Probing Traffic Trend Forecasting via Spatial-Temporal Aware Learning-Graph Attention

Xinyuan Huang hxy2211931@gmail.com
Heilongjiang University

Qianqian Ren renqianqian@hlju.edu.cn
Heilongjiang University

Editors: Berrin Yanıkoğlu and Wray Buntine

Abstract

Traffic forecasting plays an extremely important role in many applications such as intelligent transportation and smart cities. However, due to the hidden and complex dynamic spatio-temporal correlations and heterogeneity, achieving high-precision traffic prediction is a challenging task. This paper proposes a new spatio-temporal aware learning graph neural network (STALGNN) for traffic prediction. First, a temporal-aware graph generation module is designed to exploit the spatial-temporal features that the spatial graph may not be able to present. Then, a spatio-temporal joint module is designed to more effectively capture local spatio-temporal correlations. Next, a multi-scale gated convolutions module is proposed to capture gloable dynamic spatio-temporal correlations. Furthermore, STALGNN further learns explicit spatio-temporal correlations through integrated attention mechanisms and stacked graph convolutional networks to handle long-term prediction. Extensive experiments on several real traffic datasets show that the proposed method can achieve the superior performance compared with other baselines.

Keywords: Traffic forecasting, Spatio-temporal correlation, Graph convolution network, Attention mechanism

1. Introduction

Traffic forecasting plays an important part in intelligent transportation management and planning, i.e. route planning, intelligent traffic light control, etc. With the well development of data acquisition technology, the amount of traffic data is increasing rapidly. Meanwhile, The rise in traffic flow within the road network results in a heavier burden for traffic management. Therefore, accurately predicting traffic flow based on the collected historical observations is of great significance.

In recent years, deep learning methods have become popular for high-dimensional spatio-temporal traffic flow prediction. Classic models are to use convolutional neural networks (CNN) and recurrent neural networks (RNN) to handle spatio-temporal correlations in traffic networks. Although convolutional neural network methods are suitable for capturing local spatial correlations in regular spatial grids, they have some deficiencies in predicting traffic conditions in non-grid road networks with long-range spatial correlations.

STGCN (Yu et al., 2017) uses a road network to describe the spatial correlations between sensors and uses graph convolution methods to extract spatio-temporal features.

STSGCN Song et al. (2020) establishes the temporal correlations between sensors in continuous traffic flow and then performs flow prediction using graph embedding techniques. Although existing methods have achieved better prediction performance, there exists problems that may not be well addressed.

For short and long-term prediction, traffic patterns exhibits dynamic and complex correlations. On one hand, we observe similar patterns of congestion on specific roads during peak hours in the morning and evening, as well as consistent congestion patterns during different times on workdays versus holidays. On the other hand, the traffic flow is fluctuating due to various uncertain factors like waiting times at traffic lights, driver behaviors, and vehicles speed. Moreover, existing methods such as RNN/LSTM-based models Zhang et al. (2018), take a long time to process the data and may encounter gradient explosion or disappearance when processing long-term prediction. CNN-based methods need to stack convolutional layers to capture long-term correlations in long sequences. In the STGCN Yu et al. (2017) and GraphWaveNet Wu et al. (2019) models, if long-term dependencies are extracted, the expansion rate of convolution needs to be increased, which is at the price of some short-term information losing. Existing methods can not obtain satisfactory solutions for both short and long-term time prediction.

To address the above problems, we propose a new spatio-temporal aware learning graph attention neural network named STALGNN for traffic forecasting. The model can simultaneously extract local and global correlations in the traffic road network. Instead of simply using traffic data to train the model, we supplementary traffic trends to jointly train the model. Traffic flow and trends reflect different features of traffic data. The trend is the time transform of traffic flow related to the previous time step. The traffic flow distribution is irregular, but the trend distribution is more concentrated. In short, although there may be some differences in some events, things generally follow some rules. Therefore, this paper believes that trends can play an important role in traffic flow. The main contributions of this paper can be summarized as follows:

- We design a spatio-temporal attention convolution module to better model the complex spatio-temporal correlations of traffic flow data. In particular, a graph convolution is used to process the input data to better extract spatial correlations. An improved attention mechanism adaptively adjusts the weights of the graph convolutions.
- We construct the graph structure using random walk method that retains implicit spatio-temporal correlations of the traffic network.
- Instead of simply utilizing traffic flow data to train the model, we add traffic trends to jointly train the model. Trends are considered to play a certain importance in traffic flow prediction and are regarded as auxiliary information.
- We design an improved gated convolution module. Through multi-scale gated convolutions, the model's capture of long-term time dependencies in the traffic road network is further enhanced.
- Extensive experiments on real-world traffic datasets show that the model we propose efficiently improves performance compared to other baselines.

2. Related work

2.1. Graph neural network

Graph neural networks are a class of methods that use graph theory frameworks for neural network modeling. It represents the relationships or dependencies between data in the form of graphs, and learns graph data using operations such as convolution, pooling, and attention mechanisms. Graph neural networks provide a unified framework for processing graph-structured data and modeling processes evolving on graphs. It draws on the ideas of traditional CNNs, defines convolutions, pooling, and attention mechanisms on non-Euclidean graph data, can learn the topological structure and node features of the graph, obtain network embeddings, and achieve downstream tasks such as node classification, link prediction, and graph classification. In addition, graph neural networks also have the advantage of end-to-end learning, which automates the modeling process.

