SAX-STGCN: Dynamic Spatio-Temporal Graph Convolutional Networks for Traffic Flow Prediction

BIN LEI, PENG ZHANG, YIFEI SUO, AND NA LI
School of Civil Engineering, Xi’an University of Architecture and Technology, Xi’an 710055, China
Corresponding authors: Peng Zhang (zhangpeng.sx@qq.com) and Bin Lei (leibin@xauat.edu.cn)
This work was supported by the Shaanxi Transportation Science and Technology Project under Grant 20-05R.

ABSTRACT Accurate, timely, and reliable traffic flow prediction is essential for an intelligent transportation system due to the complex spatio-temporal correlation of traffic flow. The prediction model based on graph convolution neural network (GCN) has become mainstream in recent years. However, most of the prediction models based on GCN only use an adjacency matrix to characterize the spatial correlation of traffic flow, ignoring the complex and dynamic relationship between road network adjacent nodes and missing the hidden connection between the global nodes of the road network. This paper proposes a SAX-STGCN network for traffic flow prediction to solve the above problems. The SAX-STGCN model uses symbolic approximation (sax) to obtain the similarity of the historical data of the predicted node in the previous period, including adjacent nodes and non-adjacent nodes, forming a similarity matrix to replace the original adjacency matrix composed of 0 and 1, which is defined as a global sax-correlation matrix to characterize the correlation between nodes in the road network and capture the implicit spatial relationships in the road network. Then, based on the dynamic global sax-correlation matrix, GCN is used to capture the spatial correlation of traffic flow, and gated recurrent unit (GRU) is used to capture the temporal correlation of traffic flow. The prediction accuracy is better than the baseline and has long-term prediction ability.

INDEX TERMS Intelligent transportation system, traffic prediction, graph convolutional network, symbolic aggregate approximation.

I. INTRODUCTION
With the development of urbanization, the rapid growth of the urban population and motor vehicles has brought tremendous pressure on urban transportation. The urban traffic control system based on ITS has gradually developed from passive to active management. Active management’s core is responding to possible traffic conditions in advance through traffic control and guidance after sensing the changes in urban traffic. Traffic flow prediction is the first step to achieving active management, and it is also a necessary condition for various applications such as traffic control [1], path optimization [2], and vehicle scheduling [3]. At the same time, it provides a decision-making basis for traffic control and risk assessment and determines the effectiveness of traffic guidance measures. Therefore, traffic flow prediction has aroused great interest in researchers.

However, traffic flow prediction has always been challenging due to traffic flow data’s highly complex temporal and spatial dependence. The previous prediction models mainly predict the future traffic state from the perspective of temporal correlation by learning the historical traffic characteristics. These characteristics are primarily manifested as proximity, periodicity, and randomness. The spatial correlation of traffic flow is mainly displayed in transitivity and spatial similarity. Transport performance in the upstream road traffic will transfer to the downstream road to a certain extent, and the downstream road traffic condition will also affect the upstream road [4]. At the same time, due to the diffusion of the road network, the traffic condition of a particular section...
will also affect others. The spatial similarity is manifested in that traffic flows in areas with the same function will exhibit similar characteristics. With the rise of graph neural network research, more and more models [5], [6], [7] predict traffic flow based on graph convolution neural networks to characterize the spatial correlation of road networks.

However, graph convolution neural network based on an adjacency matrix can only represent the connectivity between road network nodes, ignoring the distinct influence of different nodes. At the same time, due to the randomness of traffic flow in the time dimension, the effect between road network nodes will also change with the dynamic change in traffic flow. Therefore, the fixed adjacency matrix is difficult to characterize the spatial correlation of the dynamic shift in a road network. In addition, due to the influence of city functional zoning, POI distribution, and other factors, non-adjacent nodes in the road network also have a specific hidden correlation [8].

As shown in Figure 1, these six nodes are non-adjacent nodes. In time 1, node 2 and node 3 show approximate sequence characteristics, while in time 2, node 2 and node 3 show opposite sequence characteristics. In time 3, nodes 4 and 6, which were weakly correlated, have similar sequence characteristics.

As shown in Figure 2, these six nodes are adjacent in the road network, node 1, node 3, and node 6 continue to show similar sequence characteristics, and node 2 and node 5 show other sequence characteristics. In contrast, node 4 is opposite to node 2 and node 5 in time 4 but shows similar characteristics in time 5. Therefore, the fixed adjacency matrix of 0 and 1 is challenging to characterize the influence size and dynamic node correlation between adjacent road network nodes. It is also challenging to capture the hidden spatial relationship.

Given the different correlation magnitudes and the dynamic changes of correlations between nodes, this paper proposes a symbolic aggregation representation spatio-temporal graph convolution model (SAX-STGCN). The model performs symbolic aggregation on the historical traffic flow data before the predicted node. It uses its sequence similarity to reconstruct the adjacency matrix to characterize the influence between the road network nodes to improve prediction accuracy. The main contributions of this paper are as follows:

1) For the correlation of adjacent nodes, it is proposed to use sax (Symbolic Aggregate Approximation) to symbolize the traffic state sequence and then use the similarity to represent the node correlation.

