Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 2723101, 9 pages
https://doi.org/10.1155/2022/2723101

Research Article

Traffic Flow Prediction Based on Multi-Spatiotemporal Attention Gated Graph Convolution Network

Yun Ge, Jian F. Zhai, and Pei C. Su

Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences, Beijing 102488, China

Correspondence should be addressed to Yun Ge; gaiyun@ucass.edu.cn

Received 25 May 2022; Revised 14 July 2022; Accepted 21 July 2022; Published 9 September 2022

Academic Editor: Yong Zhang

Accurate prediction of traffic flow plays an important role in ensuring public traffic safety and solving traffic congestion. Because graph convolutional neural network (GCN) can perform effective feature calculation for unstructured data, doing research based on GCN model has become the main way for traffic flow prediction research. However, most of the existing research methods solving this problem are based on combining the graph convolutional neural network and recurrent neural network for traffic prediction. Such research routines have high computational cost and few attentions on impaction of different time and nodes. In order to improve the accuracy of traffic flow prediction, a gated attention graph convolution model based on multiple spatiotemporal channels was proposed in this paper. This model takes multiple time period data as input and extracts the features of each channel by superimposing multiple gated temporal and spatial attention modules. The final feature vector is obtained by means of weighted linear superposition. Experimental results on two sets of data show that the proposed method has good performance in precision and interpretability.

1. Introduction

With the development of urbanization process, people's demand for transportation are increasing day by day. Whether to build an effective transportation system has become an important factor in restricting development of city. Accurate prediction about traffic condition plays a very important role in people's daily travel planning, urban traffic planning, and traffic management and strategy. In order to improve the efficiency of transportation and reduce the time cost for transportation activities in daily work and life, this paper proposed a traffic speed prediction model based on multi-spatiotemporal gated graph convolutional network with attention mechanism.

Based on analyzing the urban road's traffic flow situation, the velocity of vehicles on the road can be predicted. Traffic speed prediction can not only provide managers with scientific decision-making information but also provide appropriate route guidance for urban travelers, which is an important guarantee for the unimpeded flow of urban traffic. Currently, the main traffic speed prediction model can be divided into three categories: the statistical-based methods, the machine learning-based methods, and deep learning-based methods. The statistical-based methods are constructed based on the theory of statistical forecasting and mainly contain the historical average analysis prediction method, regression difference moving average method [1] (ARIMA), Kalman filtering method [2, 3], the grey prediction model method, etc. These models usually have strict requirements on input data and these corresponding algorithm structures are relatively fixed. However, the prediction result of traffic flow can be easily affected by some random interference factors, such as traffic accidents, weather, and special events, which can make the prediction accuracy relatively low. The second type is the method based on machine learning way, which can not only model the nonlinear feature of traffic flow but also continuously adjust the model parameters by means of adaptive learning methods to obtain more accurate prediction results. Therefore, the methods based on machine learning gradually replace the statistical theory-based methods and become the next research focus in traffic flow prediction field.

Algorithms used for prediction mainly include support vector machine [4], K-nearest neighbor [5], Bayesian network [6], and other methods. The third category is the way based on deep learning model, which is also the most common used method at present. Deep learning methods can be used to learn the features about input data without mankind intervention. Such models have strong requirements on nonlinear mapping characteristic feature and less strict requirements on data than those model-driven methods, so they will be more suitable to model the uncertainty status of traffic flow and improve the prediction precision. Because of these advantages, some researchers have applied deep learning methods into the field of traffic prediction and achieved remarkable progress.

Shao et al. [7] applied the Long Short-Term Memory Network (LSTM) model into traffic flow prediction and improved the accuracy of flow prediction by calculation of the spatial characteristics. Liu et al. [8] used the Gated Recurrent Unit (GRU) model to predict urban traffic flow. Since the internal neural cell number of the GRU model is less than that of LSTM, the prediction performance is still good. Traffic flow data not only has dynamic correlation in time but also has strong dynamic correlation in space. In order to extract the temporal and spatial features effectively, Shi et al. [9] proposed Conv-LSTM model, which comprehensively uses CNN and LSTM to capture the spatiotemporal feature. Liu et al. [10] applied it to short-term traffic prediction. Yao et al. [11] put forward the spatiotemporal dynamic network (STDN) model and used CNN and LSTM to capture the spatiotemporal feature. Zhang et al. [12] proposed the spatiotemporal residual network (ST-ResNet) model which uses different residual units to model the information of time proximity, periodicity, and tendency.