Spatio-temporal graph neural networks (STGNN) aim to learn hidden spatio-temporal patterns from spatio-temporal graph data, which is becoming increasingly important for applications such as traffic speed prediction, driver behavior expectation, and human behavior recognition. The key idea of STGNN is to consider both spatial dependencies and temporal dependencies at the same time. Currently, many spatio-temporal modeling methods use graph convolutional networks to capture spatial dependencies and use recurrent neural networks or convolutional neural networks to learn temporal dependencies. However, these methods usually model spatial dependencies and temporal dependencies separately and then simply fuse them, which cannot well express the complex nonlinear interactions between the two. STGNN proposes to directly define graph convolution operations on the spatio-temporal graph, learning node features, spatial dependencies, and temporal dependencies at the same time to better model the complexity of spatio-temporal relationships. Its spatio-temporal graph convolution aggregates information from neighboring nodes at multiple time steps to learn the local spatio-temporal patterns of spatio-temporal graph data, and predicts the entire spatio-temporal graph based on this. STGNN uses a unified mechanism to learn spatial dependencies and temporal dependencies at the same time, which can better establish a complex mapping between them and achieve end-to-end spatio-temporal modeling and prediction. This has stronger expressive power and advantages than first learning spatial structure and time series separately, then simply fusing them.

2.2. Traffic data prediction

Traffic prediction is a fundamental and critical technology in intelligent transportation systems Zhang et al. (2011). It has received extensive attention and research in recent decades. Early research mainly used linear time series analysis methods, such as the general time series analysis method VAR Lu et al. (2016) model. As an extension of the autoregressive model, the VAR model can consider the linear correlation between multiple time series. However, these linear models perform poorly in modeling and predicting traffic data because traffic data changes are complex, and correlations are often nonlinear. To weaken the limitations of the linear assumptions, machine learning-based methods such as support vector regression Wu et al. (2004) and k-nearest neighbor methods Van Lint and Van Hinsbergen (2012) were later proposed. When provided with manually extracted high-level features, these methods can learn more complex dependencies and provide better prediction results than linear models. However, manual feature extraction is a labor-intensive and time-consuming process.

STGCN and GraphWaveNet are works that apply graph convolutional networks to spatio-temporal data modeling. They respectively use graph convolution processing network structure information in the spatial dimension and one-dimensional convolution network learning time series in the time dimension, and then simply add the outputs of the two to obtain the state of each node at the current time step. Such methods can learn the spatial dependence and temporal dependence of spatio-temporal data respectively, but do not consider the complex mapping between them. Therefore, their ability to model the overall spatio-temporal relationship of spatio-temporal data is relatively limited.

3. Problem Definition

In this section, we provide necessary definitions in the paper, and formalize the problem of traffic prediction.

Definition 1.(Transportation road network.) The transportation road network is represented as an undirected graph $\mathcal{P} = (V, E_r)$, where V and E_r represent the sets of sensors and connected road sections between sensors, respectively. The transportation road network $\mathcal{P}$ represents the spatial relationship between sensors.

Definition 2.(Graph signal matrix.) Given the transportation road network $\mathcal{P} = (V, E_r)$ and a continuous sequence of time intervals, the observed graph signal on $\mathcal{P}$ is denoted as $\boldsymbol{x}_{\mathcal{P}} \in \mathbb{R}^{|V|}$, where the i-th element of $\boldsymbol{x}_{\mathcal{P}}^{(t)}$ represents the traffic flow observed by the i-th sensor at the t-th time step.

Definition 3.(Traffic prediction problem.) At time step t, given the transportation road network $\mathcal{P}$ and graph signal $\{\boldsymbol{x}_{\mathcal{P}}^{(t-T+1)}, \cdots, \boldsymbol{x}_{\mathcal{P}}^{(t)}\}$ for the past T time steps, the traffic prediction problem can be described as finding a mapping function f to predict the traffic flow of the future time steps T'.

$$\left(X_{\mathcal{P}}^{(t-T+1)}, \ldots, X_{\mathcal{P}}^{(t)}\right) \xrightarrow{f} \left(X_{\mathcal{P}}^{(t+1)}, \ldots, X_{\mathcal{P}}^{(t+T')}\right). \tag{1}$$

4. Methodologies

In this section, we focus on introducing the technical details of the proposed framework. Figure. 1 shows the overall architecture. In the model, we design a spatial and temporal module to jointly model the temporal characteristics of traffic data at different time steps and the spatial dependencies between multiple spatial regions. To this end, we combine a pair of temporal convolutional networks and graph convolutional networks as the basic framework. To model the temporal feature of traffic data, we use two one-dimensional causal convolutions with two gate units. To model the spatial dependencies between multiple spatial regions, we integrate attention mechanisms and stacked graph convolutional networks.