2) Aiming at the implicit spatial correlation in the road network, the sax method is used to represent the traffic state sequence of non-adjacent nodes in the road network to represent the implicit spatial correlation in the road network.

3) Given the dynamic change of the correlation between the road network nodes, the two methods (1) and (2) are combined. The traffic state data of the previous period of the predicted node is used to construct a dynamically changing global sax-correlation matrix, which is used to dynamically represent the spatial correlation of adjacent nodes and implicit spatial correlation.

Use the graph convolutional neural network that introduces the dynamic global sax-correlation matrix to capture the spatial correlation, and then use the GRU to capture the temporal characteristics of the traffic state to form a SAX-STGCN model to obtain the final prediction result.

II. RELATED WORK
Traffic flow forecasting research has a long history. Forecasting methods are mainly divided into statistical models,
machine learning, and deep learning. Statistical models mainly include the historical average method (HA), auto regressive integrated moving average (ARIMA) [9] and its variants [10], [12]. These models have the advantages of low computational complexity but rely on the assumption of data stability. The actual traffic flow data often have strong randomness and complex variability, so the effect in practical application is not ideal. Machine learning models can further model complex data, such as support vector regression (SVR) [13], K-nearest neighbor (KNN) [14], etc. However, the performance of such models depends on the design of feature engineering, and the overall nonlinearity is limited. With the rapid progress of deep learning technology, various deep learning models are used to alleviate these limitations. Early work included multi-layer perceptron (MLP) [15] and deep belief network (DBN) [16], which improved prediction accuracy. Subsequently, models such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs) [17], and gated recurrent units (GRUs) [18] have emerged, widely used in studying time series such as traffic flow prediction. It is worth mentioning that EnLSTM-WPEO [19] uses ensemble learning of diverse LSTMs and PEO-based NNCT weight integration, which effectively improves the prediction accuracy and alleviates the time lag problem. However, these models ignore the spatial correlation of traffic flow data only from the perspective of temporal correlation.

From spatio-temporal correlation, some scholars have used convolutional neural networks (CNNs) [20] and deeper residual networks (ResNets) [21] in computer vision for traffic flow prediction. CNN is usually used for traffic flow prediction by grid division to establish a regional-level spatial matrix of traffic flow, stack the spatio-temporal matrix of traffic flow according to time, and predict using the time series model. The spatio-temporal correlation of traffic flow is comprehensively considered. The representative work includes STRResNet [22], CNN-LSTM [23], FCL-Net [24], etc.

Although CNN can model the spatial correlation of traffic flow to a certain extent, the road network is a typical graph structure. Its spatial topology is often lost when using grid processing, so it cannot describe the spatial-temporal correlation of the road network in essence. With the rise of graph depth learning, GCN has become the most advanced method in many fields. GCN can effectively extract the dependency between road network nodes as a natural graph structure. Therefore, many GCN-based models have been widely used in traffic flow prediction in the past few years, such as AST-GCN [25], T-GCN [4], AST-GCN [26], TGC-LSTM [27], MRA-BGCN [28], and other models have achieved good results in traffic prediction. However, in these models, the Laplacian matrix of graphs is strictly defined by GCN [29]. The adjacency matrix of the fixed graph composed of 0 and 1 is used, and only the connectivity between nodes is considered. Some scholars use threshold Gaussian Kernel to process the distance matrix of adjacent road network nodes and construct the distance adjacency matrix based on adjacent road network nodes, such as DNRNN [30], KW-GCN [31], etc.

The distance matrix is more expressive than the connected matrix. Nevertheless, it is challenging to characterize the spatial correlation of dynamic traffic flow as a fixed matrix.

Therefore, this paper proposes to dynamically construct the graph structure of the road network according to the input traffic flow features instead of the original adjacency matrix. The sax method is used to describe the flow correlation between adjacent nodes. At the same time, to capture the hidden spatial relationship for non-adjacent nodes, the sax method is also used to describe the spatial correlation. A dynamic global sax-correlation matrix is constructed to characterize the spatial correlation better. It can enhance the expressive ability of the model, thereby improving the prediction accuracy.

III. METHOD

A. PROBLEM DEFINITION

In this paper, we take the traffic speed prediction as an example to illustrate our work.

Definition 1: Road network $G$. We describe the road network as a graph $G = (V, A, E)$, $V = \{v_1, v_2, \ldots, v_N\}$ represents the set of nodes in the road network, and $N$ represents the number of nodes in the road network. Nodes can be sensors deployed on the road network or road intersections or sections. $E$ represents the set of edges, and if node $v_i$ is adjacent to node $v_j$, there is edge $e_{ij}$ between them; $A \in \mathbb{R}^{N \times N}$ represents the adjacency matrix of the road network topology graph. The adjacency matrix can be represented as $0$ and $1$ of binary according to different tasks, that is, $0$ means that there is no connection between the two nodes while $1$ means connectivity.