Zhao et al. [13] proposed a temporal graph convolutional network (T-GCN) model based on combining GCN model with GRU model. The GCN model was used to learn complex topological structure for capturing spatial feature, and GRU model was used to learn temporal feature of traffic flow changing data. Yu et al. [14] proposed a spatiotemporal graph convolutional network (STGCN) model, which uses one-dimensional CNN model to capture the time dynamic feature and the GCN model was used to obtain the spatial feature of local traffic data. In order to capture the dependence between temporal and spatial feature, Li et al. [15] improved the gated GRU unit and proposed diffused convolution gated loop unit (DCGRU). Combined with encoder and decoder, the DCRNN model for Seq2Seq was proposed. In view of the traffic flow data being time-dependent, Guo et al. [16] used three different components to extract feature from historical data. Song et al. [17] used three different continuous time slices to construct local spatiotemporal models and used sliding windows to segment time periods into three parts. By stacking multiple graph convolution layers, a spatiotemporal synchronous graph convolution network (STSGCN) was established to extract long-term spatiotemporal feature. Although the T-GCN model uses two-layer graph convolution network to aggregate the spatial information about two level neighbors, it still ignores deeply excavating the spatial correlation between higher-order neighbor nodes. Therefore, K-order Chebyshev graph convolution which can cover k-order neighbor nodes was used to complete the spatial convolution operation and extract the spatial feature of higher-order neighbor nodes. In addition, T-GCN model uses a single time series to perform prediction work without mining time dependence between different slices. The spatiotemporal information can also be used in other fields. Wang et al. [18] use spatiotemporal correlation information to reconstruct traffic data. Wang et al. [19] perform passenger flow prediction via dynamic hypergraph convolution networks. Yu et al. [20] proposed a low-rank dynamic mode decomposition model for short traffic flow prediction.

Although these methods have been able to predict traffic flow very precisely, there are still some areas that can be improved. The existing methods can be improved from the following two aspects: improving the scope of neighborhood scale and considering the influence of data with different time periods on future traffic. The traffic flow status in any node on the traffic network can be affected not only by the first-level neighborhood nodes, but also by the second-level neighborhood nodes. The change rule of traffic flow is periodic. The traffic flow on the road is generally large during working hours, and small during other times. Traffic information with different time periods has different influence on the status change of traffic flow in the future. It is of great help to improve the prediction of traffic flow to comprehensively consider the changing rules of traffic flow in different time periods.

Therefore, this paper extracts three different time series datasets which are monthly data, daily data, and weekly data to fully capture temporal characteristics. In general, this paper proposes a multichannel gated spatiotemporal graph convolution with attentional mechanism, which puts three different time series datasets into the model and gets the feature by stacking multiple gated spatiotemporal blocks. The forecasting work was finished by combining all the three different feature vectors with the help of weighted linear combination operation. The main contributions of this paper can be summarized as follows:

(i) We developed a multichannel gated spatiotemporal graph convolution network to learn the dynamic feature of traffic flow data. Specifically, a multichannel feature extraction and fusion framework was proposed. The temporal feature of the traffic data was fully exploited.

(ii) A novel spatiotemporal calculation module was designed by adding attention mechanism. It helps the model to pay more attention to import the feature in each channel.

(iii) Extensive experiments are carried out on read traffic data, which can verify the effectiveness of the model proposed in this paper. The performance of this

prediction model has a certain progress compared to existing methods.

The rest of this paper is organized as follows. Section 2 describes the related work on traffic flow forecasting and the development of graph neural networks. Section 3 introduces the detailed architecture of proposed forecasting network with gated graph neural network and attention. Section 4 presents the experiment setting and the experiment results. Finally, Section 5 concludes the work and presents the findings of this research.

2. Related Work

In this section, we will briefly introduce corresponding theories and definitions referring to the proposed model.

2.1. Graph Neural Network. Convolutional neural network is a feed-forward neural network based on convolutional operation, which can efficiently compute feature information from structured data such as image, speech, and text. However, there are a lot of unstructured data in daily life, such as social network data, human skeleton data, traffic flow data, and other data without regular structure. The traditional CNN models cannot effectively model such unstructured data. In order to effectively capture the local spatial feature of these data, a graph convolution network model for unstructured data was proposed. Graph convolutional network is a kind of neural network structure which is popular in recent years. It is a kind of neural network which extends the convolution operation to graph structure data. Compared with traditional convolutional network models which can only be used in structured data computation, graph convolutional networks are special in capturing unstructured data. The road network structure in reality is typical unstructured data. Thus, the local feature of traffic data can be extracted effectively based on using graph neural network.

The existing graph convolution operation-based methods mainly can be divided into two types: the way based on spatial domain and the way based on frequency domain. The spatial domain-based operation can be defined by aggregating the feature information about adjacent nodes in the graph. The frequency domain-based operation uses Fourier transform to realize the convolution calculation in frequency domain.

