) METR-LA



(b) PEMS-BAY

Fig. 8 Ablation study of two data sets for different evaluation functions

References

[1] YE J, ZHAO J, YE K, et al. How to build a graph-based deep learning architecture in traffic domain: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(5): 3904-3924.

[2] YU G, ZHANG C. Switching ARIMA model based forecasting for traffic flow[C]//2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada: IEEE, 2004: 1-4.

[3] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.

[4] XIE Y, ZHANG Y, YE Z. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition[J]. Computer-Aided Civil and Infrastructure Engineering, 2007, 22(5): 326-334.

[5] SUN S, ZHANG C, YU G. A Bayesian network approach to traffic flow forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2006, 7(1): 124-132.

[6] ZHANG L, LIU Q, YANG W, et al. An improved k-nearest neighbor model for short-term traffic flow prediction [J]. Procedia-Social and Behavioral Sciences, 2013, 96: 653-662.

[7] ZHANG W, YU Y, QI Y, et al. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning [J]. Transportmetrica A: Transport Science, 2019, 15(2): 1688-1711.

[8] ZHENG H, LIN F, FENG X, et al. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22 (11): 6910-6920.

[9] NARMADHA S, VIJAYAKUMAR V. Spatio-temporal vehicle traffic flow prediction using multivariate CNN and LSTM model[J]. Materials Today: Proceedings, 2023, 81: 826-833.

[10] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(9): 3848-3858.

[11] GUO S, LIN Y, FENG N, et al. Attention based spatial-

temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press, 2019: 922-929.

[12] ZHANG W, ZHU F, LV Y, et al. AdapGL: an adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks[J]. Transportation Research Part C: Emerging Technologies, 2022, 139: 103659.

[13] TA X, LIU Z, HU X, et al. Adaptive spatio-temporal graph neural network for traffic forecasting[J]. Knowledge-Based Systems, 2022, 242: 108199.

[14] ZHAO W, ZHANG S, ZHOU B, et al. STCGAT: spatio-temporal causal graph attention network for traffic flow prediction in intelligent transportation systems[EB/OL]. (2022-03-21) [2024-01-04]. https://arxiv.org/pdf/2203.10749v1.

[15] WANG C, TIAN R, HU J, et al. A trend graph attention network for traffic prediction[J]. Information Sciences, 2023, 623: 275-292.

[16] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04) [2024-01-04]. https://arxiv.org/pdf/1710.10903.

[17] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. (2014-09-03) [2024-01-04]. https://arxiv.org/pdf/1406.1078v3.

[18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[19] KAMARIANAKIS Y, PRASTACOS P. Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches[J]. Transportation Research Record, 2003, 1857(1): 74-84.

[20] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014: 3104-3112.

[21] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: AAAI Press, 2017: 3634-3640.

[22] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [EB/OL]. (2017-06-06) [2024-01-04]. https://arxiv.org/pdf/1707.01926v1.

[23] WANG T, NI S, QIN T, et al. TransGAT: a dynamic graph attention residual networks for traffic flow forecasting[J]. Sustainable Computing: Informatics and Systems, 2022, 36: 100779.

ZHANG Xijun, born in 1980. He received his Ph. D, M. S. and B. E. degrees in Lanzhou University of Technology in 2021, 2009 and 2003 respectively. His research interests include big data processing, data mining and analysis, embedded systems, machine learning and intelligent transportation system.