STGMN: A gated multi-graph convolutional network framework for traffic flow prediction

Qingjian Ni[1] [ORCID] · **Meng Zhang[1]**

Abstract

Accurate traffic flow prediction is crucial for the development of intelligent transportation. It can not only effectively avoid traffic congestion and other traffic problems, but also provide a data basis for other complex tasks. The rapid development of social technology and the increasingly complex traffic environment lead to the emergence of massive traffic data. Traffic flow prediction as a spatial-temporal prediction problem has been widely concerned, but the traditional forecasting methods often ignore the spatial-temporal dependence, difficult to meet the prediction requirements. Therefore, this paper proposes a novel spatial-temporal model based on an attention one-dimension convolutional neural network (1D-CNN) and a gated interpretable framework, which models historical traffic data from the perspectives of time and space respectively. The core of the model proposed in this paper is to construct spatial-temporal blocks. First, a 1D-CNN based on channel attention mechanism and "inception" structure is proposed to extract temporal correlation. Then, considering the complexity of the actual traffic network, an interpretable multi-graph gated graph convolution framework is proposed to extract the spatial correlation. Finally, extensive experiments are carried out on real data sets, which prove the effectiveness of the proposed model, and it is very competitive compared with some state-of-the-art methods.

Keywords Traffic flow prediction · Graph convolution · Spatial-temporal correlations

1 Introduction

With the continuous development of society, the increasingly complex traffic situation has brought great challenges to the construction of an intelligent transportation system (ITS). The growth of population and the widespread popularity of transportation tools have led to traffic jams, traffic accidents, and other social events. As a key part of intelligent transportation, traffic flow prediction can effectively help solve similar incidents and coordinate traffic management. Traffic flow refers to some traffic flow states on the road composed of pedestrians, running vehicles and roads, etc. By using historical traffic flow data to predict the future, it can help people avoid congestion during the journey and choose convenient and safe routes. The research of traffic flow prediction started very early. Currently, the problem of traffic flow prediction based on spatial-temporal data is widely concerned by researchers [1–3].

In the past few decades, scholars have proposed many methods to predict traffic flow [4], which can be broadly divided into two categories: 1. Traditional prediction models based on statistical methods. 2. Prediction models based on machine learning methods. The first category includes AR (auto-regressive model), ARMA (auto-regressive moving average model), ARIMA (auto-regressive integrated moving a sverage model) [5]. Kumar et al. [6] proposed a traffic flow forecasting model based on seasonal ARIMA (SARIMA), after the model performs the necessary differentiation to stabilize the input time series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) were plotted to identify the appropriate order of the SARIMA model. The second category includes support vector machines (SVM), artificial neural networks (ANN) and Bayesian networks, etc. [7, 8]. This kind of method can well capture the complex nonlinear relationship in traffic time series and has been widely studied in the past years. Li et al. [9] proposed a new dynamic radial basis function (RBF)

✉ Qingjian Ni
  nqj@seu.edu.cn

  Meng Zhang
  220191763@seu.edu.cn

[1] School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

neural network to predict outbound passenger volume and improve passenger flow control. However, with the development of society and technology, the traditional shallow neural network models are not performing well in the face of the increasingly complex traffic networks and huge traffic data volume, and they are usually only applicable to the traffic prediction of a single station. In the face of spatial-temporal data, they can not well mine the spatial-temporal correlation. While the rise of deep learning has just solved this problem. Through the continuous attempts of researchers, deep learning has achieved a huge breakthrough in the field of traffic flow prediction [10]. Models such as deep neural network (DNN), convolutional neural network (CNN), recurrent neural network (RNN) have been well applied in the field of traffic. Deep learning models can excavate the temporal and spatial correlation of traffic data [11–15]. Yu et al. [16] proposed a spatial-temporal recurrent convolutional network (SRCN), which inherited the advantages of deep convolutional neural networks (DCNN) and long short-term memory networks (LSTM), through DCNN to obtain traffic network dependence, LSTM to obtain time dependence. However, although most current deep learning-based methods can improve the mining ability of nonlinear relations through complex structures, there are still many shortcomings in the acquisition of spatial correlation. Although RNN can well mine the changes in time series, it has the problems of slow training and easy to ignore the spatial correlation of data [17]. Although CNN can extract the relevance in temporal and spatial at the same time through the convolution kernel, it organizes the traffic network into the form of grids and ignores the spatial information brought by the network structure.

In the problem of spatial-temporal prediction of traffic flow, the traffic network can be regarded as a graph structure. Compared with CNN, which decomposes the network into a grid, the graph neural network (GNN) proposed in recent years [18] has received extensive attention from scholars because it can better mine the spatial correlation in the traffic data [19–21], such as T-GCN [22], STGCN [23], ASTGCN [24], DCRNN [25]. These models construct graph models based on real traffic networks to obtain spatial dependence of traffic data. Fang et al. [26] proposed the global temporal and spatial n network (GSTNet), which constructed a spatial-temporal block that contained a multi-resolution time module and a globally related space module to obtain the temporal and spatial correlation respectively. Song et al. [27] proposed that the synchronous modeling mechanism of the spatial-temporal synchronous convolutional network model (STSGCN) can effectively capture complex spatial-temporal information while ignoring the heterogeneity in data.

In order to better capture the dependencies of complex spatial-temporal traffic data, this paper proposes a spatial-temporal gated multi-graph network (STGMN). The model contains multiple spatial-temporal blocks. Each spatial-temporal block obtains multi-level temporal correlation by constructing an attention 1DCNN based on the inception mechanism, and captures complex spatial correlation by constructing a graph convolution framework based on gated multi-graphs.

From the perspective of temporal dependence, the traffic flow tends to change periodically, and the traffic flow will also be affected by the previous moments. Therefore, we use filters of different sizes in the model and control their output through the attention mechanism, which can supplement the time dependence from multiple scales. From the perspective of space, as the traffic network becomes more and more complex, the change of traffic flow is obviously affected by its topological structure, and the traffic flow data between adjacent stations and stations with upstream and downstream relationships are obviously closely related. However, in past studies, researchers usually only build the model with one kind of graph network. In this paper, a multi-graph mechanism is proposed, and different graphs are convoluted and stacked through an interpretable framework, which greatly improves the performance of the graph convolution model.

In summary, the contributions of this article include:

1. A novel spatial-temporal prediction model is proposed, it integrates temporal and spatial modules to extract temporal and spatial correlations respectively.
2. A multi-resolution CNN framework based on attention mechanism is proposed as a temporal block. In this module, convolution kernels of different sizes are used to extract the spatial dependence of traffic data at different scales, and the features obtained are fused and adjusted by the channel attention mechanism.
3. An interpretable multi-graph gated graph convolution framework is proposed as a spatial block. In this module, the deep spatial information of traffic data is gradually mined through the graph convolution layer of different graph structures by residual stacking. Also, a gating mechanism is proposed to control the output.
4. This paper verifies the feasibility of the model on real datasets, and compares the prediction results of the proposed model with the results of some models proposed in recent years, and proves that the model in this paper has obtained better results.