According to graph theory, the properties of graph structure can be obtained by calculating Laplacian eigenvalue and eigenvector about adjacency matrix, and the spectrum convolution result on graph can be obtained by calculating the convolution of signal $x \in R^N$ and graph convolution kernel g_θ. The purpose of graph convolution is to predict the state of the node at the next moment according to current status of the node in a graph, which can be defined as

$$H^{(l+1)} = f\left(H^{(l)}, A\right), \tag{1}$$

where $H^{(l)}$ denotes all the note status at time l, A denotes the adjacent matrix, and $f(\cdot)$ denotes mapping function. Different mapping function represents different GCN models. Usually, the node status in the next moment can be obtained by calculating the linear combination of its adjacent notes through multiplying the adjacent matrix with the current status matrix, and the final expression can be defined as follows:

$$H^{(l+1)} = \sigma\left(AH^{(l)}W\right). \tag{2}$$

The weight matrix was used to perform linear mapping operation and the function $\sigma(\cdot)$ was used to calculate the nonlinear mapping operation. The function of the adjacency matrix and state matrix multiplication was used to calculate the addition of adjacency nodes in a matrix manner. However, the information of node itself has not been taken into account. The direct way to solve this problem is adding an identity matrix to the adjacency matrix so as to add the self-loop information of each node into the adjacency matrix. In addition, with the accumulation of the GCN operations, the dimension difference of status information between nodes in the graph will become large. In order to maintain the stability of the operation, the matrix information needs to be normalized before each calculation. Graph convolution operators usually adopt graph Laplacian matrix as the substitution of adjacency matrix, and graph convolution calculation function can be defined as follows:

$$H^{(l+1)} = \sigma\left(D^{-(1/2)}\widetilde{A}D^{-(1/2)}H^{(l)}W^{(l)}\right), \tag{3}$$

where $\widetilde{A} = A + I_N$ represents a new adjacency matrix with self-loop information, $\widetilde{D} = \sum_i \widetilde{A}_{ij}$, $H^l \in R^{N\times F}$ denotes the nodes information in l-th layer, $H^0 = X$, X denotes the initial status of graph nodes, and $W^l \in R^{F\times F}$ denotes the weight value in the l-th layer. Each calculation of graph convolution is the extraction of first-order neighborhood information. Multiorder neighborhood information can be realized by superimposing several convolutional layers.

2.2. Spatiotemporal Attention Mechanism. Graph convolutional neural network can capture the local spatial correlation between adjacent nodes in graph, but different adjacent points have different impact on the current node. The key idea of spatial attention mechanism is to pay adaptive attention to the characteristics of the most relevant nodes according to the input data. In time slice, the information of road network is changing dynamically all the time. Therefore, using spatial attention mechanism and temporal attention mechanism to adaptively capture the node information with higher correlation in each dimension will be of great help to improve the prediction accuracy.

In this paper, soft attention [21] mechanism is used to calculate attention weight. It can extract features from the input sequence and adaptively calculate the importance of each node from the road network information at different time. Firstly, the information of all nodes at time t was aggregated into a vector. The aggregated information

includes the spatial characteristics and node information of the road network at time t can be expressed as follows:

$$q_t = \text{relu}\left(\sum_{i=1}^{N} W h_{ti}\right), \tag{4}$$

where W denotes the trainable parameters and h_{ti} denotes hidden state value of the i-th node in time t. The attention values about all nodes can be formulated as follows:

$$\alpha_t = \text{Sigmoid}\left(U_s \tanh\left(W_h h_t + W_q q_t + b_s\right) + b_u\right), \tag{5}$$

where $\alpha_t = (\alpha_{t1}, \alpha_{t2}, \ldots, \alpha_{tN})$ and α_{ti} denotes the attention value of the i-th nodes at t time. U_s, W_h, and W_q denote trainable parameters; b_s and b_u denote bias vector. This attention mechanism firstly spliced the aggregated information of all nodes at time t with the information of all nodes at the same time and then obtained the attention weight of each node relative to all nodes through the full connection layer. In order to calculate the nonlinear mapping information of nodes at different time, this paper uses the structure of two fully connected layers to calculate attention value. The second hidden state h_{ti} of the i-th node at t time can be calculated by $(1 + \alpha_{ti}) \cdot h_{ti}$ and the weighted graph state will be input into next layer.