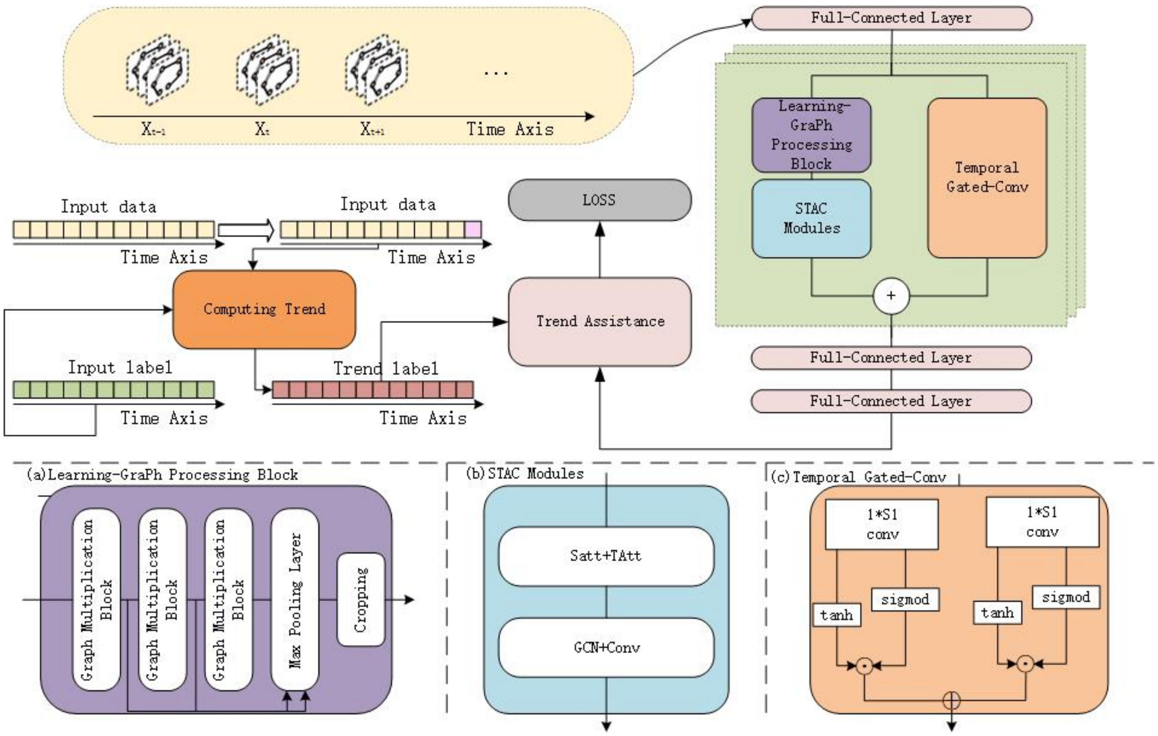

Figure 1: The overall structure of STAMGNN, consisting of four modules: multi-graph processing modules, STAC modules and temporal-gated convolutions.

4.1. Learning-Graph Processing

4.1.1. Temporal Graph

The temporal graph can be represented as $\mathcal{G}_T = (V, E_t)$, where V represents the set of nodes in the transportation road network, and E_t indicates connection of nodes at time step t. Let us take the period of a week as an example, which is divided into a set of consecutive time intervals with equal length. Assuming there are T such time intervals within a week, then for each sensor, a T-dimensional feature vector is constructed. The i-th element in the feature vector represents the average traffic flow recorded by the node in the i-th time interval of this week. In the time feature vector, the neighbors of adjacent nodes are determined according to the Euclidean distance metric.

4.1.2. Spatial-temporal Graph Embedding

Given the transportation road network $\mathcal{P}$ and the temporal graph $\mathcal{G}_T$, we need to determine an embedding function to represent correlations between nodes. The embedding function is expressed as: $\boldsymbol{h} : V \to \mathbb{P}^n$, mapping each sensor in the road network to an n-dimensional feature vector. In this sense, we obtain an embedding function $\boldsymbol{h}(\cdot)$ and a spatio-temporal learning graph $\mathcal{G}_{TS} = (V, E)$, the spatio-temporal dependence between each sensor is well mapped in E, and well preserved in the embedding function.

Inspired by the efficiency of random walks Grover and Leskovec (2016) and negative sampling Han et al. (2021), we optimize the representation procedure. Starting from the first node $\boldsymbol{v}_0 \in V$ in $\mathcal{G}_{TS}$, we generate a random walk path $p = \langle v_0, v_1, \cdots, v_L \rangle$ of length L. Through a predetermined threshold $\Delta(\Delta \ll L)$, we can obtain a set of neighbors from path p to v_i:

$$\mathcal{N}_p(v_i) = \{v_k \in p \| k - i \mid \leq \Delta, k \neq i\}, \tag{2}$$

where $i = \{0, 1, 2, \ldots, L\}$, $\mathcal{N}_p(v_i)$ is a sample of neighbors of node v_i.Traditional graph embedding methods generate random walks based on topological structure to search neighborhood sets, without considering temporal correlations. In this paper, we use a new sampling strategy, which considers the temporal correlation in random walks.