Definition 2: Traffic speed matrix $X^{T \times N}$. $T$ represents the length of the historical speed sequence, $N$ represents the number of nodes, $X^t_i$ represents the speed of node $i$ at the time $t$, $X^t_i$ represents the historical speed sequence of node $i$, $X^t$ represents the speed characteristics of all nodes in the time $t$ road network $G$.

Definition 3: Traffic forecasting task. The essence of traffic prediction is to learn a mapping function $f(\cdot)$, and map the traffic state of the future $Q$ time steps through the given historical traffic state of the road structure $G$ and $P$ time steps, as shown in the Figure 3, and Eq.1:

$$[Y^1, \ldots, Y^Q] = f(X^1, X^2, \ldots X^P; G)$$  \hspace{1cm} (1)

B. SAX PROCESS

There are few studies using time series similarity to characterize the spatial correlation of road network. DC-STGCN [32] directly uses the Pearson correlation coefficient to characterize the correlation of traffic flow at road network nodes, but does not use the time series representation method. Although the complexity is low, it is easily affected by mutation points and cannot well measure the similarity trend of time series [33].
Traffic flow data is specific time series data with complex and high dimensions. Feature extraction and representation of time series data requires dimensionality reduction and denoising based on maintaining and reflecting sequence morphology and basic information as much as possible to reduce the computational cost and improve the efficiency of data processing, mining, and analysis. There are many feature extraction and representation methods [33], [34], [35], among which symbolic aggregate approximation (sax) [36] is considered an atypical and effective symbolic representation method.

In this paper, the traffic flow data are processed by sax, and on this basis, the similarity of traffic flow between nodes is calculated to characterize the correlation between nodes. The construction process of the sax adjacency matrix and global sax-correlation matrix is as follows:

Assuming that the historical sequence length of the input model is \( P \) and the traffic velocity matrix is \( X \in \mathbb{R}^{P \times N} \).

Step 1: Normalize the historical sequence at each node.

\[
\mu_i = \frac{1}{P} \sum_{p=1}^{P} X_{p}^i \\
\sigma_i = \sqrt{\frac{1}{P-1} \sum_{s=1}^{P} (X_{s}^i - \mu_i)^2} \\
\hat{X}_{s}^i = \frac{X_{s}^i - \mu_i}{\sigma_i}
\]

where \( \mu_i \) and \( \sigma_i \) represent the mean and standard deviation of the traffic speed at sensor \( i \) in \( P \) time periods, respectively, and \( \hat{X} \) represents the standardized historical traffic speed matrix.

Step 2: PAA(Piecewise aggregate approximation) is performed on the input data, dividing the speed sequence with length \( P \) into \( W \) segments, and obtaining the speed matrix \( C \in \mathbb{R}^{W \times N} \) obtained after PAA conversion, as shown in Eq.(5):

\[
\left\{
\begin{array}{l}
C_{h}^{i} = \frac{1}{P} \sum_{j=1}^{P} X_{j}^{i} \\
\text{where } C_{h}^{i} \text{ represents the speed detection average value of the } h \text{ segment of the sensor } i, \text{ where } h = 1, 2, \ldots, W.
\end{array}
\right.
\]

Step 3: Discrete sequence. Suppose the symbol set is \( \Pi \) and the number of elements is \( o \). The Gaussian probability density curve is divided into interval \([\theta_{k-1}, \theta_{k}]\), where \( k = 1, 2, \ldots, o \), and the area under each interval curve is \( 1/o \).

Step 4: Symbolized sequence. Mapping \( C \) to the corresponding symbol indicates, as shown in Eq.6:

\[
\left\{
\begin{array}{l}
\tilde{C}_{h}^{i} = \Pi_k \\
\theta_{k-1} < C_{h}^{i} \leq \theta_{k}
\end{array}
\right.
\]

where the symbolized sequence is \( \tilde{C} \).

Step 5: Obtain sequence similarity after symbolization, obtain global sax adjacency matrix:

\[
\hat{A}_{ij} = \left\{ \begin{array}{ll}
\text{Dist} (s_{i}, s_{j})^{-1} & \text{if } i \neq j \\
1 & \text{if } i = j
\end{array} \right.
\]

\[
\text{Dist} \left( \tilde{C}_{i}, \tilde{C}_{j} \right) = \sqrt{\frac{P}{W} \sum_{h=1}^{W} \left( \tilde{C}_{h}^{i} - \tilde{C}_{h}^{j} \right)^2}
\]

where \( \tilde{C}_{h}^{i} - \tilde{C}_{h}^{j} \) can be obtained by querying the symbol set distance table.

\( \hat{A} \) describes the influence of traffic flow between adjacent nodes from the perspective of time series similarity and is used to describe the implicit spatial relationship between non-adjacent nodes, replacing the original 0 and 1-valued adjacency matrix.