2.3. Gated Convolution Network. Gated linear unit was proposed by Dauphin et al., which is a convolutional neural network model with gated mechanism. This model was mainly used to replace the recurrent neural network in natural language processing model. Compared with the gated unit in RNN model, this unit has the advantages of lower complexity, faster gradient propagation efficiency, and being less prone to gradient disappearance or gradient explosion. In addition, the gated linear unit can also process the input data in parallel, which can improve the accuracy of the model as well as the computational efficiency. Let X denote layer input, h_l denote output of this layer which also represents the hidden states of this layer, W and V denote two different convolution cores, and b and c denote two bias parameters; the gated convolution model can be expressed as follows:

$$h_l(X) = (X * W + b) \otimes \sigma(X * V + c). \tag{6}$$

The output of the model was realized through dot product calculation between linear mapping result vector and nonlinear mapping vector. The linear mapping vector can be obtained by multiplying the input vector X with parameter vector W. The nonlinear mapping vector can be calculated by multiplying the input vector X with parameter vector V at first. Then, the nonlinear mapping function can be obtained by using nonlinear function $\sigma(\cdot)$. Because the output of function $\sigma(\cdot)$ can only be 1 or 0, the function of multiplying these two vectors is to perform gated selection operation for each node.

3. Methodology

In this section, we will describe the framework of the proposed method. The traffic flow information was extracted in three channels separately. In each channel, there are two spatiotemporal blocks to fetch space-temporal feature. Each ST block is composed of a spatial block and a temporal block which are used to fetch the spatial feature and temporal feature separately. The input of the three channels corresponds to the traffic flow data containing three impassable periods, respectively. The model structure is shown in Figure 1.

3.1. Problem Definition. The goal of traffic prediction is to predict the traffic information in a certain time based on the historical traffic information on the road. This paper takes traffic speed forecasting as the main objective of the study. This prediction work is performed based on traffic flow data on the road which was collected by traffic sensors distributed throughout the network. Typically, traffic flow data refers to the number of vehicles that pass through a sensor during a specified period of time. The topology structure composed of all sensors in the road network was defined as $G = (V, E, A)$. The vector $V = \{v_1, v_2, \ldots, v_N\}$ denotes vertex set. Assume that only one sensor was placed on each road and the road in the road network can be represented by the sensor. Let N denote the number of the codes and E denote the set of edges in the network. The adjacent matrix $A = R^{N \times N}$ was used to denote the connection between nodes. The feature matrix $X_t \in R^{N \times P}$ denotes the flow status in time t, and P denotes the length of feature vector. The traffic flow prediction problem can be defined as follows: given the traffic flow's status at the time t and other historical data, the $t+1$ time traffic flow data can be calculated in the form of the following equation:

$$\left[X_{t+1}, X_{t+2}, \ldots, X_{t+p}\right] = f\left(G; \left(X_{t-n}, X_{t-n+1}, \ldots, X_t\right)\right), \tag{7}$$

where t is the length of historical time series and n is the length of time series that need to be predicted.

3.2. Graph Convolution on the Traffic Data. Since the structure of the road network is an irregular structure, the traffic flow data generated by vehicles on the road network is also irregular, and it is very suitable to use GCN model to calculate the feature of traffic flow. Because the standard graph convolution computation is too huge, the Chebyshev inequality was often used to get the approximate solutions, and the approximate equation can be formulated as follows:

$$\theta * \varphi x = \theta(L)x \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x, \tag{8}$$

where θ is the graph convolution kernel, $T_k(\tilde{L}) \in R^{N \times N}$ is k-order Chebyshev inequality, $L = 2(L/\lambda_{\max}) - I_N \in R^{N \times N}$,