Given the transportation road network $\mathcal{P}$, the temporal graph $\mathcal{G}_T$ and a random walk path $p_i = \langle v_0, v_1, \cdots, v_L \rangle$ staying at node v_i. According to the probability, the sampling strategy determines the next node v_{i+1} to visit:

$$\Pr(v_{i+1} \mid p_i) \propto \begin{cases} \pi(\tau_i, v_{i+1}), & \text{if } (v_i, v_{i+1}) \in \mathcal{P} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\pi(\tau_i, v_{i+1})$ represents the weight assigned to edge (v_i, v_{i+1}).

$$\pi(p_i, v_{i+1}) = \begin{cases} a, & \text{if } d_{\mathcal{P}}(v_{i+1}) = 0 \text{ and } (v_0, v_{i+1}) \in \mathcal{G}_{\mathcal{T}} \\ 1, & \text{if } d_{\mathcal{P}}(v_{i+1}) = 1 \text{ and } (v_0, v_{i+1}) \in \mathcal{G}_{\mathcal{T}} \\ b, & \text{if } d_{\mathcal{P}}(v_{i+1}) = 2 \text{ and } (v_0, v_{i+1}) \in \mathcal{G}_{\mathcal{T}} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $d_{\mathcal{P}}(v_{i+1})$ represents the shortest path distance between nodes v_{i+1} and v_{i-1} in the road network. According to the above formula, we ensure that the traffic flow patterns of all nodes are similar to the first node in the random walk. Moreover, this sampling strategy can preserve the topological structure of the road network.

4.1.3. Spatial-temporal Correlation Modeling

According to the previous section, we have obtained an embedding function that represent spatio-temporal correlations. Next, we model the spatio-temporal correlations of the graph. Specifically, the correlation between node v_i and v_j is expressed as:

$$\boldsymbol{M}(i, j) = (\boldsymbol{h}(v_i), \boldsymbol{h}(v_j)) \tag{5}$$

According to the spatio-temporal correlation definition, we calculate a k-nearest neighbor set for each node using Euclidean distance.Finally, normalize it to obtain the learned spatio-temporal association graph $\mathcal{G}_{\mathcal{TS}}$.Next, the hidden spatio-temporal correlations are extracted by the following graph matrix multiplication, modules. In the graph multiplication module, the gating mechanism with LSTM is used. The gated linear unit uses its nonlinear activation for generalization in graph multiplication. The graph multiplication module is represented as:

$$h^{l+1} = \left(\mathcal{G}_{\mathcal{TS}} h^l W_1 + b_1\right) \odot \sigma\left(\mathcal{G}_{\mathcal{TS}} h^l W_2 + b_2\right) \tag{6}$$

In the above formula, h^l represents the i-th hidden state, $W_1, W_2, b_1 and b_2$ are the model parameters in GLU, $\odot$ is the Hadamard product, σ is the sigmoid activation function.

4.2. STAC Module

The purpose of the spatio-temporal attention convolution(STAC) module to capture the dynamic spatial and temporal dependencies, which consists of temporal attention, spatial attention and graph convolution modules.

4.2.1. Temporal attention mechanism

On the time axis, traffic conditions are correlated between different time windows over time, and the degree of correlations also varies in different situations. The attention mechanism can automatically learn the relationship and importance between different time slices, dynamically assign attention weights to different time slices, and achieve adaptive information selection and filtering. This allows the model to select time slices that are more relevant and important to obtain more accurate information. Temporal attention is expressed as follows:

$$\mathbf{E} = \mathbf{V}_e \cdot \sigma\left(\left(\left(\mathcal{X}_{\mathcal{P}}\right)^T \mathbf{U}_1\right) \mathbf{U}_2 \left(\mathbf{U}_3 \mathcal{X}_{\mathcal{P}}\right) + \mathbf{b}_e\right), \qquad \mathbf{E}'_{i,j} = \frac{\exp\left(\mathbf{E}_{i,j}\right)}{\sum_{j=1}^{T_r-1} \exp\left(\mathbf{E}_{i,j}\right)} \tag{7}$$

In the above formula, $\mathbf{V}_e, \mathbf{b}_e \in \mathbb{R}^{T \times T}, \mathbf{U}_1 \in \mathbb{R}^N, \mathbf{U}_2 \in \mathbb{R}^{F \times N}, \mathbf{U}_2 \in \mathbb{R}^F$ are learnable parameters, where T represents the time step and F represents the number of channels of the data. The temporal association matrix $E_t ime$ is determined by the changing input. The value of element $E_{(i,j)}$ in this matrix represents the strength of dependence between time i and time j. Finally, the temporal association matrix is normalized by the softmax function. We directly use the normalized temporal attention matrix as input and dynamically adjust the input by merging relevant information.

4.2.2. Spatial attention mechanism

In space, the influence of different traffic factors leads to a high degree of dynamics in the spatial dependence of traffic data. Therefore, this paper uses an attention mechanism Feng et al. (2017) to adaptively calculate the dynamic spatial dependence between nodes.The spatial attention is represented as follows:

$$\mathbf{S} = \mathbf{V}_s \cdot \sigma\left(\left(\mathcal{X}_{\mathcal{P}} \mathbf{W}_1\right) \mathbf{W}_2 \left(\mathbf{W}_3 \mathcal{X}_{\mathcal{P}}\right)^T + \mathbf{b}_s\right), \qquad \mathbf{S}'_{i,j} = \frac{\exp\left(\mathbf{S}_{\mathrm{i,j}}\right)}{\sum_{j=1}^{N} \exp\left(\mathbf{S}_{\mathrm{i,j}}\right)} \tag{8}$$

In the above formula, $\mathbf{V}_s, \mathbf{b}_s \in \mathbb{R}^{N \times N}$, $\mathbf{W}_1 \in \mathbb{R}^T$, $\mathbf{W}_2 \in \mathbb{R}^{F \times T}$, $\mathbf{W}_2 \in \mathbb{R}^F$are learnable parameters. According to the current input, dynamically calculate the spatial attention matrix S. The value of element $S_{(i,j)}$ represents the strength of association between node i and node j. Then use the softmax function to ensure that the sum of the attention weights of the nodes is 1. When performing graph convolution, dynamically adjust the weights between nodes together with the adjacency matrix and the spatial attention matrix.