The content of this article is organized as follows: Section 2 describes and defines the traffic flow problem and introduces the graph convolutional network. Section 3 details the model proposed in this article. Section 4 conducts experiments on the proposed model on real datasets and compares it with other models. Section 5 contains a summary of the article and discusses future work.

2 Preliminaries

In this section, we give a brief description of traffic flow and traffic flow prediction. At the same time, the application of current deep learning methods (especially graph convolution) in traffic flow prediction is analyzed and discussed.

2.1 Traffic flow prediction problem

As an important part of intelligent transportation system (ITS), traffic flow prediction has been concerned and studied by researchers. The traffic flow prediction problem is essentially a spatial-temporal sequence prediction problem, which uses historical traffic data to predict the future.

Use $X_i \in \mathbb{R}^{N \times C}$ to represent the data of N stations at time point i, where C represents the number of features. Use graph $G = (V, E)$ to denote the topological structure of the transportation network, V denotes the stations on the road, E represents the connectivity between stations.

The goal of the traffic flow prediction problem is to learn a mapping function f, which can predict future data based on the input historical data and graphs. In summary, a traffic flow problem that uses historical data with a length of n in the past and traffic network information to predict future T steps can be expressed as follows:

$$\left[X_{t-n+1}, X_{t-n+2}, ..., X_t, G\right] \xrightarrow{f} \left[X_{t+1}, X_{t+2}, ..., X_{t+T}\right] \tag{1}$$

Figure 1 represents the acquisition process of a data set. Data samples were constructed based on known historical traffic flow data, and continuous 2T traffic data were taken, of which T data were taken as input (x) and the other data as output (y). In addition, considering the importance of the traffic network relationship in the traffic flow prediction problem, the graph information containing site correlations is also used as part of the input.

It should be noted that traffic flow forecasting, as a spatial-temporal forecasting problem, is facing the following challenges: 1. As a forecasting problem, how to better capture the time dependence; 2. How to combine the real traffic network and forecasting to mine the spatial relationship. 3. How to model the influence of other factors, such as weather, holidays.

2.2 Graph convolution network for traffic flow prediction

In recent years, the use of deep learning methods to solve complex traffic flow prediction problems has received widespread attention [28]. In the past, researchers usually used the traditional convolutional neural network based on map grids to deal with spatial dependence [29, 30], but obviously, the real transportation network is a more complex graph topology, and the number of neighbors at each station may be different. CNN which cannot process data with a non-Euclidean structure, cannot effectively extract features under such a network structure. Therefore, to make convolution applicable to graph-structured data, graph convolution is proposed. In recent years, researchers have begun to integrate graph convolution into traffic flow prediction framework to solve the spatial dependence problem [31].

2.3 Graph convolution network

GCN can aggregate the features of each node with those of its neighborhood, which means, it can aggregate the local features in the spatial-temporal network. Therefore, graph convolution operation can make good use of graph relations. A basic graph operation [32] can be expressed in the following (2).

$$H^{l+1} = g(X, A) = \sigma\left(\hat{A} H^l W_0\right) \tag{2}$$

Where $H^l \in \mathbb{R}^{N \times C \times T}$ represents the input. Let a graph be defined as $G(V, E)$ with vertices V and edges E, edge weights are given as e_{ij}. A represents the adjacency matrix, which is a matrix representing the adjacency between vertices, $A_{ij} = e_{ij}$. $\tilde{D}$ represents the corresponding degree matrix, consisting of node degrees. $\tilde{A} = A + I_N$ is the graph plus the

Fig. 1 Acquisition process of data sets
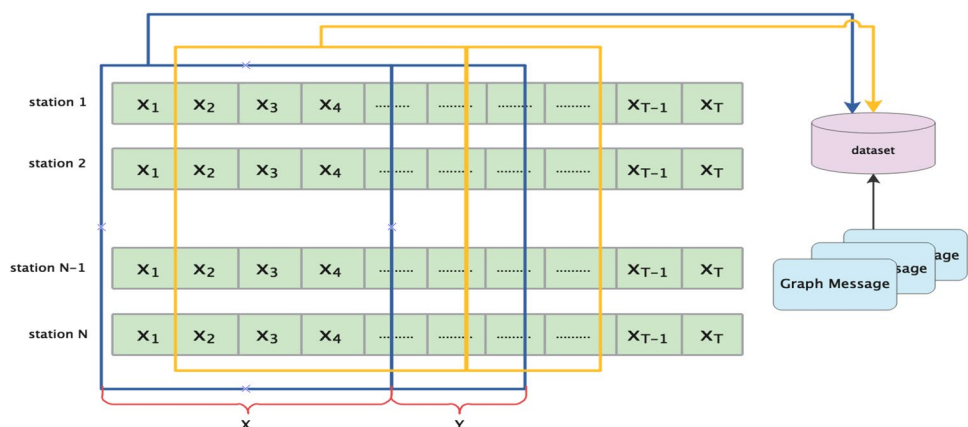
self-connected adjacency matrix, $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ represents the normalized graph, $W_0 \in \mathbb{R}^{C \times C'}$ is the neural network parameter, and σ is the activation function, such as ReLU.

3 Proposed model

Section 3 introduces the proposed model STGMN in detail. The model is mainly composed of stacked spatial-temporal blocks, which mainly include two parts: temporal module and spatial module. The temporal module extracts multi-scale time dependence through multi-resolution 1DCNN, and adaptively adjusts channel weight through channel attention. The spatial module uses an interpretable multi-graph convolution framework to mine the effective information from the relationships between stations.

3.1 The architecture of STGMN

The architecture of STGMN model proposed in this paper is shown in Fig. 2. The model mainly includes two spatial-temporal blocks and an output block. In STGMN, each spatial-temporal block includes a temporal module based on attention multi-resolution convolutional network and a spatial module based on gated multi-graph convolutional network. The output block contains an attention-based convolution layer and a complete connection layer.

The novelty of the proposed model is that it separately constructs modules to extract spatial-temporal and temporal dependencies. For the mining of temporal relationships, a channel-focused multi-resolution CNN model is used to improve the temporal data processing capability of CNN. For more complex spatial relationships, an interpretable multi-graph framework is used, which deeply considers the spatial information of the traffic network. The detailed structure of each module is described in the following sections.

3.2 Temporal module

In the temporal dimension, the traffic conditions at different moments are different for the future. For CNN, the time range of extraction is determined by specifying the size of the filter, but the correlations obtained by models are different in different time ranges, so this paper proposes a multi-resolution feature extraction approach.

The idea of "Inception" structure [33] is to increase the width of the network by using filters of different sizes to obtain features of different scales. Based on this idea, a multi-resolution convolutional network module combined with channel attention mechanism is proposed in this paper. It mainly consists of a multi-scale convolutional module, a channel attention module, and an output module. Its specific structure is shown in Fig. 2. First of all, the multi-scale convolutional module selects one-dimensional CNN to extract time dependence, and in order to improve the mining ability of CNN, filters of different sizes are used for operation. Then, the dependencies of the output channels are modeled by channel attention. Finally, an output module containing the convolution layer and the attention layer is used to control the length of the output sequence of the temporal module.