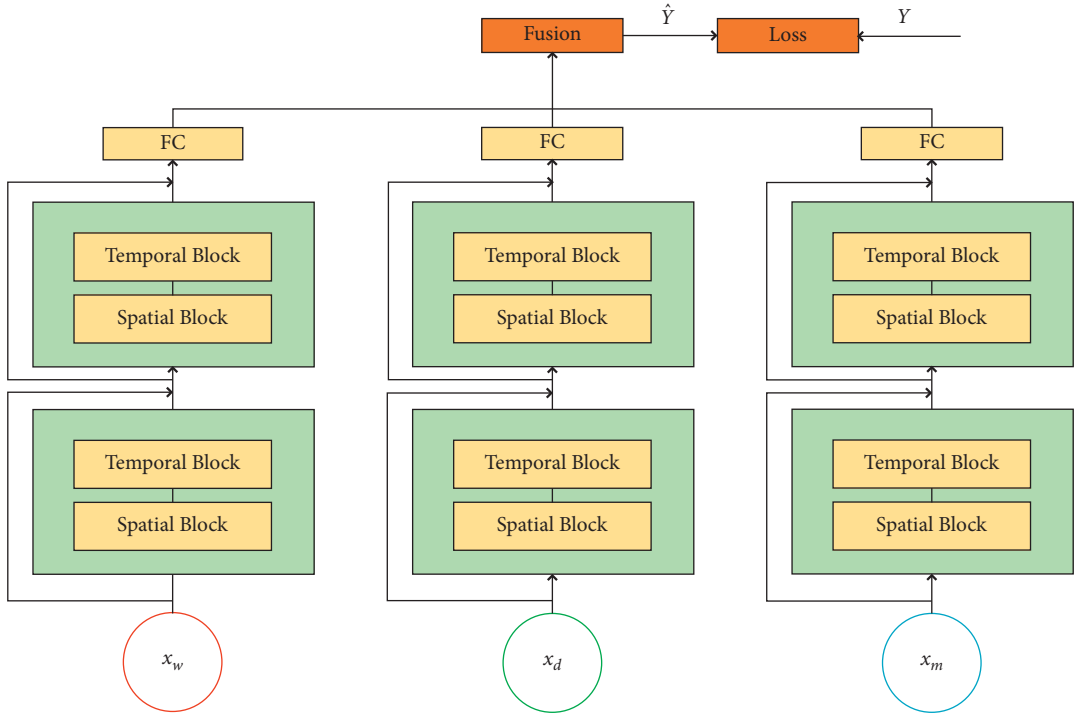

Figure 1: The framework of the proposed model.

$\lambda_{\max}$ is the largest eigenvalue of the Laplace matrix, k is the size of convolution kernel, and the k-th Chebyshev inequality's recursive definition is $T_k(\tilde{L}) = 2xT_{k-1}(x) - T_{k-2}(x)$, while $k = 0$ and $T_0(x) = 1$.

In order to effectively learn local spatial dynamic feature, the spatial attention matrix $W \in R^{N\times N}$ was multiplied with the k-order Chebyshev inequality $T_k(\tilde{L})$ based on dot product. The concrete equation can be formulated as follows:

$$\theta * \varphi x = \theta(L)x \approx \sum_{k=0}^{K-1} \theta_k \left(T_k(\tilde{L}) \odot W\right)x. \tag{9}$$

In this paper, k-order Chebyshev inequality was applied to extract feature of road network information. The k-order convolution operator of Chebyshev graph convolution can cover the features of k-order neighborhood nodes.

3.3. Multiperiod Flow Data Series. In order to capture the temporal dynamic characteristics of traffic flow, this paper uses three different spatiotemporal components to extract the characteristics of historical traffic data. This paper constructs three different traffic flow data sequences with three different periods: week, day, and hour.

3.3.1. The Weekly Periodic Series. The weekly periodic series data X_w was composed of traffic data sampled in weeks. They have the same weekly properties and time intervals as the forecast period. In terms of the variation trend and peak value of traffic conditions, the traffic flow on weekdays is similar to that on weekdays, but not on nonweekdays. Therefore, training with weekly periodic data can help us capture differences between weekdays and nonweekdays data.

3.3.2. The Daily Periodic Series. The daily periodic series X_d was composed of the traffic data sampled in days. Due to the regularity of people's activity track, the traffic flow shows periodic fluctuation. For example, the morning and evening rush hours on weekdays may have similar traffic volumes. Therefore, daily correlation data were added to extract temporal and spatial dynamic correlation.

3.3.3. The Minutely Periodic Series. The minutely periodic series X_m was composed of the traffic data sampled in minutes.

The sequence that has the greatest impact for the future traffic is the traffic situation in the adjacent period. If the current traffic flow on the adjacent road is large, the possibility of congestion at the next moment of this section will be large.

All these three data series have the same structure and can be calculated in the same way. There are two spatiotemporal blocks in the model and a fully connected layer in the end. The spatiotemporal block was composed of spatial block and temporal block. Each block has an attention module. In order to avoid the decrease of training accuracy,

we introduce residual learning module between spatiotemporal blocks. In the end of forecasting model, the outputs of the three channels will be merged by a parameter matrix to form the final feature vector.

3.4. Gated Convolution for Feature Extracting. Graph convolution model can be used to extract spatial information of traffic data effectively. However, traffic flow data is a typical time flow data. Effectively extracting the characteristics of traffic flow information on the time axis is of great help to improve the accuracy of prediction. In this paper, gated convolution model is used to extract temporal and spatial features of traffic information. Compared with RNN model, the gated convolution model has a simpler structure and smaller computational time. In order to capture the characteristics of traffic data on the time axis, we apply gated convolution operation on each time axis to capture the dynamic characteristics of traffic flow data.

3.5. Multichannel Data Merge. Each spatiotemporal convolution module consists of a graph convolution module for spatial information and a gated convolution model for time domain. The gated convolution module captures the features of the time axis along the time axis. The outputs of different channels have different weight in prediction. In this paper, we combine them based on linear combination operation. The fusion equation is shown as follows:

$$\widehat{Y} = W_w \odot \widehat{Y}_w + W_d \odot \widehat{Y}_d + W_m \odot \widehat{Y}_m, \tag{10}$$

where $\odot$ denotes the element-wise Hadamard product, $\widehat{Y}_w$ denotes the output of the channel weekly data, $\widehat{Y}_d$ denotes the output of the channel daily data, and $\widehat{Y}_m$ denotes the output of the channel minutely data. W_w, W_d, and W_m are weighted parameters corresponding to different channel data. In this paper, we take 0.4, 0.2, and 0.4 as the default weight parameter values, because traffic flow status in former time has more impact on the traffic data in the next time.