4.2.3. GCN for Spatial-temporal Features

After the spatio-temporal attention module, more weight is given to valuable information in the traffic data, and the information adjusted by the attention module is used as input to the spatio-temporal graph convolution module. The spatio-temporal convolution module designed in this paper includes a spatial graph convolution and a temporal graph convolution. The spatial graph convolution extracts spatial dependence from the node neighborhood. The temporal graph convolution extracts temporal dependence from adjacent time nodes.

Spatial graph convolution: In the research problem of traffic flow prediction, the traffic road network is essentially a graph structure, and the features of each node can be regarded as signals on the graph Shuman et al. (2013). Because spectral graph theory generalizes convolutional operations from grid-based data to graph-structured data, in order to make full use of the topological characteristics of the traffic road network, this paper uses graph convolution based on spectral graph theory to directly process the data on each time slice, utilizing the spatial correlation between nodes on the traffic road network. The spectral graph method transforms the graph into algebraic form and analyzes the topological structure of the graph.

In spectral graph theory, a graph can be represented by its corresponding Laplacian matrix. By analyzing the Laplacian matrix and its eigenvalues, the properties of the graph structure can be obtained. The Laplacian matrix of a graph is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, and its normalized form is $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{N\times N}$, where $\mathbf{A}$ is the adjacency matrix, $\mathbf{I}_N$ is the identity matrix, and the degree matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix composed of node degrees, $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The eigenvalue decomposition of the Laplacian matrix is $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$, where $\boldsymbol{\Lambda} = \operatorname{diag}\left([\lambda_0, \ldots, \lambda_{N-1}]\right) \in \mathbb{R}^{N\times N}$ is a diagonal matrix and $\mathbf{U}$ is the Fourier basis. Taking the traffic flow at time t as an example, the signal on the entire graph is $x = \mathbf{x}_{\mathcal{P}}^t \in \mathbb{R}^N$, and the graph Fourier transform of this signal is defined as $\hat{x} = \mathbf{U}^T x$. According to the properties of the Laplacian matrix, $\mathbf{U}$ is an orthogonal matrix, and the corresponding Fourier inverse transform is $x = \mathbf{U}\hat{x}$. Graph convolution is a kind of convolution operation that replaces the classical convolution operator with a linear operator diagonalized in the Fourier domain. On this basis, the signal x on the graph G is filtered using the kernel function g_θ:

$$g_\theta *_G x = g_\theta(\mathbf{L})x = g_\theta\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\right)x = \mathbf{U}g_\theta(\boldsymbol{\Lambda})\mathbf{U}^T x \tag{9}$$

Where $*_G$ represents a graph convolution operation. Since the convolution operation of graph signals is equal to the product of these signals transformed to the spectral domain through the graph Fourier transform Simonovsky and Komodakis (2017), the above formula can be understood as respectively transforming g_θ and x to the spectral domain, then multiplying their transform results, and then performing Fourier The final result of the convolution operation is obtained by inverse transformation.

Temporal graph convolution: After obtaining the neighborhood information of each spatial node through the spatial graph convolution operation, further stack the temporal convolution layer to update the node signal by combining the information on adjacent time slices.

$$\mathcal{X}_{\mathcal{P}} = \operatorname{Re} LU \left(\Phi * \left(\operatorname{Re} LU \left(g_\theta *_G \hat{\mathcal{X}}_{\mathcal{P}} \right) \right) \right) \in \mathbb{R}^{F \times N \times T} \tag{10}$$

In the above formula, $*$ is the operation of standard convolution, Φ is the parameter of the temporal convolution kernel, and the ReLU is selected as the activation function. The spatio-temporal convolution module can capture the spatio-temporal dependencies of traffic flow data well. The spatio-temporal attention module and the spatio-temporal convolution module constitute the STAC module. Finally, multiple spatio-temporal blocks are stacked to further extract dynamic spatio-temporal dependencies over a larger range.

4.3. Temporal Gated Convolution Module

This paper designs a new Dual gated unit (D-GTU) convolution module to capture long-term time dependency information in traffic flow data. The specific structure of this module is shown in Figure 1(c). It mainly consists of two gated convolution unit (GTU) Dauphin et al. (2017) modules.