3.2.1 Multi-resolution convolutional network

In recent years, CNN, RNN, and other models have been widely applied to the analysis of time series with good results. RNN makes data processing with the order, gives the network memory, very suitable for processing sequence data. However, the structure of CNN is simple, and when one-dimensional CNN is used to process the sequence data, it can not only effectively mine the temporal correlation, but also greatly improve the training speed. Therefore, this paper chooses to use CNN to obtain the temporal dependence of traffic flow. At the same time, to further excavate the time relationship in the sequence, the idea of "Inception" structure is introduced, and a multi-resolution convolutional neural network model is adopted, in which filters of different sizes are used to obtain features of different scales.

We use multi-resolution blocks to capture the temporal relationship. The key idea is to increase the width of the network and make the convolutional layer have filters of different sizes, which can help to better capture the temporal

Fig. 2 The model structure of STGMN
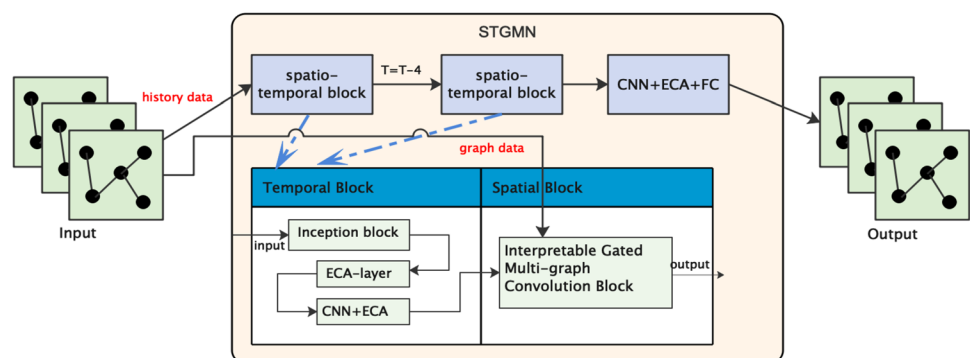
dimension dependence of traffic data. The input data of the data block is processed by using filters of different sizes. At the same time, the residual connection mechanism is introduced to splice the input and different convolution calculation results together as the output. The structure of the multi-resolution block is shown in Fig. 3.

Figure 3 shows the multi-resolution block structure of the proposed model's two spatial-temporal blocks. The first spatial-temporal block contains three filters: $1 \times 1, 1 \times 3$, and 1×5. Because the sequence length T decreases after passing through the first spatial-temporal block, there is no filter of size 1×5 in the multi-resolution block-2. Taking the multi-resolution block-1 as an example, the formula is as follows:

$$X^{(r)} = ReLU\left(\Phi_1 * X^{(r-1)} \oplus \Phi_2 * X^{(r-1)} \oplus \right.$$
$$\left. \Phi_3 * X^{(r-1)} \oplus X^{(r-1)}\right) \quad (3)$$

Where $X^{(r-1)} \in \mathbb{R}^{C_{r-1} \times N \times T_{r-1}}$ represents the input, $X^{(r)} \in \mathbb{R}^{C_r \times N \times T_r}$ represents the output, $*$ represents the convolutional operation, Φ_1, Φ_2, Φ_3 respectively represent three different sizes of filters, $(C_r/2 \times 1 \times 1)$, $(C_r/4 \times 1 \times 3)$ and $(C_r/4 \times 1 \times 5)$, and by padding ensure $T_r = T_{r-1}$.

3.2.2 Attention based convolutional network

Multi-resolution block combines the calculation results of different filters only in the form of the splice, so the channel attention mechanism (ECA) is introduced to capture the correlation between channels [34]. Channel attention has great potential in improving the performance of deep convolutional neural networks. The local cross-channel interaction strategy without dimension reduction proposed by ECA-Net can not only learn effective channel attention, but also greatly reduce the complexity of the attention model. The weights of each channel in ECA are calculated as follows:

$$\alpha = \sigma(W_k y) \quad (4)$$

$$g\left(X^{(r)}\right) = \frac{1}{NT} \sum_{i=1, j=1}^{N,T} X_{i,j}^{(r)} \quad (5)$$

Where $g(X^{(r)}) \in \mathbb{R}^C$ is the global average pool (aggregated feature), σ is the Sigmoid function, W_k is a $C \times C$ parameter matrix, which is set as follows:

$$\begin{bmatrix} w^{1,1} & \cdots & w^{1,1} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w^{2,2} & \cdots & w^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w^{C,C-k+1} & \cdots & w^{C,C} \end{bmatrix} \quad (6)$$

The weight calculation of channel y^i only considers its interaction with k neighbors, and the formula is as follows:

$$\alpha_i = \sigma\left(\sum_{j=1}^{k} w_i^j y_i^j\right), \ y_i^j \in \Omega_i^k \quad (7)$$

where Ω_i^k indicates the set of k adjacent channels of y_i.

A more efficient approach would be to have all channels share the same weight.

$$\alpha_i = \sigma\left(\sum_{j=1}^{k} w^j y_i^j\right), \ y_i^j \in \Omega_i^k \quad (8)$$

The ECA module obtains cross-channel interaction locally, so it needs to set an appropriate K value manually, and manual verification will consume a lot of computing resources. Therefore, an adaptive strategy is proposed in ECA-NET. Given channel dimension C, neighborhood size k can be adaptively determined by:

$$k = \psi(C) = \left|\frac{\log_2}{\gamma} + \frac{b}{\gamma}\right|_{odd} \quad (9)$$

where γ and b are set to 2 and 1, respectively.

ECA module can effectively improve the performance of convolutional neural network. Combined with the Inception block in this paper, ECA module can reflect the value of the dependence of different time scales and improve the mining ability of the model in the temporal dimension. We will get the weight to adjust the input, the calculation formula is as follows:

$$\hat{X}^{(r)} = \omega X^{(r)} = \omega(X_1, X_2, ..., X_{C_r}) \in \mathbb{R}^{C_r \times N \times T_r} \quad (10)$$

Fig. 3 Architecture of Multi-resolution blocks
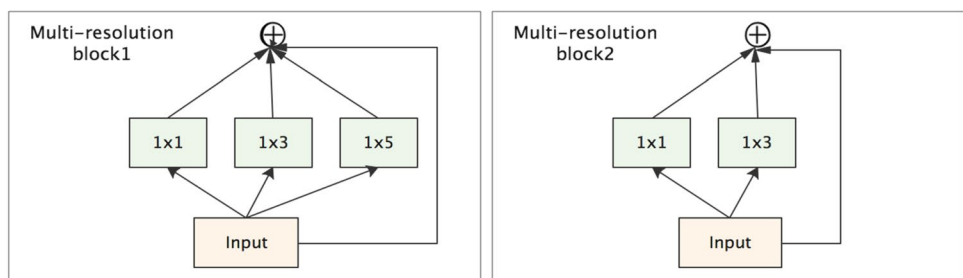
3.2.3 Summary of the temporal module

In the multi-resolution temporal module proposed in this paper, a multi-scale filter is used to process the data in order to find the temporal correlation in the data. A lightweight channel attention mechanism is introduced to analyze the importance of each channel and give different weights to each channel without burdening the network structure. Finally, a CNN layer is added to compress the data in the time dimension and serve as the input of the subsequent spatial module.