3.6. Loss Function. The goal of model training is to minimize the error between the actual traffic speed and the predicted value on the road. In this paper, Y_t and $\widehat{Y}_t$ were used to represent the actual traffic speed and predicted speed, respectively. The loss function of MSTAGCN was shown in the following equation:

$$\text{loss} = \left\| Y_t - \widehat{Y}_t \right\| + \lambda L_{\text{reg}}. \tag{11}$$

In this formulation, the first term was used to measure the error between the actual speed and the predicted value. The second term represents L_{reg}, and the regularization term, which helps to avoid the overfitting problem, is a hyperparameter.

4. Results and Discussion

4.1. Datasets. The experiment datasets used in this paper are PeMS04 and PeMS08 which belong to Caltrans performance evaluation system (PeMS, https://www.pems.dot.ca.gov). The geographic information and time information are contained in the data. The PEMS04 is the traffic flow data collected from San Francisco Bay, which includes 3,848 sensors on 29 roads. We pick out the experiment data from 307 sensors. The time range of the dataset is from January 1 to February 28 in 2018 which covers 59 days. The PEMS08 was the traffic flow data collected from SAN Bernardino, which includes 1,979 sensors on 8 roads. We pick out the data from 170 sensors as experiment data.

4.2. Data Preprocessing. The data in these two datasets are sampled in every five minutes. Each sensor contains 288 data records per day, and each record contains three features. They are the traffic flow, average vehicle speed, and occupancy rate responding to sensors during that time period. The spatiotemporal data were divided into training set, validation set, and test set in the ratio of 6 : 2 : 2. At the same time, range normalization was carried out for each feature to keep the data value between [0,1]. The specific calculation formula is as follows:

$$x^* = \frac{x - \min}{\max - \min}. \tag{12}$$

By using the distance between different sensors, the adjacency matrix of the graph was established using the threshold Gauss kernel. The calculation process of the threshold Gaussian kernel can be formulated as follows:

$$W_{ij} = \begin{cases} e^{\left(\text{dist}\left(v_i, v_j\right)^2 / \sigma^2\right)}, & \text{dist}\left(v_i, v_j\right) < s, \\ 0, & \text{dist}\left(v_i, v_j\right) \geq s, \end{cases} \tag{13}$$

where W_{ij} represents the weight of the edge between sensor v_i and sensor v_j, $\text{dist}(v_i, v_j)$ represents the distance between sensor v_i and sensor v_j, σ^2 is the variance of the distance, and s is the threshold. As there are almost no sensors over 1000 meters in the dataset, the threshold s is 1000.

4.3. Evaluation Metrics Subheadings. To evaluate the performance of the proposed model, we choose three metrics to evaluate the difference between real traffic value Y_t and estimated value $\widehat{Y}_t$, which was shown in the following equations.

(1) Root Mean Square Error (RMSE) is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_j^i - \widehat{y}_j^i\right)^2}. \quad (14)$$

(2) Mean Absolute Error (MAE) is calculated as follows:

$$\text{MAE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|y_j^i - \widehat{y}_j^i\right|. \quad (15)$$

(3) Mean Absolute Percentage Error (MAPE) is calculated as follows:

$$\text{MAPE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\frac{y_j^i - \widehat{y}_j^i}{y_j^i}\right|, \quad (16)$$

where y_j^i denotes the real traffic flow data value in the i-th time, $\widehat{y}_j^i$ denotes the predict value, M denotes the number of samples, and N denotes the number of roads. Specifically, the rule of metrics measuring prediction error is as follows: the smaller the error, the higher the accuracy of the prediction.

4.4. Experiment Settings. To verify the validity of the model, the MSTAGCN model proposed in this paper was compared with the classical GRU model and the recently proposed DCRNN, T-GCN, ASTGCN, and STSGCN models. Table 1 shows the hyperparameter settings of each model and the word layers means the number of hidden layers. The word units represents the number of computing units in each hidden layer and all models in the experiment are composed of the same number of units. k denotes the order of graph convolution, and T_w, T_d, and T_m represent the length of weekly, daily, and minutely sequence.