The input of the temporal gated convolution module is $\mathcal{X}_{\mathcal{P}}$. The general GTU uses a convolution kernel with doubled channel number, where the size of the convolution kernel is $1 \times S$. The GTU in time can be defined in the following form:

$$\Gamma *_\tau \mathcal{X}_{\mathcal{P}} = \phi(A) \odot \sigma(B) \tag{11}$$

In the above formula, $*_\tau$ is the gated convolution operator, ϕ is the tan activation function, σ is the sigmoid activation function, A and B are the first half and second half of the channel size of $\mathcal{X}_{\mathcal{P}}$ respectively. By stacking gated convolution to expand the temporal receptive field, the model's ability to extract long-term dependencies in the data is improved. Therefore, we designed D-GTU, D-GTU as follows:

$$\begin{aligned} X_{\text{out}} = & \text{M-GTU}\left(\mathcal{X}_{\mathcal{P}}\right) = \left(\left(\text{Pooling}\left(\Gamma_1 *_\tau \mathcal{X}_{\mathcal{P}}\right) \times \right.\right. \\ & \left.\left. \text{Pooling}\left(\Gamma_2 *_\tau \mathcal{X}_{\mathcal{P}}\right)\right)\right) \end{aligned} \tag{12}$$

In the above formula, Γ_1 and Γ_2 represent convolution kernels of $1 \times S_1$ and $1 \times S_2$ respectively. The D-GTU designed in this paper has advantages in using multi-scale convolution to extract long-term temporal features of traffic data.

4.4. Trend Assistance

Existing traffic flow prediction methods generally use traffic volume to train the model, but there are many random external factors affecting the traffic flow data collected by sensors in the traffic road network, such as the geographical location of the city, the degree of damage to surrounding roads, etc. The distribution of traffic flow data shows an irregular state, and the amount of traffic data used for training is limited, so it is difficult to fit accurately. However, we found that the trend of traffic flow data is concentrated.

Based on the above findings, we designed a new auxiliary training method that focuses on the trend of the data rather than the real data. Essentially, traffic flow prediction is to find a mapping function f to predict the next T' graph signals based on the previous T graph signals. Therefore, for traffic flow data, there are $X \to Y$ training samples.

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_T] \in \mathbb{R}^{|V| \times T} \tag{13}$$

$\boldsymbol{X}$ is a matrix of T consecutive graph signals.

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \cdots, \boldsymbol{y}_T] \in \mathbb{R}^{|V| \times T'} \tag{14}$$

$\boldsymbol{Y}$ is the predicted matrix of the next T' graph signals. The traffic flow trend of the i-th sensor $v_i \in V$ in the road network at time $T + t$ is represented as Tr_{it}:

$$Tr_{it} = \frac{y_{it} - x_{iT}}{x_{iT}}. \quad (t = 1, 2, 3, \cdots, T') \tag{15}$$

The trend Tr_{it} is the relative change of y_{it} relative to the last observed flow of the sensor v_i. Finally, given a training sample $X \to Y$, a trend matrix Tr with the same shape as Y is obtained.

5. Experiments

In order to evaluate the performance of our model, extensive experiments are conducted on real-world traffic datasets. The statistical information of three datasets is summarized in Table 1.

5.1. Datasets

We conducted experiments on three real traffic datasets, namely PEMS03, PEMS04, and PEMS08. The data were extracted from the California Caltrans Performance Measurement System (PeMS) Chen et al. (2001) in three different areas of California. The Caltrans Performance Measurement System collects real-time data every 30 seconds. The original traffic flow data is aggregated at 5-minute intervals. There are 12 time steps per hour in the traffic data. Each sensor in PEMS04 and PEMS08 contains three types of traffic-related data: flow, occupancy, and speed. In this paper, we use traffic flow data.

DataSets	Days	Sensors	Edges	Data
PEMS03	91	358	547	26208
PEMS04	59	307	340	16992
PEMS08	62	170	295	17856

Table 1: Datasets

5.2. Baseline methods

We use the following 11 advanced methods as baselines for comparion experiments. (1)FC-LSTM Sutskever et al. (2014) has fully connected LSTM hidden units, which is a special type of RNN mmodel. (2)TCN Bai et al. (2018) is effective in learning local and global time correlations. (3) DCRNN Li et al. (2017) integrates graph convolutions into an encoder-decoder gated recurrent unit. (4)STGCN Yu et al. (2017) integrates graph convolutions into one-dimensional convolutional units. (5)ASTGCN Guo et al. (2019) is an attention-based spatial-temporal graph convolutional network. (6) STSGCN Song et al. (2020) directly extracts local spatio-temporal associations and uses multiple modules to capture spatio-temporal heterogeneity of the network. (7)STFGNN Li and Zhu (2021) proposes a new spatio-temporal fusion graph to model spatio-temporal correlations. (8)STGODE Fang et al. (2021) connects continuous differential equations with node representations of the transportation road network, enabling the construction of deeper networks. (9)Z-GCNET Chen et al. (2021) introduces time-aware zigzag patterns into the graph structure and designs zigzag topological layers for graph convolutional networks. (10)AGCRN Bai et al. (2020) incorporates learnable embeddings of nodes into graph convolutions.(11) DMSTAGNN Lan et al. (2022) utilizes spatio-temporal perceptions from historical traffic data instead of relying on predefined static adjacency matrices.