3.3 Spatial module

In the spatial dimension, for transportation networks, traffic conditions at neighboring locations affect each other. In recent studies, attention has been paid to the help of spatial relationships for traffic flow prediction, but the use is still relatively homogeneous and often limited to inter-site connectivity, but complex traffic networks often contain deeper dependencies, so this module aims to explore how to better use spatial information for data feature mining.

In this paper, an innovative Spatial module of multi-graph convolutional neural network is proposed, which combines the architecture of fully connected deep neural network (N-BEATS) proposed in [35] with the mechanism of multi-graph structure. Its architecture is shown in Fig. 4. The idea of the module is to stack the graph convolution blocks using different graph information based on the residual connection and to control the output of each small module through the gating mechanism. Next, we will analyze the module from the perspective of multi-graph mechanism, stack mechanism, and gating mechanism.

3.3.1 Multi-graph mechanism

When using graph convolution to capture spatial-temporal correlations in traffic data, it is very important to construct an appropriate graph. However, in the face of the complex traffic network environment, one graph is obviously not enough to describe the relationship between stations, so this section shows the different types of correlation between traffic networks and how to build a graph based on these correlations. The following three graph generation methods are used in the model:

(1) Distance graph

First, the most basic distance relationship is used to describe the relationship between the stations. In the real traffic network, the connection and distance between the stations are very valuable for reference. The traffic flow variations of stations close to each other may be similar, so the distance is chosen to represent the weight between stations. The calculation formula for the weights w_{ij} of the distance matrix A_{dis} is as follows:

$$w_{ij} = \begin{cases} exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & i \neq j \ and \ exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \epsilon \\ 0, & otherwise \end{cases} \tag{11}$$

Where σ^2 and ϵ are thresholds used to control matrix distribution and sparsity. We use the reciprocal of distance to describe the relationship between stations. The closer the distance between stations is, the greater the weight is; otherwise, the smaller the weight is.

(2) Correlation graph based on DTW

In addition to using the distance relationship between stations to construct the graph matrix, we also propose to use the historical data relationship of each station to construct the graph and use DTW (Dynamic Time Warp Distance) algorithm to calculate the correlation between the historical data of stations. The basic idea of DTW is to regularize the time axis through the numerical similarity of time series, and then to find the optimal corresponding relationship between the two time series.

When using historical data for each station to evaluate similarities between stations, there is often a "time lag" problem.

Fig. 4 Architecture of Spatial module. The center of the graph is the core of the multi-graph mechanism, which can receive multiple graph structures and connect stacked convolution blocks through residuals. The right side of the graph is a gated structure, which controls the output of the multi-graph structure part
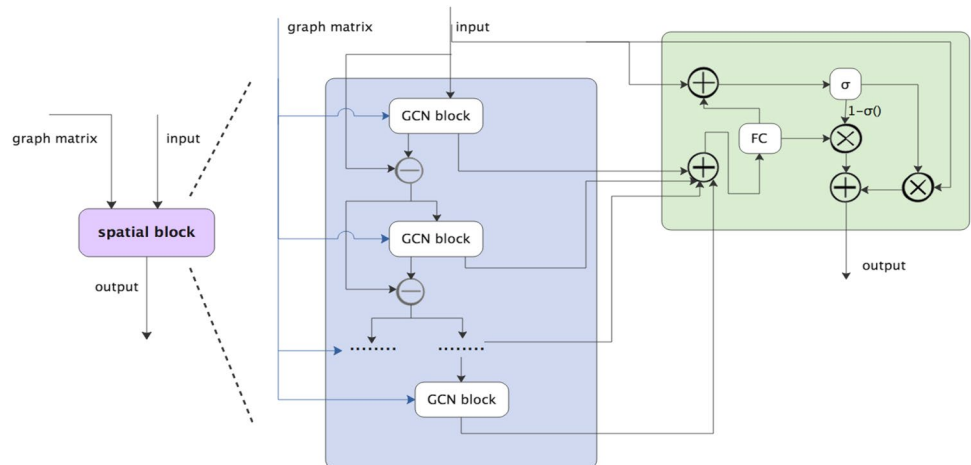
Although the two stations have similar fluctuations on the traffic flow change curve, the different geographical locations will cause different time steps. Assume that there are historical traffic flow data curves for two stations in Fig. 5, two curves can be observed on the change if there are a lot of similarities, but their fluctuations have phase difference, if only used Euclidean distance to measure their relationship, obviously the result is bad. Therefore, DTW can be a good solution to his problem. DTW is defined as follows, there are two time series, $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ respectively, where m and n are the lengths of X and Y. The cumulative distance is constructed as follows:

$$D(i,j) = d(x_i, y_j) + min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\} \tag{12}$$

Where $d(x_i, y_j) = \sqrt{(x_i - y_j)^2}$, denotes the distance between x_i and y_j. For the final process of cumulative distance is, in fact, in the distance matrix B to find an optimal path W, to minimum cumulative distance and the distance measure is defined as:

$$DTW(X, Y) = min\{D(m, n)\} \tag{13}$$

W has also been called a warping path, which is a correspondence between two sequences. W must satisfy the constraints of boundary, continuity and monotony, (1) Boundary constraints: the starting point of path W is $W_1 = (x_1, y_1)$ and the end point is $W_K = (x_K, y_K)$; (2) Constraints of continuity and monotony: if $W_k = (x_k, y_k)$, then W_{k+1} must be one of $(x_{k+1}, y_k), (x_{k+1}, y_{k+1}), (x_k, y_{k+1})$.

After calculating the shortest cumulative distance D between the historical data of each station by DTW, the larger the D is, the smaller the correlation between the sequences is. Therefore, $1/D$ is taken as the weight of our correlation graph Ar.
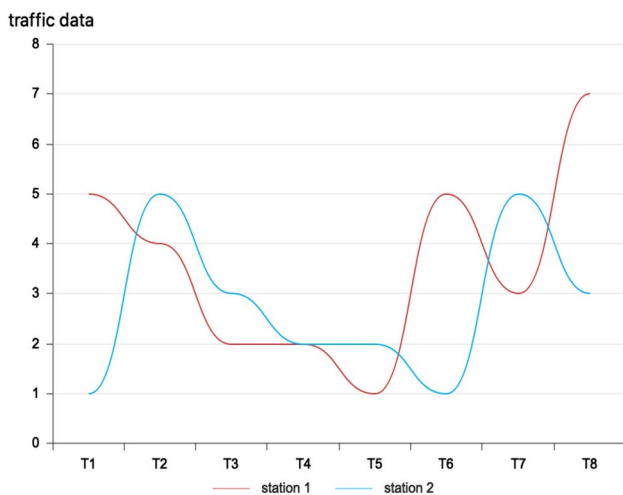
(3) Adaptive graph

In addition, considering that the first two graph structures are both looking for the correlation between traffic stations from the perspective of traffic flow data, we also use the adaptive adjacency matrix proposed by [36] in our model. It has the advantage that it does not require any prior knowledge, but is acquired through continuous learning during training, which exactly meets our needs for complex traffic networks In the real environment, the stations may be affected by geographical location, environment, humanity, and other factors, and there is a complex dependent relationship. Therefore, at the end of multi-graph framework of the spatial-temporal block, we use the adaptive graph to carry out the final convolution operation. Through it, we can excavate more hidden spatial information. The formula is as follows:

$$\tilde{A}_{adp} = SoftMax(ReLU(E_1 E_2^T)) \tag{14}$$

Where $E_1, E_2 \in \mathbb{R}^{N \times C}$ are the parameters generated by random initialization. The SoftMax function is applied to normalize the adaptive adjacency matrices.