C. FRAMEWORK

The model framework of SAX-STGCN is shown in Figure 4. It consists of the sax Process module and STGCN module, where the sax Process module is used to extract dynamic global sax-correlation matrix, ST Block consists of GCN and GRU, GCN is used to extract spatial features, GRU is used to extract temporal features, and get the final prediction result.
D. SPATIAL DEPENDENCE MODELING

Graph neural network (GNN) is a deep learning method mainly used to model graph structure data. It learns the relationship between graph structure and node attributes through information transmission between nodes and neighbors. An essential model of GNN is graph convolutional networks (GCN). GCN introduces convolution operation from Euclidean space into a graph structure, which can be expressed in Eq.9-10:

\[
L^{sym} = D^{-1/2}(D - A)D^{-1/2} = I - D^{-1/2}AD^{-1/2} \quad (9)
\]

\[
H^{(l)} = \sigma \left( L^{sym}H^{(l-1)}\theta^{(l-1)} \right) \quad (10)
\]

where \(L^{sym}\) represents the renormalized Laplacian matrix to effectively prevent numerical instability; \(H^{(l-1)}\) represents the output of \(l - 1\) layer convolution; \(\theta^{(l-1)}\) represents the parameters of \(l - 1\) layer convolution.

In this paper, two-layer graph convolution is used to capture spatial correlation:

\[
\hat{y} = \sigma \left( L^{sym} \text{ReLU} \left( L^{sym}XW_{(0)} \right) W_{(1)} \right) \quad (11)
\]

E. TEMPORAL DEPENDENCE MODELING

Gated Recurrent Unit (GRU) [37] is an extended model for RNN. GRU is generally easier to train than LSTM, and the effect is equivalent, so it is widely used in time series modeling. Unlike LSTM, GRU does not introduce new memory units but controls information access and forgetting by using update gate \(Z_t\) and reset gate \(R_t\). The GRU procedure can be formulated as Eq.12-15:

\[
R_t = \sigma \left( X_t \cdot W_{xr} + H_{t-1}W_{hr} + b_r \right) \quad (12)
\]

\[
Z_t = \sigma \left( X_t \cdot W_{xz} + H_{t-1}W_{zh} + b_z \right) \quad (13)
\]

\[
\hat{H}_t = \tanh \left( X_t \cdot W_{xh} + (R_tH_{t-1})W_{bh} + b_h \right) \quad (14)
\]

\[
H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \hat{H}_t \quad (15)
\]

where \(gc(\cdot)\) represents the graph convolution operation, \(W\) and \(b\) are learnable parameters, and \(H\) represents the hidden state.

F. SYMBOL AGGREGATION APPROXIMATION SPATIAL-TEMPORAL GRAPH CONVOLUTION NETWORK

In this paper, the dynamic global sax-correlation matrix is constructed using time series similarity to replace the original adjacency matrix. The global sax correlation matrix is used to characterize the implicit spatial relationship in the road network. The specific steps of the SAX-STGCN model are as follows Eq.16-19:

\[
R_t = \sigma \left( W_{xr} \cdot gc \left( X_t^l, \hat{A}, P \right) + H_{t-1}W_{hr} + b_r \right) \quad (16)
\]

\[
Z_t = \sigma \left( W_{xz} \cdot gc \left( X_t^l, \hat{A}, P \right) + H_{t-1}W_{zh} + b_z \right) \quad (17)
\]

\[
\hat{H}_t = \tanh \left( W_{xh} \cdot gc \left( X_t^l, \hat{A}, P \right) + (R_t \odot H_{t-1})W_{bh} + b_h \right) \quad (18)
\]

\[
H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \hat{H}_t \quad (19)
\]

where \(gc(\cdot)\) represents the graph convolution operation, \(W\) and \(b\) are learnable parameters, and \(H\) represents the hidden state.

Traffic flow prediction aims to make the prediction results as close as possible to the actual traffic flow state. The loss function used in this model is as follows:

\[
Loss = \begin{cases} 
0.5(y - \hat{y})^2 & \text{if } |y - \hat{y}| < n \\
|y - \hat{y}| - 0.5 \ast \beta & \text{otherwise} 
\end{cases} \quad (20)
\]

\(y\) and \(\hat{y}\) represent the actual value and predicted value, respectively, the value of \(\beta\) is 1.0.

IV. EXPERIMENTS

A. DATA DESCRIPTION

This experiment takes two widely used actual datasets to verify the prediction effect of SAX-STGCN:

1) SZ-TAXI [4]: This dataset comprises taxi tracks in Shenzhen from January 1 to January 31, 2015. 156 main roads in Luohu District are selected for data collection. The experimental data mainly includes two parts. One is a 156 \(\times\) 156 adjacency matrix to describe the spatial relationship between roads. The other is the feature matrix, which describes the speed change on each road with time. Each row represents a road, and each column shows the traffic speed on the road during different periods. The traffic speed on each road is recorded every 15 minutes.

2) PEMSBAY: The dataset is collected by California Transportation Agencies Performance Measurement System (PEMS), containing 325 sensors, and records the road network traffic speed for six months from January 1, 2017, to May 31, 2017. Like SZ-TAXI, each row represents the traffic speed of the sensor at different periods, each column represents an observation point, and the traffic speed of each observation point is recorded every 5 minutes.