5.3. Experimental Settings

To compare our model with the baseline methods, we divide the three datasets into training sets, validation sets and test sets in a 7:1:2 ratio. This paper uses the historical traffic flow of one hour to predict the future one hour of traffic flow. Therefore, $T = T' = 12$ is set in the experiments. In the random walk strategy, the parameters a and b are both set to 1. When constructing the spatio-temporal correlation graph, this paper considers the 10 nearest neighbors of each sensor. In addition, the sensor embedding size is set to 128, the random walk length L is 25, and the window threshold Δ is 10.

5.4. Experiment results and analysis

5.4.1. Comparison Results

Dataset	Metric	FC-LSTM	TCN	DCRNN	STGCN	ASTGCN	STSGCN	AGCRN	STFGNN	STGODE	Z-GCNET	DMSTAGNN	STALGNN
PEMS03	MAE	21.33	19.31	18.18	17.49	18.05	17.48	*15.98	16.77	16.5	16.64	15.57	14.81
	MAPE(%)	22.33	19.86	18.91	17.15	17.02	16.78	*15.23	16.3	16.69	16.39	14.68	14.29
	RMSE	25.11	33.24	30.31	30.12	30.13	29.21	*28.25	28.34	27.84	28.15	27.21	23.81
PEMS04	MAE	26.24	23.11	24.07	22.7	21.85	21.19	19.83	19.83	20.84	19.5	19.3	19.27
	MAPE(%)	19.3	15.48	17.12	14.59	14.11	13.9	12.97	13.02	13.77	12.78	12.7	12.65
	RMSE	40.49	37.25	38.12	35.55	34.54	33.65	32.26	31.88	32.82	31.61	31.46	31.4
PEMS08	MAE	22.2	22.69	17.86	18.02	18.7	17.13	15.95	16.64	16.81	15.76	15.67	15.15
	MAPE(%)	15.02	14.04	11.45	11.4	11.64	10.96	10.09	10.6	10.62	10.01	9.94	9.61
	RMSE	33.06	35.79	27.83	27.83	28.66	26.8	25.22	26.22	25.97	25.11	24.77	24.37

Table 2: Performance comparison of different approaches on three datasets.

Table 2 shows the experimental results of STALGNN and baseline methods. It can be seen that our STALGNN achieves the best results on all metrics of the three datasets. The prediction results of traditional time series analysis methods are usually not ideal, indicating that these methods have limited ability to model nonlinear and complex traffic data. By comparison, deep learning based methods usually obtain better prediction results than traditional time series analysis methods. Since our spatio-temporal correlation graph is a learnable graph structure, it can help the model capture the spatio-temporal dependencies between nodes. In addition, the spatio-temporal attention mechanism and stacked graph convolutions we use can better capture the dynamic changes in the data and significantly improve the prediction performance.

Comparing the evaluation metrics of the PEMS04 dataset, it can be seen that STALGNN performs better on the PEMS03 and PEMS08 datasets. This may be because the road network is sparse in the PEMS04 dataset, resulting in an inaccurate learning spatio-temporal correlation graph. The high traffic missing rate can also lead to poor model performance.

5.4.2. Ablation Study

To verify the effectiveness of individual modules in STALGNN, we performed ablation experiments on the Pems08 dataset, the experimental data are shown in Table 3, and we make the following variants of STALGNN: (1) Rem-STAC: Completely remove the spatio-temporal attention mechanism and graph convolution module; (2) Rem-Trend: Remove traffic trend auxiliary information; (3) Rem-DGTU: Remove the dual gated units. Based on the above three model variants, we conducted ablation experiments on the PEMS03, PEMS04 and PEMS08 datasets respectively. Figures. 2, Figures. 3 and Figures. 4 show the MAE, MAPE and RMSE measurement results. It can be seen that the performance of our STALGNN is superior to other variants, which also verifies the effectiveness of each module in our model.

Model	MAE	MAPE	RMSE
Rem-STAC	15.24	10.53	25.89
Rem-TREND	15.42	9.89	24.68
Rem-GTU	25.13	15.38	38.44
Rem-Graph	16.43	10.24	25.94
STALGNN	15.15	9.61	24.37

Table 3: Ablation Study on PEMS08 dataset

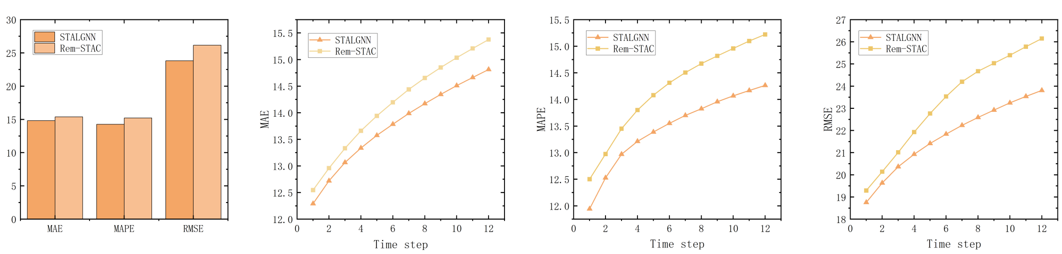

Figure 2: Ablation experiment of STAC

Rem-STAC: The spatio-temporal attention mechanism combined with spatio-temporal convolution captures the dynamic spatio-temporal characteristics of traffic data, thereby improving model performance. The measurement results of the STAC module ablation experiment are shown in Figure 2.