3.3.2 Stack mechanism

In order to mine the information of each graph as much as possible, this paper proposes a novel multi-graph prediction framework based on the structure of N-BEATS, as shown in the middle of Fig. 4. Each GCN block in the spatial module corresponds to a graph convolution operation, and the specific stacking mechanism between the GCN blocks is shown in Fig. 6. For the first GCN block in the module, its input is x and a graph matrix, and its output is $\tilde{x}$ and $\tilde{y}$. x is the

Fig. 5 Comparison chart of traffic flow data fluctuation curve of two stations
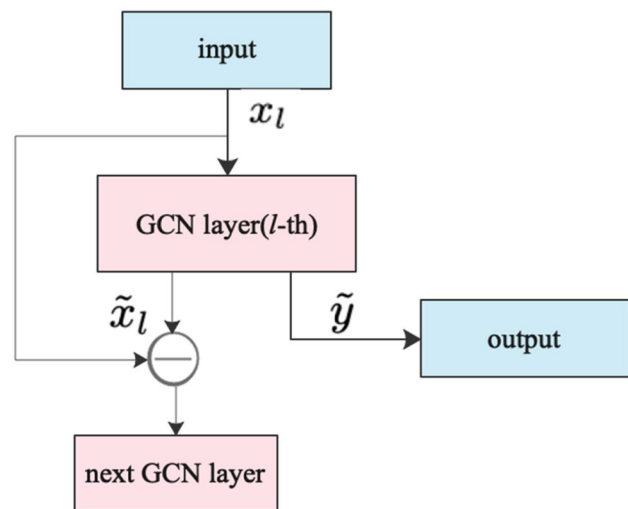

Fig. 6 Architecture of GCN block. It describes the residual connection mode of the GCN block, where the GCN block takes the input from the upper layer and with the output $\tilde{x}$ and $\tilde{y}$ as the input to the lower layer and as part of the output of the space module, respectively

historical traffic data, $\tilde{y}$ is the forecast result for the future, and $\tilde{x}$ can be understood as the "backcast" of the input x, which is the best estimate of the input x. For the rest of the blocks, they also need an input x_l, which is a residual output of the previous blocks. x_l, $\tilde{x}_l$ represent the input and output of the l-th block respectively, and the input of the next block is calculated as follows:

$$x_{l+1} = x_l - \tilde{x}_l \tag{15}$$

Because of the complexity of the traffic network, the single graph structure is often not enough to describe its relevance, so the multi-graph mechanism emerges. However, in past studies, scholars' applications of the multi-graph mechanism have usually included two approaches: 1. Multiple graphs were fused into one graph and then calculated. 2. Convolve the graph in simple layer series and parallel. Although these methods take into account the complexity of the traffic network, the simple integration can not well excavate the spatial dependence. Therefore, this paper uses an interpretable multi-graph stacking framework.

Each GCN block in the module uses different graph matrices and is stacked in order by residual connection. The convolution operation for each graph can effectively mine out the spatial information in the graph. At the same time, using the residual connection mode, it can be understood that each block removes the information that can be explained well from the initial input, and the following blocks can better mine the remaining hidden information. The output of each GCN block (except the last one) has two parts: $\tilde{x}$ and $\tilde{y}$, where $\tilde{x}$ is a review of the past and represents the mining performance of the GCN block on the input, and $\tilde{y}$ is a prediction of the future.

In addition, three kinds of graph structures are used in the proposed framework, and their stacking order in the spatial block is arranged as follows: distance graph, DTW based correlation graph, and adaptive graph.

First of all, the basic distance graph is used, which contains the connectivity, distance, and other information between stations. It is the most basic description of the traffic network, so it is put in the first place. Then we use DTW based graph to mine the correlation between stations from the historical data of each station, which can find the similarity between stations from the perspective of data. Finally, in order to capture more hidden spatial information, an adaptive graph construction method is used. The graph can be continuously learned in the process of model training to better capture the remaining hidden relationships.

$$out_1 = r = \sigma(X \oplus FC(out_1 \oplus out_2 \oplus ... \oplus out_k)) \tag{16}$$

3.3.3 Gating mechanism

Spatial block through continuous stack GCN block to realize the depth of the spatial and temporal information capture. In the process of deepening the network, the models are constantly aggregated based on different graph information (that is, different neighbor information) which may lead to unstable performance. Therefore, this paper combines the gating mechanism to construct a "learning" gate to carry out selective learning. Gating mechanism has been widely used in deep learning, especially in the variant of RNN, which has been proven to have strong transmission control capabilities.

This paper uses a gated residual block to control the final result, as shown at the right of Fig. 4. Firstly, the prediction results of all GCN blocks is splicing together, and the final output is controlled through the "learning" gate. Suppose there are k GCN blocks, and their output is out_1, out_2, ..., out_k, X is the input of the spatial block, and the gated block can be expressed as:

$$r = \sigma(X \oplus FC(out_1 \oplus out_2 \oplus ... \oplus out_k)) \tag{17}$$

$$Output = r \otimes FC(out_1 \oplus out_2 \oplus ... \oplus out_k) + (1 - r) \otimes X \tag{18}$$

3.3.4 Summary of the spatial module

In this paper, we use an interpretable multi-graph framework to explore spatial-temporal dependencies. The basic idea is to stack graph convolutional networks using different graph structures through residual connections. In this model, the information that can be well represented by the current graph convolutional layer is continuously removed from the input, which facilitates the mining of hidden information by the next graph convolutional layer and avoids the confusion of data.

3.4 Model summarization

The structure of STGMN is shown in Fig. 2, which stacks computing modules through a serial structure. After mining the spatial-temporal dependency in traffic data through two spatial-temporal blocks, the data is passed into the final output block and the output is calculated. The output block contains a CNN layer using channel attention and a fully connected layer to control the output dimension. The specific calculation process of STGMN is as follows:

1. Input traffic data and connectivity, distance and other data between stations, preprocess the data and generate the required graph structure.

2. Start to train the model. The model deals with spatial-temporal dependence through two spatial-temporal blocks. Firstly, spatial-temporal block captures multi-scale temporal dependence through temporal module, and then obtains multi-level spatial dependence through multi-graph spatial module;

3. Transfer the result of spatial-temporal blocks to output block, and calculate the final prediction result.

Finally, MSE loss is used to train the model, and the formula is as follows:

$$L(MSE) = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y_n})^2 \tag{19}$$

4 Experimental results and analysis

In this section, the proposed model experiments on real datasets, and the experimental results are analyzed and compared.