B. EVALUATION METRICS

The following indicators are used to evaluate the model’s performance by evaluating the prediction results:

1) Mean Absolute Error (MAE):

\[
MAE = \frac{1}{n} \sum_{t=1}^{n} |y^t - \hat{y}^t| \quad (21)
\]

2) Root Mean Squared Error (RMSE):

\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y^t - \hat{y}^t)^2} \quad (22)
\]

C. PARAMETER SETTINGS

The CPU of the experimental platform adopts Intel Xeon(R) Platinum 8259CL CPU@2.50GHz, and the GPU adopts an NVIDIA Tesla T4 graphics card. During the experiment, the Z-score normalization operation was performed on the data
set, and the data set was divided into 80% training set and 20% test set. The initial learning rate of the SAX-STGCN model is 0.001, the weight decay is 5e-5, and the batch size is 8. SAX-STGCN uses the sax method to calculate the similarity of historical traffic flow data before prediction to obtain a dynamic global sax-correlation matrix. Therefore, selecting the length of historical traffic flow is necessary for constructing the dynamic correlation matrix.

In our experiments, the historical sequence lengths are selected from [4], [8], [12], [16], [20], [24], and the changes in prediction accuracy are analyzed. As shown in Figure 5, the horizontal axis represents the length of the historical sequence, and the vertical axis represents the results of RMSE and MAE corresponding to different lengths. Figure 5(a) shows that the accuracy varies with different historical sequence lengths in the SZ-TAXI dataset. When the historical sequence length is 8, the model performs the best, and the accuracy decreases as the sequence length increases, so the sequence length 8 is chosen. Figure 5(b) shows the variation of accuracy with different historical sequence lengths in the PEMS-BAY dataset, with the best performance when the historical sequence length is 16.

Similarly, Figure 6 shows the effect of the number of hidden layer units on the model performance at different numbers. The horizontal axis represents the number of hidden layer units, and the vertical axis represents the RMSE and MAE values. The number of hidden layer units is selected from [4], [8], [16], [32], [64], [128]. It can be seen from Figures 6(a) and 6(b) that in the SZ-TAXI and PEMS-BAY datasets, the best results are obtained when the number of hidden layer units is both 16.

Finally, we determined that in the SZ-TAXI dataset, the length of the historical sequence is 8, and the number of hidden layer units is 16. In the PEMS-BAY dataset, the length of the historical sequence is selected as 16, and the number of hidden layer units is 16.

D. BASELINES
SAX-STGCN model will be compared with the following eight baseline models:

1) Historical average model (HA): HA is a historical average model that uses historical average as a prediction value. The historical period length of this paper is one day, and the average value of three historical periods is used as the prediction result.

2) Autoregressive integral moving average (ARIMA): It is a classical time series prediction model. The prediction parameters p, D and Q are set to 2, 1 and 2, respectively.

3) Support vector regression (SVR): The prediction results are obtained through historical data training. The results of this paper use a linear kernel function, and the penalty factor is set to 0.001.

4) Gated recurrent unit (GRU): It is a variant of RNN, which introduces a gating mechanism to effectively solve the gradient disappearance and gradient explosion problems in long sequence prediction. The number of Gru hidden units was 64 and 32 in SZ-TAXI and PEMS-BAY, respectively, and the learning rate was set to 0.001.

5) Time graph convolution model (T-GCN): In this model, GCN is used to capture the spatial correlation of traffic
flow, and GUR is used to capture the temporal correlation. The number of hidden layer units in the SZ-TAXI data set is set to 100, and the learning rate is set to 0.001. The number of hidden layer units in the PEMS-BAY is set to 64, and the learning rate is set to 0.001.

6) Attribute-Augmented spatiotemporal convolutional graph convolutional network (AST-GCN): It obtains the POI information of nodes in the SZ-TAXI data set, defines historical weather data as external factors, and designs an attribute enhancement unit to encode external factors and integrate them into the spatiotemporal map convolution model. The number of hidden layer units of AST-GCN is set to 100, and the learning rate is set to 0.001.

7) Diffusion convolutional recurrent neural network (DCRNN): It combines a cyclic neural network and diffusion convolution to model traffic flow’s inflow and outflow relationship. SZ-TAXI and PEMS-BAY experiments in this paper use the original paper parameter settings.

8) Multi-Range attentive bicomponent graph convolutional network (MRA-BGCN): It constructs a node graph according to the distance of the road network, introduces a multi-range attention mechanism, aggregates information in different neighborhood ranges, and automatically learns the importance of different ranges. The number of BGCGRU layers of MRA-BGCN is set to 2, the number of hidden cells is set to 64, The maximum hop of the bicomponent graph convolution is set to 3, the initial learning rate is set to 0.01, and the attenuation rate is set to 0.6 per 10 epochs.