4.5. Experiment Result. The experimental results are shown in Table 2. In The PEMS08 dataset, MSTAGCN model is always superior to other benchmark models in terms of accuracy. In EMS04 dataset, MST-GCN has the smallest prediction error compared with other forecasting methods and has slightly larger errors in MAE and MAPE result. In the RMSE evaluation results, the proposed method has larger error than STSGCN methods. Due to the simplest model structure, the GRU model has the worst performance in both datasets. The lower prediction results of the former spatial analysis-based model demonstrate that those methods have not effectively model the nonlinear information of the traffic data. In general, the deep learning-based methods have better performance than those non-deep learning models and the convolution operation plays an importance role in improving the accuracy of prediction. The convolution operation can effectively capture the local feature in both the spatial information and the temporal information. Simultaneously using spatial and temporal information is the other effective prediction improving routine. As we can see, the last four methods have better performance than other methods. Besides, the MSTAGCN performs better than other methods, indicating that the multichannel mechanisms applied in the proposed model are effective in capturing the changing routine characteristic. Our MSTAGCN achieves better performance than the previous models proving the feature about traffic changing is nonlinear and single input information cannot provide sufficient information for feature learning.

Figures 2 and 3 exhibit the prediction performance on these two datasets. The GRU model only considers the temporal characteristics and does not take advantage of the spatial information of road network. The accuracy of GRU is not as good as that of temporal correlation method. GRU only considers the temporal correlation and does not use the spatial correlation of road network, so the accuracy of GRU is not as good as that of the method using temporal and spatial correlation. The DCRNN and T-GCN model spatial and temporal feature information separately, but they only use a single time window to extract long-term dependence without considering impaction caused by the periodicity of different time windows. ASTGCN and STSGCN both use different spatiotemporal components to extract corresponding correlation from time windows, but they ignored the correlation between different time period channels. So, the prediction precision will be relatively reduced. In this paper, the MSTAGCN method considers impaction from different periodic data on traffic forecasting work and uses multichannel structure to fuse the spatiotemporal components, so as to capture the long-term spatiotemporal dependence between different periodic traffic data. Therefore, the prediction accuracy of the proposed model is better than that of the existing models, and the prediction effect is better.

Table 1: Hyperparameter settings for different models.

Models	Layers	Units	k	T_w	T_d	T_m
GRU	3	500	—	—	—	5
DCRNN	2	64	3	—	—	5
T-GCN	3	64	2	—	—	5
ASTGCN	2	64	3	24	12	5
STSGCN	4	64	3	—	—	12
MSTAGCN	3	64	3	2	6	5

Table 2: Performance comparison of different models of traffic flow prediction.

Model	PEMS04			PEMS08		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
GRU	24.34	43.47	16.59	19.01	35.12	13.23
DCRNN	24.06	34.7	16.00	19.36	31.94	11.18
T-GCN	23.71	34.74	16.37	22.98	32.57	11.88
ASTGCN	22.36	32.6	15.18	18.21	27.99	13.22
STSGCN	22.52	34.62	14.92	17.79	26.33	11.80
MSTAGCN	**22.11**	**32.96**	**14.15**	**15.85**	**23.62**	**11.44**

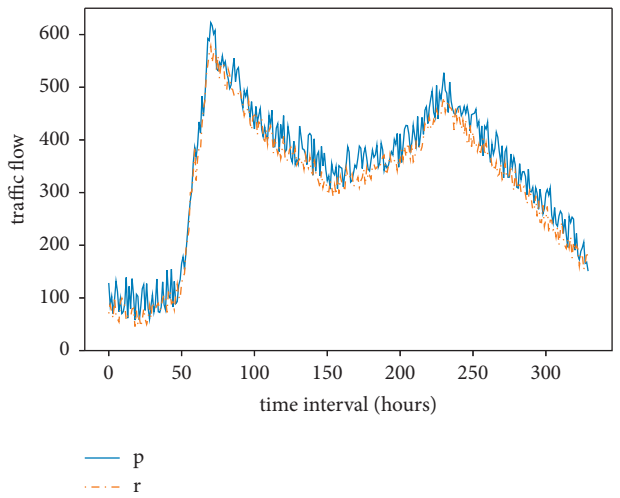

Figure 2: Prediction result on PEMS04.

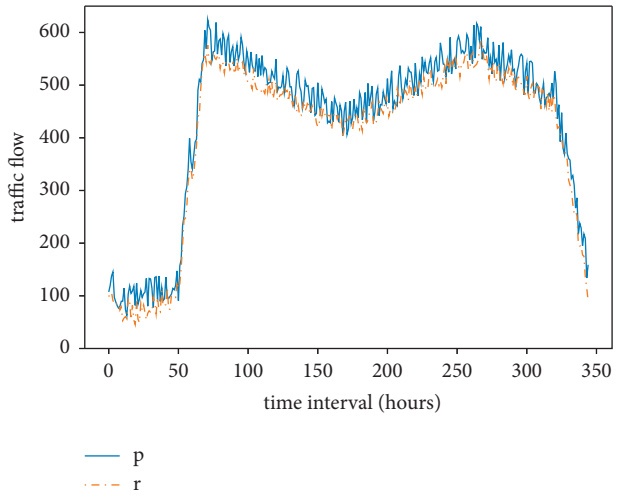

Figure 3: Prediction result on PES08.