4.1 Datasets

The datasets used in this article are from a real highway dataset in California, collected by the Caltrans Performance Measurement System (PEMS) [37]. PEMS has more than 39,000 sensors deployed in major metropolitan areas in California. The System samples at 30 seconds, and the data is aggregated to 5 minutes. Data features include traffic flow, average speed, and average lane occupancy.

PeMSD4: the data of San Francisco Bay, including a total of 3848 detectors on 29 roads, data range from January to February 2018. There are a total of 59 days of data. Data of 307 detectors are selected in this paper, the first 50 days as a training set, the last 9 days as a test set;

PeMSD8: the data of San Bernardino, containing 1979 detectors on 8 roads, data time range from July to August 2016. There are a total of 62 days of data. This paper selects the data of 170 detectors, the first 50 days are used as the training set and the last 12 days as the test set.

4.2 Experimental settings

4.2.1 Parameter settings

In the experiment of this paper, the data of the past hour are used to predict the data of the future one hour. Since the time interval of the data set is once every 5 minutes, that means, 12 historical data are used to predict the next 12 data. In the model proposed in this paper, all the convolution operations are set with 64 filters (including graph convolution and 1D

convolutional network). In the spatial module, the sequence of graphs used is distance graph, DTW diagram graph and adaptive graph. When using distance graph, Chebychev polynomials are attempted to approximate the convolution kernels (K=3). In the training stage, we use Adam to optimize all models 50 epochs, batch size is 50, and the learning rate is 0.0005.

4.2.2 Compared methods

In order to prove the validity of the model proposed in this paper, we set experiments to compare it with the following methods.

(1) HA: Historical Average method.
(2) SVM [38]: Support Vector Machine, a traditional prediction method.
(3) GRU [39]: Gated Recurrent Unit Network.
(4) LSTM [40]: Long Short-Term Memory Network.
(5) STGCN [23]: Spatio-Temporal Graph Convolutional Network model.
(6) ASTGCN [24]: Attention Based Spatial-Temporal Graph Convolutional Network model.
(7) Graph WaveNet [36]: A CNN-based method for Deep Spatial-Temporal Graph Modeling.
(8) AGCRN [41]: An Adaptive Graph Convolutional Recurrent Network to capture fine-grained spatial and temporal correlations in series.

4.2.3 Evaluation functions

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), two commonly used evaluation functions, are used in this paper to evaluate the experimental results. The formulas are as follows:

$$MAE = \frac{1}{N} \sum_{n=1}^{N} |y_n - \hat{y_n}| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y_n})^2} \tag{21}$$

4.3 Experimental results

4.3.1 Comparison of models

Table 1 respectively shows the prediction effects of the proposed model (STGMN) and other compared models on PeMSD4 and PeMSD8. The experimental results prove that

Table 1 Experimental results on PeMSD4 and PeMSD8

Model	PeMSD4		PeMSD8	
	RMSE	MAE	RMSE	MAE
HA	55.17	37.46	45.79	31.09
SVM	45.71	32.29	38.22	27.23
GRU	43.61	28.70	32.02	21.05
LSTM	43.59	28.53	31.80	20.66
STGCN	35.99	23.84	27.71	18.35
ASTGCN	35.14	22.86	27.38	17.77
AGCRN	33.60	20.89	26.23	16.30
WaveNet	36.52	23.37	27.11	18.06
STGMN*	**32.85**	**20.42**	**24.58**	**15.87**

Bold entries indicate the best results of each evaluation index

the proposed model can effectively extract spatial-temporal features and achieve good results in traffic flow prediction. From Table 1, we can clearly observe that:

- Compared with traditional prediction methods such as HA and SVM, the model based on deep learning achieves better results on both datasets.
- Compared with LSTM and GRU which only consider the time relationship, the graph-based models including STGCN and ASTGCN have been greatly improved, which proves the usefulness of the graph-based method in the field of traffic flow prediction.
- STGMN achieves the best result among all models under all metrics, and the improvement effect is obvious.

To sum up, we may arrive at the following conclusions:

- In the face of increasingly complex traffic conditions, we should not only take into account the time dependence, but also make reasonable use of the spatial information of the traffic network when forecasting traffic flow. Compared with traditional forecasting methods, models based on spatial and temporal information mining, such

as STGCN, ASTGCN, WaveNet and STGMN, have achieved better results.

- As a spatial-temporal model, STGMN proposed in this paper adopts the method of constructing spatial-temporal information mining structure compared with similar STGCN, ASTGCN, WaveNet, etc. However, compared with other methods, the temporal module of STGMN adopts 1D-CNN based on the idea of Inception, using different filters to mine time characteristics, and using attention mechanism to combine the characteristics of data at different scales. The spatial module uses a novel, interpretability of the multi-graph architecture, making full use of a variety of graph structure of the space information. These improvements enable STGMN to achieve higher computational efficiency and precision and show good performance on data sets.

In order to further analyze the performance of STGMN, we show the prediction errors of all models in PeMSD4 and PeMSD8 at each step in Figs. 7 and 8 respectively. It can be clearly seen that STGMN has the best effect in both datasets, and the prediction result of each step is the best among all models. Compared with other models, HA, GRU, and LSTM have the worst prediction effect, and the prediction performance decreases significantly with the increase of prediction length, which proves the validity of the spatial-temporal models. The prediction performance of STGCN, ASTGCN, and WAVENet is similar, and they can all obtain excellent results in the first step. However, the stability of the model is not enough, and the performance degradation rate is obviously higher than that of STGMN and AGCNRN. STGMN has achieved significant improvement in all aspects. Compared with AGCRN, which has the closest prediction results with STGMN, although AGCRN has a stable prediction effect, it performs a poor prediction effect at the beginning, especially in PeMSD4 data, the prediction result of the first two steps of AGCRN is even lower than that of LSTM