E. EXPERIMENTAL RESULTS
Table 1 shows the performance comparison between the SAX-STGCN model and the baseline model. Deep learning-based methods (GRU, T-GCN, AST-GCN, MRA-BGCN, DCRNN) have higher accuracy than traditional methods. In the SZ-TAXI dataset, compared with SVR and ARIMA, the RMSE of SAX-STGCN is reduced by about 45.7% and 18.8%, and the MAE is reduced by 48.6% and 18.8%, respectively, at the 15-minute prediction horizon. From the spatiotemporal perspective, the MAE of SAX-STGCN using a dynamic sax-correlation matrix is reduced by 5.55% and 5.16%, respectively, compared with T-GCN and AST-GCN using a fixed adjacency matrix. At the longer 60-minute prediction horizon, it also improves. Compared with T-GCN, MRA-BGCN and DCRNN, MAE is reduced by 4.86%, 3.86%, 16.2% and 17.2%, respectively.

In the PEMS-BAY data set, the above characteristics are also shown. Compared with HA, SVR and ARIMA, the RMSE of SAX-STGCN at the 15-minute prediction horizon is reduced by about 53.8%, 28.1% and 21.8%, respectively, and the MAE is respectively decreased by 48.6% and 18.8% and 32.0%. Compared with T-GCN, MRA-BGCN and DCRNN, MAE is improved by 4.86%, 3.86%, 16.2% and 17.2%, respectively.

At the longer 60-minute prediction horizon, the improvements were 12.3%, 3.0%, and 7.2%, respectively.

Through the comparison of the above two datasets, we can find that SAX-STGCN almost achieves the best performance under both RMSE and MAE. SAX-STGCN significantly improved the short-term prediction of the 15-minute prediction horizon. With the increase in the prediction time, the improvement in the 60-minute prediction horizon is slightly reduced. In addition, compared to the smaller SZ-TAXI dataset, SAX-STGCN has a more significant improvement over the larger PEMS-BAY dataset, and we speculate that with the increase in the amount of training data, the model can learn more. There are many changes in the traffic status of the road network, so it has been improved.

| Dataset   | Time | Metric | HA   | ARIMA | SVR   | GRU   | T-GCN | AST-GCN | MRA-BGCN | DCRNN | SAX-STGCN |
|-----------|------|--------|------|-------|-------|-------|-------|---------|----------|-------|-----------|
|           |      | RMSE   |      |       |       |       |       |         |          |       |           |
| SZ-TAXI   | 15min| 4.29   | 7.24 | 4.84  | 3.99  | **3.92** | 4.03 | 4.32   | 4.50     |       | 3.93      |
|           |      | MAE    | 2.78 | 4.98  | 3.04  | 2.59  | 2.71  | 2.70   | 2.72     | 2.74 | 2.56      |
|           | 30min| 4.29   | 6.78 | 4.96  | 4.09  | 3.96  | 4.05  | 4.43   | 4.56     |       | 3.95      |
|           |      | MAE    | 2.78 | 4.67  | 3.28  | 2.69  | 2.73  | 2.73   | 3.21     | 3.23 | 2.64      |
|           | 60min| 4.29   | 6.77 | 5.21  | 4.07  | 4.00  | 4.10  | 4.56   | 4.64     |       | 3.99      |
|           |      | MAE    | 2.78 | 4.66  | 3.51  | 2.93  | 2.88  | 2.85   | 3.27     | 3.31 | 2.74      |
| PEMS-BAY  | 15min| 5.59   | 3.30 | 3.59  | 4.13  | 2.78  | *     | 2.72   | 2.95     |       | 2.58      |
|           |      | MAE    | 2.88 | 1.62  | 1.85  | 2.10  | 1.35  | *      | 1.29     | 1.38 | 1.26      |
|           | 30min| 5.59   | 4.42 | 5.18  | 4.65  | 3.77  | *     | 3.67   | 3.97     |       | 3.52      |
|           |      | MAE    | 2.88 | 2.11  | 2.48  | 2.33  | 1.71  | *      | 1.61     | 1.74 | 1.51      |
|           | 60min| 5.59   | 6.50 | 7.08  | 4.98  | 4.65  | *     | 4.46   | 4.74     |       | 4.38      |
|           |      | MAE    | 2.88 | 3.38  | 3.28  | 2.55  | 2.19  | *      | 1.98     | 2.07 | 1.92      |
F. ABSTRACTION EXPERIMENTS

To verify the contribution of the dynamic global sax-correlation matrix to the model, we design ablation experiments. We divide the matrices used to capture spatial correlations into two groups, dynamic or fixed and global or non-global.

Among them, non-global refers to only using sax processing to obtain the similarity of adjacent nodes to replace the original adjacency matrix, while global refers to processing all nodes. The dynamic matrix refers to the adjacency matrix or correlation matrix used to predict the traffic flow next time, which is obtained by sax processing the traffic flow data at the previous period of the prediction node. The fixed matrix refers to the matrix obtained by sax processing the data on the first day of the dataset. The PEMS-BAY dataset contains 288 data points a day and takes 16-time data as a group, totaling 18 groups. The sax- adjacency matrix is obtained respectively, and then the mean value is obtained to obtain the fixed sax-adjacency matrix. SZ-TAXI contains 96 data points daily, divided into 6 groups, and the repeat above operations.