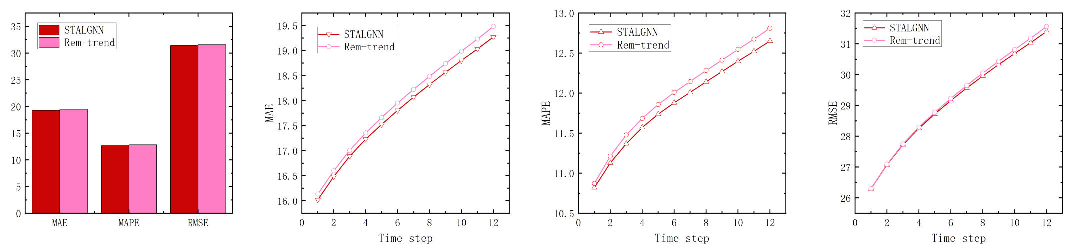

Figure 3: Ablation experiment of Trend

Rem-TREND: Adding trend auxiliary information to train the STALGNN model can improve the performance of the model on all datasets. Due to the consistent trend, training model with the trend feature an improve the perforamnce of the model. The measurement results of the ablation experiment of trend auxiliary information are shown in Figure 3.

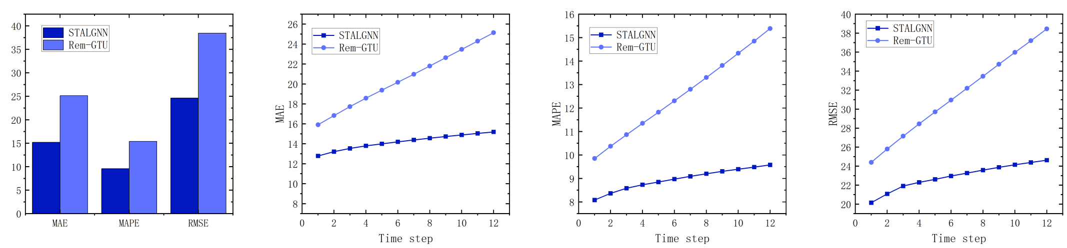

Figure 4: Ablation experiment of GTU

Rem-GTU: For the temporal-gated convolution module, it can compensate for the STALGNN model's ability to extract long-term time dependencies, thereby improving the performance of STSGNN. The measurement results of the time-gated convolution module ablation experiment are shown in Figure 4.

6. Conclusion

In this paper, we propose a new spatio-temporal aware learning attention graph neural network and successfully apply it to traffic forecasting. The model uses a new method for generating spatio-temporal correlation graphs that can effectively capture hidden spatial dependencies. It combines spatio-temporal attention mechanisms and spatio-temporal convolutions to capture the spatio and temporal features of traffic flow data simultaneously.

Trend auxiliary information is supplemented when calculating the final loss. In addition, a dual temporal-gated convolution module is designed to extract long-term temporal dependencies of traffic data. Experiments on three real datasets show that the performance of the proposed model is superior to other baseline models.

Acknowledgments

This work was supported by Fundamental Research Foundation of Universities in Heilongjiang Province of China (2021KYYWF0010)

References

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33:17804–17815, 2020.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Chao Chen, Karl Petty, Alexander Skabardonis, Pravin Varaiya, and Zhanfeng Jia. Freeway performance measurement system: mining loop detector data. *Transportation Research Record*, 1748(1):96–102, 2001.

Yuzhou Chen, Ignacio Segovia, and Yulia R Gel. Z-gcnets: Time zigzags at graph convolutional networks for time series forecasting. In *International Conference on Machine Learning*, pages 1684–1694. PMLR, 2021.

Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. In *International conference on machine learning*, pages 933–941. PMLR, 2017.

Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 364–373, 2021.

Xiaocheng Feng, Jiang Guo, Bing Qin, Ting Liu, and Yongjie Liu. Effective deep memory networks for distant supervised relation extraction. In *IJCAI*, volume 17, pages 1–7, 2017.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 922–929, 2019.

Peng Han, Jin Wang, Di Yao, Shuo Shang, and Xiangliang Zhang. A graph-based approach for trajectory similarity computation in spatial networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 556–564, 2021.

Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International Conference on Machine Learning*, pages 11906–11917. PMLR, 2022.

Mengzhang Li and Zhanxing Zhu. Spatial-temporal fusion graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4189–4196, 2021.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

Zheng Lu, Chen Zhou, Jing Wu, Hao Jiang, and Songyue Cui. Integrating granger causality and vector auto-regression for traffic prediction of large-scale wlans. *KSII Transactions on Internet and Information Systems (TIIS)*, 10(1):136–151, 2016.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3693–3702, 2017.

Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu 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*, volume 34, pages 914–921, 2020.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

JWC Van Lint and CPIJ Van Hinsbergen. Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1):22–41, 2012.

Chun-Hsin Wu, Jan-Ming Ho, and Der-Tsai Lee. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, 5(4):276–281, 2004.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

Jiani Zhang, Xingjian Shi, Junyuan Xie, Hao Ma, Irwin King, and Dit-Yan Yeung. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. *arXiv preprint arXiv:1803.07294*, 2018.

Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.