Fig. 7 Comparison of each step error of all models on dataset PeMSD4
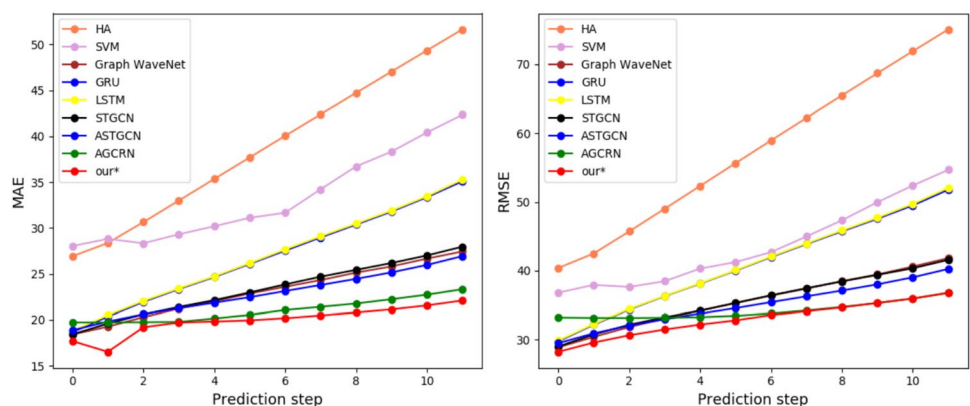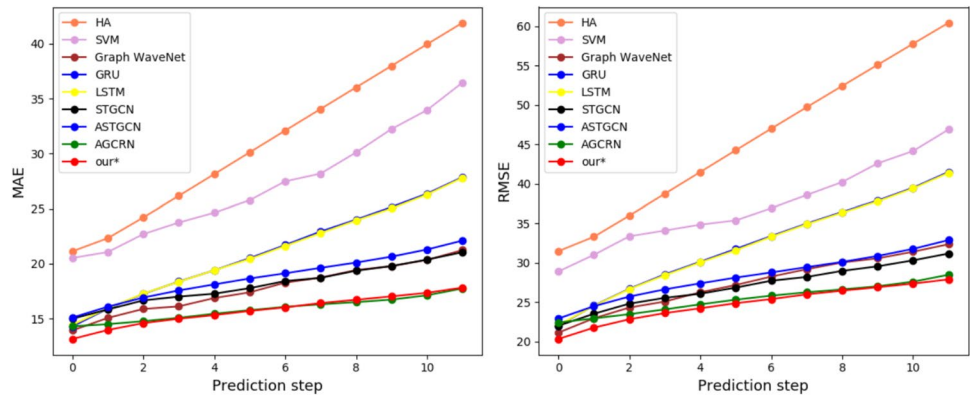
Fig. 8 Comparison of each step error of all models on dataset PeMSD8

and RNN, while the prediction curve of STGMN is obviously more stable and its performance degradation is slower.

4.3.2 Experiments of multi-graph mechanism

4.4 Experiments of multi-graph mechanism

In STGMN, a multi-graph mechanism is proposed. In order to verify its effectiveness and the effect of three different graphs selected by it, four models are respectively tested by taking the PeMSD8 dataset as an example.

- the model which uses graph convolutional network with distance graph only (STGMN-1).

- the model which uses the multi-graph mechanism with Chebychev (K=3) distance graph (STGMN-2).
- the model which uses distance graph and DTW correlation graph simultaneously (STGMN-3).
- and the model using three graphs (distance graph, DTW correlation graph, and adaptive graph) simultaneously (STGMN).

In our experimental setting, in order to compare the improved effect of three different map information on spatial information mining, we first set STGMN-1, which only has the most basic information of site connection and distance, and is widely used in the map convolutional network by researchers. Meanwhile, we set STGMN-2, which introduces third-order Chebychev on the basis of STGMN-1. In addition, STGMN-3 also introduces DTW diagram information,

Fig. 9 Experimental results of different multi-graph mechanisms
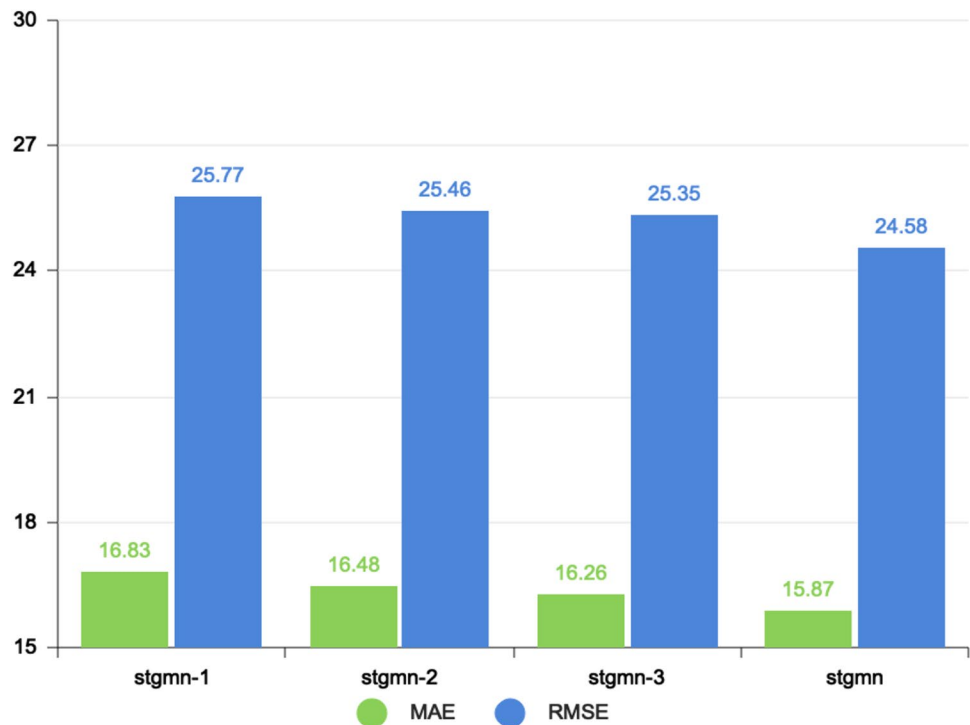
and STGMN simultaneously uses three graph structures, which is a complete model setting.

The experimental results are shown in Fig. 9. It can be clearly observed that compared with STGMN-1, the performance of STGMN has been significantly improved, which proves the effectiveness of the multi-graph mechanism proposed in this paper. In addition, it can also be seen that the three graph structures used in this paper have improved the prediction performance of the model to some extent. In addition, it can be seen that from STGMN-1 to STGMN, as graph information is added to the model, the prediction effect of the model is continuously improved, which proves that the three graph structures used in this paper can realize further mining of spatial information on the basis of the original model, and also proves the effectiveness of the multi-graph mechanism proposed in this paper. With more and more information available, the interference between graphs can be avoided by mining the information of graph structure step by step. In conclusion, the proposed multi-graph mechanism and the chosen graph structure are very effective and can help the model to better mine different spatial information and further improve the prediction accuracy of the model.

4.4.1 Experiments of different forecast length

Finally, in order to verify the effectiveness of the model in scenarios with different prediction durations, four prediction lengths were set for experiments in this paper. Table 2 shows the results of the model in this paper in predicting the next 15 minutes (3 steps), the next 30 minutes (6 steps), the next 45 minutes (9 steps), and the next 1h (12 steps) on two datasets respectively. We can observe the short-term forecast result (15 minutes) is obviously better than that of 1h prediction results, proving that the proposed model on the short-term forecast has obvious advantages. In addition, the prediction result of 45 minutes is close to 1h, which proves that the prediction performance

of this model is stable, and it can still maintain a relatively good prediction result in the face of long-term traffic prediction problems.

5 Conclusion

In the face of increasingly complex traffic conditions, this paper proposes a gated multi-graph attention spatial-temporal model (STGMN) for traffic flow prediction. In STGMN, for temporal dependence, we propose a temporal learning module based on attention mechanism and "Inception" structure. By enlarging the width of the network layer, the receptive field of CNN at different scales is given, which effectively improves the temporal capture ability of the model. For spatial dependence, we propose a spatial learning module based on gated multi-graph convolution, which can effectively improve the performance of the graph neural network with the multi-graph mechanism and the stacked framework. Finally, the validity, superiority and stability of the proposed model are verified by experiments on traffic flow datasets PeMSD4 and PeMSD8.