Finally, we set up five different matrices to verify their effectiveness. They are: (1) Fixed adjacency matrix composed of original 0 and 1, (2) Fixed sax-adjacency matrix, (3) Fixed global sax-correlation matrix, (4) Dynamic sax-adjacency matrix, and (5) Dynamic global sax-correlation matrix.

In Table 2, it can be seen from the comparison results of the ablation experiments of the two datasets that compared with the adjacency matrix composed of the original 0 and 1, the matrix processed by the sax method has an impact on the performance of the model. Specifically, among the fixed matrices, the global sax-correlation matrix has a small contribution to the model improvement and even reduces the model performance. On the contrary, using the sax-adjacency matrix improves the model’s performance. The dynamic matrix can improve the model, whether the dynamic global sax-correlation matrix or the sax-adjacency matrix. The dynamic global correlation matrix can improve the model the most.

In summary, using the sax method to dynamically calculate the similarity of traffic flow to construct the correlation matrix does improve the model. It also confirms our hypothesis that the correlation between road network nodes changes dynamically with traffic flow. In contrast, the road network’s similarity of traffic flow between nodes can represent the correlation of road network nodes. The fixed global sax-correlation matrix shows a negative improvement in the model. We analyze that because it is obtained by the summed average of multiple global sax-correlation matrices, each pair of road network nodes will have a correlation and remain fixed; this does not correspond to reality and thus exhibits a negative lift. In contrast, a fixed sax-adjacency matrix contributes to model performance improvement. Furthermore, it also shows that the sax method does contribute to the model.
In order to better illustrate the prediction performance of this model, a comparative experiment is designed in this paper, the fixed adjacency matrix and the dynamic global sax correlation matrix are brought into the model for comparison, and a road in PEMS-BAY is selected for visual analysis as shown in Figure 7-9, where STGCN means that sax processing is not applicable. Only a fixed adjacency matrix consisting of 0 and 1 is used. We will analyze it from the following three perspectives:

1) Prediction horizons. The prediction results of traffic speed in the next 15, 30, and 60 minutes using historical data of the past hour are shown in Figures 7-9. The visualization results show that the dynamic sax-correlation matrix model can better fit the actual data in different prediction horizons. Specifically, short-term forecasting performance is better, especially 15 minutes short-term forecasting advantage, is more prominent; by contrast, long-term forecasting performance is weaker than short-term forecasting. We analyze that the reason is that the traffic flow changes dynamically over time, and the correlation between each traffic node also changes over time. Therefore, according to the sax-correlation matrix constructed at the last moment, the traffic flow state cannot be well characterized over time.

V. CONCLUSION
This paper proposes a model for traffic flow prediction called SAX-STGCN. Aiming at the traditional adjacency matrix that only characterizes the connectivity of the road and the fixed adjacency matrix is difficult to characterize the dynamically changing traffic flow graph structure of the road network. This paper uses the sax method to process the traffic flow from the perspective of time series similarity to obtain the global sax-correlation matrix. The matrix replaces the original matrix to characterize node correlations better and capture implicit spatial correlations, and the global sax-correlation matrix changes dynamically with traffic flow. In addition, this paper shows the vital role of the sax method in constructing an adjacency matrix through ablation experiments. Compared with the baseline methods, the prediction results show that SAX-STGCN performs better in different prediction durations. However, the performance of SAX-STGCN in long-term prediction is not as good as in short-term prediction, and it is challenging to predict sudden changes in traffic flow. Our future work will introduce more external features such as weather and significant events. We will use deep learning methods to embed both traffic flow timing data and external features and try to solve these problems.

REFERENCES
[1] X. Song, H. Kanasugi, and R. Shibasaki, “Deepttransport: Prediction and simulation of human mobility and transportation mode at a citywide level,” in Proc. 25th Int. Joint Conf. Artif. Intell., 2016, pp. 2618–2624.
[2] Z. Cao, S. Jiang, J. Zhang, and H. Guo, “A unified framework for vehicle rerouting and traffic light control to reduce traffic congestion,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 7, pp. 1958–1973, Jul. 2017.
[3] F. Rodrigues, I. Markou, and F. C. Pereira, “Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach,” Inf. Fusion, vol. 49, pp. 120–129, Sep. 2019.
[4] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, “T-GCN: A temporal graph convolutional network for traffic prediction,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 9, pp. 3848–3858, Sep. 2020.
[5] J. Ye, J. Zhao, K. Ye, and C. Xu, “How to build a graph-based deep learning architecture in traffic domain: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 5, pp. 3904–3924, May 2022.
[6] H. Yuan and G. Li, “A survey of traffic prediction: From spatio-temporal data to intelligent transportation,” Data Sci. Eng., vol. 6, no. 1, pp. 63–85, Mar. 2021.
[7] K. Lee, M. Eo, E. Jung, Y. Yoon, and W. Rhee, “Short-term traffic prediction with deep neural networks: A survey,” IEEE Access, vol. 9, pp. 54739–54756, 2021.
[8] J. Ye, J. Zhao, K. Ye, and C. Xu, “How to build a graph-based deep learning architecture in traffic domain: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 5, pp. 3904–3924, May 2022.
[9] B. L. Smith and M. J. Demetsky, “Traffic flow forecasting: Comparison of modeling approaches,” J. Transp. Eng., vol. 123, no. 4, pp. 261–266, Jul. 1997.
[10] N. L. Nihan and K. O. Holmesland, “Use of the box and Jenkins time series technique in traffic forecasting,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.
[11] B. Williams and L. Hoel, “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.
[12] M. Van Der Voort, M. Doughehtry, and S. Watson, “Combining Kohonen maps with ARIMA time series models to forecast traffic flow,” Transp. Res. C, Emerg. Technol., vol. 4, no. 5, pp. 307–318, 1996.
REFERENCES