5. Conclusions

In this paper, we proposed a multi-spatiotemporal attention gated graph convolution network (MSTAGCN) to capture the spatiotemporal feature about traffic flow data. Firstly, in order to deeply explore the temporal and spatial correlation of nodes, the Chebyshev convolution and gated loop unit were combined to obtain a larger receptive field. Secondly, three periodicity datasets with different time window were picked up to provide comprehensive traffic information. Finally, the MSTAGCN model was constructed by fusing multiple spatiotemporal components with encoder-decoder network structure. The experimental results about highway datasets PEMS04 and PEMS08 in Caltrans performance evaluation system show that the performance of the new model is significantly better than other models, and it can be applied to the actual road network to improve traffic prediction precision efficiency. In the next step, datasets about urban road networks will be collected to explore the adaptability of the model under complex urban road networks.

Data Availability

All data and program files included in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant no. NSFC 61602486).

References

[1] C. Han, S. Su, and C. Wang, "Real-time adaptive prediction of short-term traffic flow based on ARIMA model," *Journal of System Simulation*, vol. 16, no. 7, pp. 1530–1532, 2004.

[2] S. V. Kumar, "Traffic flow prediction using kalman filtering technique," *Procedia Engineering*, vol. 187, pp. 582–587, 2017.

[3] Q. Shen, Y. Wang, and Y. Huang, "Grey verhulst-markov model with improved initial value and its application," *Statistics & Decisions*, vol. 36, no. 7, pp. 30–33, 2020.

[4] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2001–2013, 2018.

[5] C. X. Tao and M. Xie, "Short-term traffic flow prediction method based on nonparametric regression of k-nearest neighbors," *Systems Engineering Theory and Practice*, vol. 30, pp. 376–384, 2010.

[6] J. Wang, W. Deng, and J. Zhao, "Short-term traffic flow forecast based on bayesian network multi-method combination," *Transportation Systems Engineering and Information*, vol. 11, no. 4, pp. 147–153, 2011.

[7] H. X. Shao and B. H. Soong, "Traffic flow prediction with long short-term memory networks (lstms)," in *Proceedings of the 2016 IEEE Region 10 Conference*, 2016, Article ID 29862989.

[8] M. Liu, J. Wu, and Y. Wang, "Traffic flow prediction based on deep learning," *Journal of System Simulation*, vol. 30, no. 11, pp. 4100–4105, 2018.

[9] X. J. Shi, Z. R. Chen, H. Wang, Y. Dit-Yan, W. Wai-kin, and W. Wang-chun, "Convolutional Lstm Network: A Machine Learning Approach for Precipitation Now Casting [EB/OL]," 2015, https://arxiv.org/abs/1506.04214.

[10] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Nanjing, China, 2017.

[11] H. X. Yao, X. F. Tang, H. Wei, G. Zheng, and Z Li, "Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction," *AAAI*, vol. 33pp. 5668–5675, Honolulu, 2019.

[12] J. Zhang, Y. Zheng, and D. Qi, "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1655–1661, AAAI, San Francisco, USA, 2016.

[13] L. Zhao, Y. J. Song, C. Zhang et al., "T-GCN: a temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.

[14] B. Yu, H. T. Yin, and Z. X. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, June 2018.

[15] Y. Li, R. Yu, S. Cyrus, and L. Yan, "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," in *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, Vancouver, BC, Canada, 2018.

[16] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 922–929, Palo Alto, CA, 2019.

[17] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 914–921, Palo Alto, CA, 2020.

[18] Y. Wang, Y. Zhang, X. Piao, H. Liu, and K. Zhang, "Traffic data reconstruction via adaptive spatial-temporal correlations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1531–1543, 2019.

[19] J. Wang, Y. Zhang, Y. Wei, Y. Hu, X. Piao, and B. Yin, "Metro passenger flow prediction via dynamic hypergraph convolution networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7891–7903, 2021.

[20] Y. Yu, Y. Zhang, S. Qian, S. Wang, Y. Hu, and B. Yin, "A low rank dynamic mode decomposition model for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, 2021.

[21] A. Vaswani, N. Shazeer, N. Parmar, N. G. Aidan, K. Lukasz, and P. Illia, "Attention is all you need," 2017, https://arxiv.org/abs/1706.03762?context=cs.