In future work, we will consider adding more external data information in the experiment to further improve the effectiveness of the model. At the same time, we will continue to explore the model structure, further improve the prediction effect, and try to introduce the model structure in this paper to other practical applications.

References

1. Lana I, Del Ser J, Velez M, Vlahogianni EI (2018) Road traffic forecasting: Recent advances and new challenges. IEEE Intelligent Transportation Systems Magazine 10(2):93–109
2. Ye J, Zhao J, Ye K, Xu C (2020) How to build a graph-based deep learning architecture in traffic domain: A survey. IEEE Trans Intell Transp Syst :1–21
3. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) Graph neural networks: A review of methods and applications. AI Open 1:57–81
4. Lee K, Eo M, Jung E, Yoon Y, Rhee W (2021) Short-term traffic prediction with deep neural networks: A survey. IEEE Access 9:54739–54756
5. Luo X, Niu L, Zhang S (2018) An algorithm for traffic flow prediction based on improved sarima and ga. KSCE Journal of Civil Engineering 22(10):4107–4115
6. Kumar SV, Vanajakshi L (2015) Short-term traffic flow prediction using seasonal arima model with limited input data. European Transport Research Review 7(3):1–9

Table 2 Experimental results of different forecast length on PeMSD4 and PeMSD8

datasets		15min	30min	45min	1h
PeMSD4	MAE	18.86	19.55	20.20	20.42
	RMSE	29.73	31.15	32.36	32.84
PeMSD8	MAE	14.15	14.64	15.37	15.87
	RMSE	21.66	22.87	24.13	24.58

7. Cong Y, Wang J, Li X (2016) Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm. Procedia Engineering 137:59–68

8. Makridakis S, Spiliotis E, Assimakopoulos V (2018) Statistical and machine learning forecasting methods: Concerns and ways forward. PloS ONE 13(3):e0194889

9. Li H, Wang Y, Xu X, Qin L, Zhang H (2019) Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network. Applied Soft Computing 83:105620

10. Lv Z, Xu J, Zheng K, Yin H, Zhao P, Zhou X (2018) Lc-rnn: A deep learning model for traffic speed prediction. Int Joint Conf Artif Intell :3470–3476

11. Yin X, Wu G, Wei J, Shen Y, Qi H, Yin B (2021) Multi-stage attention spatial-temporal graph networks for traffic prediction. Neurocomputing 428:42–53

12. Guo S, Lin Y, Li S, Chen Z, Wan H (2019) Deep spatial-temporal 3d convolutional neural networks for traffic data forecasting. IEEE Transactions on Intelligent Transportation Systems 20(10):3913–3926

13. de Medrano R, Aznarte JL (2020) A spatio-temporal attention-based spot-forecasting framework for urban traffic prediction. Applied Soft Computing 96:106615

14. Peng H, Du B, Liu M, Liu M, Ji S, Wang S, Zhang X, He L (2021) Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. Information Sciences 578:401–416

15. Xiao Y, Yin H, Zhang Y, Qi H, Zhang Y, Liu Z (2021) A dual-stage attention-based conv-lstm network for spatio-temporal correlation and multivariate time series prediction. International Journal of Intelligent Systems 36(5):2036–2057

16. Yu H, Wu Z, Wang S, Wang Y, Ma X (2017) Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors 17(7):1501

17. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY (2020) A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems 32(1):4–24

18. Zhang Z, Cui P, Zhu W (2020) Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering 34(1):249–270

19. Liang Y, Ke S, Zhang J, Yi X, Zheng Y (2018) Geoman: Multi-level attention networks for geo-sensory time series prediction. Int Joint Conf Artif Intell :3428–3434

20. Zheng C, Fan X, Wang C, Qi J (2020) Gman: A graph multi-attention network for traffic prediction. Proceedings of the AAAI Conference on Artificial Intelligence 34(01):1234–1241

21. Yao H, Tang X, Wei H, Zheng G, Li Z (2019) Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. Proc AAAI Conf Artif Intell 33(01):5668–5675

22. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2019) T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems 21(9):3848–3858

23. Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. Int Joint Conf Artif Intell :3634–3640

24. Guo S, Lin Y, Feng N, Song C, Wan H (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 33(01):922–929

25. Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. Int Conf Learn Representations :1–16

26. Fang S, Zhang Q, Meng G, Xiang S, Pan C (2019) Gstnet: Global spatial-temporal network for traffic flow prediction. Int Joint Conf Artif Intell :2286–2293

27. Song C, Lin Y, Guo S, Wan H (2020) Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 34(01):914–921

28. Tedjopurnomo DA, Bao Z, Zheng B, Choudhury F, Qin A (2020) A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. IEEE Transactions on Knowledge & Data Engineering 01:1

29. Zhang W, Yu Y, Qi Y, Shu F, Wang Y (2019) Short-term traffic flow prediction based on spatio-temporal analysis and cnn deep learning. Transportmetrica A: Transport Science 15(2):1688–1711

30. Bogaerts T, Masegosa AD, Angarita-Zapata JS, Onieva E, Hellinckx P (2020) A graph cnn-lstm neural network for short and long-term traffic forecasting based on trajectory data. Transportation Research Part C: Emerging Technologies 112:62–77

31. Zhang Q, Chang J, Meng G, Xiang S, Pan C (2020) Spatio-temporal graph structure learning for traffic forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 34(01):1177–1185

32. Kipf TN, Welling M Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations

33. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1–9

34. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020) Eca-net: Efficient channel attention for deep convolutional neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp 11531–11539

35. Oreshkin BN, Carpov D, Chapados N, Bengio Y N-beats: Neural basis expansion analysis for interpretable time series forecasting. International Conference on Learning Representations

36. Wu Z, Pan S, Long G, Jiang J, Zhang C (2019) Graph wavenet for deep spatial-temporal graph modeling, International Joint Conference on Artificial Intelligence. pp 1907–1913

37. Chen C, Petty K, Skabardonis A, Varaiya P, Jia Z (2001) Freeway performance measurement system: mining loop detector data. Transportation Research Record 1748(1):96–102

38. Cortes C, Vapnik V (1995) Support-vector networks. Machine learning 20(3):273–297

39. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. NIPS 2014 workshop on deep learning

40. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780

41. Bai L, Yao L, Li C, Wang X, Wang C (2020) Adaptive graph convolutional recurrent network for traffic forecasting. 34th conference on neural information processing systems, vol 33. pp 17804–17815

Qingjian Ni received his Ph.D. degree from Southeast University in 2009. Currently, he is an Associate Professor at the School of Computer Science and Engineering, Southeast University. His research interests mainly focus on artificial intelligence and its applications.

Meng Zhang received the B.S. degree from Soochow University, in 2019. She is currently pursuing M.S. degree in the School of Computer Science and Engineering, Southeast University. Her research interests include spatio-temporal analysis, graph representation learning and and their applications.