[13] C.-H. Wu, J.-M. Ho, and D. T. Lee, “Travel-time prediction with support vector regression,” IEEE Trans. Intell. Transp. Syst., vol. 5, no. 4, pp. 276–281, Dec. 2004.

[14] X. L. Zhang, H. E. Guo-Guang, and L. U. Hu-Pu, “Short-term traffic flow forecasting based on K-nearest neighbors non-parametric regression,” J. Transp. Eng., vol. 24, no. 2, pp. 178–183, Feb. 2009.

[15] D. Park and L. R. Rilet, “Forecasting freeway link travel times with a multilayer feedforward neural network,” Comput. Aided Civil Infrastruct. Eng., vol. 14, no. 5, pp. 357–367, 1999.

[16] W. Huang, G. Song, and H. Hong, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5, pp. 2191–2201, Apr. 2014.

[17] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long short-term memory network for traffic speed prediction using remote microwave sensor data,” Transp. Res. C, Emerg. Technol., vol. 54, pp. 187–197, May 2015.

[18] R. Fu, Z. Zhang, and L. Li, “Using LSTM and GRU neural network methods for traffic flow prediction,” in Proc. 31st Youth Academic Ann. Conf. Chin. Assoc. Autom. (YAC), Nov. 2016, pp. 324–328.

[19] F. Zhao, G.-Q. Zeng, and K.-D. Lu, “EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization,” IEEE Trans. Veh. Technol., vol. 69, no. 1, pp. 101–113, Jan. 2020.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[22] J. Zhang, Y. Zheng, and D. Qi, “Deep spatio-temporal residual networks for citywide crowd flows prediction,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 1655–1661.

[23] M. Cao, V. O. K. Li, and V. W. S. Chan, “A CNN-LSTM model for traffic speed prediction,” in Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring), May 2020, pp. 1–5.

[24] J. Ke, H. Zheng, H. Yang, and X. M. Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” J. Transp. Res. C, Emerg. Technol., vol. 85, pp. 591–608, Dec. 2017.

[25] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, “Attention based spatial-temporal graph convolutional networks for traffic flow forecasting,” in Proc. AAAI, vol. 33, 2019, pp. 922–929.

[26] J. Zhu, Q. Wang, C. Tao, H. Deng, L. Zhao, and H. Li, “AST-GCN: Attribute-augmented spatiotemporal graph convolutional network for traffic forecasting,” IEEE Access, vol. 9, pp. 35973–35983, 2021.

[27] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, “Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 11, pp. 4883–4894, Nov. 2020.

[28] W. Chen, L. Chen, Y. Xie, W. Cao, and X. Feng, “Multi-range attentive Z. Cui, K. Henrickson, R. Ke, and Y. Wang, “Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting,” in Proc. IEEE Int. Conf. Pattern Recognit. (ICPR), Aug. 2018, pp. 1018–1023.

[29] C. Pan, J. Zhu, Z. Kong, H. Shi, and W. Yang, “DC-STGCN: Dual-channel based graph convolutional networks for network traffic forecasting,” Electronics, vol. 10, no. 9, p. 1014, Apr. 2021.

[30] T. W. Liao, “Clustering of time series data-a survey,” Pattern Recognit., vol. 38, no. 11, pp. 1857–1874, 2005.

[31] A. Abanda, U. Moro, and J. A. Lozano, “A review on distance based time series classification,” Data Mining Knowl. Discovery, vol. 33, no. 2, pp. 378–412, 2019.

[32] YIFEI SUO received the B.S. degree in traffic engineering from Chang’an University, Xi’an, China, in 2020. She is currently pursuing the M.S. degree in transportation engineering. Her research interests include network representation learning, spatio-temporal data mining, and big data mining.

[33] NA LI received the B.S. degree from the School of Civil Engineering, Xi’an University of Architecture and Technology, Xi’an, China, in 2021. She is currently pursuing the M.S. degree in transportation engineering. Her research interests include network representation learning, spatio-temporal data mining, big data mining, and machine learning.