IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 23, 2020, accepted January 2, 2021, date of publication January 6, 2021, date of current version January 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3049556

Modeling Global Spatial–Temporal Graph Attention Network for Traffic Prediction

BIN SUN[1,2], DUAN ZHAO[1,2], XINGUO SHI[3], AND YONGXIN HE[1,2]

[1]National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou 221008, China
[2]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, China
[3]Information Center, Shandong Energy Zibo Mining Group Company, Ltd., Zibo 255300, China

Corresponding author: Duan Zhao (zhaoduan1027@163.com)

This work was supported by the National Key Research and Development Program of China under Grant 2017YFC 0804402.

ABSTRACT Accurate and efficient traffic prediction is the key to the realization of intelligent transportation system (ITS), which helps to alleviate traffic congestion and reduce traffic accidents. Due to the complex dynamic spatial-temporal dependence between traffic networks, traffic prediction is extremely challenging. In previous studies, convolution neural network (CNN) and graph convolution network (GCN) were used to model spatial correlation. However, the non-Euclidean correlation of road network reduces the effect of convolution operator modeling. In addition, only considering the traffic interaction around the concerned points simplifies the influence of traffic network. In order to address the above problems, this article proposes an end-to-end global spatial-temporal graph attention network (GST-GAT), which uses the ''global interaction + node query'' to model the dynamic spatial-temporal correlation of traffic. In the encoder, the long short-term memory (LSTM) component flexibly transforms the traffic dynamic spatial-temporal graph into feedforward differentiable features. Global traffic interaction is proposed to summarize traffic network context changes and integrate all node features at each moment through a forward calculation. Then, each node computes the influence of traffic global interaction on a single node in parallel, and the spatial-temporal interaction information is adaptive fused by gating fusion mechanism. Finally, the end-to-end network structure is used to train the rich mixed feature coding to generate the traffic prediction status of each node. Experiments on public transportation data sets show that GST-GAT performs better than previous work in terms of accuracy and inference speed.

INDEX TERMS Intelligent transportation system, global spatial-temporal graph, traffic prediction.

I. INTRODUCTION

With the development of urbanization in recent years, ''urban diseases'' such as traffic congestion and parking difficulty have become the primary problems faced by urban traffic management. In the past, the most direct solution was to increase infrastructure and transportation supply. However, transportation supply is not unlimited and construction is expensive. Refined urban management is the most effective measure to alleviate current urban traffic congestion. In order to alleviate traffic congestion and reduce traffic accidents, the construction of intelligent transportation system has become a future trend [1]. The intelligent transportation system aims to establish a safe, coordinated and smart transportation network [2]. Cyber-physical system (CPS) is the cornerstone of intelligent transportation system construction, it integrates technologies such as intelligent perception, deep computing, reliable communication and precise control to connect the virtual digital world with the physical world [3]. Accurate traffic prediction is the key to the realization of intelligent transportation system, which helps the government or company to make better management decisions. For example, taxi demand forecast [4] helps operating companies (such as didi and Uber) optimize vehicle scheduling routes, spatial population flow forecast [5] helps public security management, subway network traffic forecast [6] helps operation and maintenance departments optimize scheduling schemes, and road traffic forecast [7] helps transportation departments better control and manage traffic conditions. Nowadays, with the deployment of sensor networks in the transportation sector, rich spatial-temporal traffic data can be recorded (such as traffic flow or speed) [8]. How to mine the space-time rules of big data to make traffic prediction more accurately has aroused great interest of researchers.

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai.

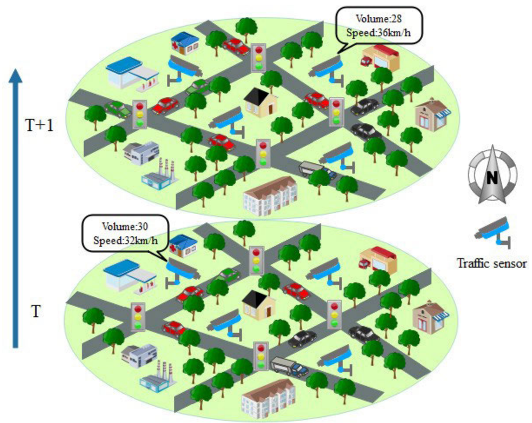


FIGURE 1. Traffic detection sensor on the road network.

Traffic prediction is a typical spatial-temporal data prediction problem, which presents the characteristics of spatial static and temporal dynamic. As shown in Fig. 1, many advanced sensor technologies (such as induction coils, GPS, cameras, bluetooth and RFID devices) have been used to collect traffic information, and the fixed sensor nodes on the traffic network generate a series of data updated periodically. However, due to the complex and dynamic spatial-temporal dependence between different sensor nodes on the traffic road network, traffic prediction is a very challenging task. On the one hand, the topology of urban traffic network affects the traffic condition of each monitoring node. There is a wide range of influences between nodes. The upstream road affects the downstream road through the transmission effect, and the downstream road affects the upstream road through the feedback effect [9]. Traffic accidents, road maintenance and other special circumstances will have a significant local impact on the condition of the entire traffic network [10]. The complex topology and different context information of road network also bring challenges to traffic prediction. On the other hand, the nonlinear dynamic correlation also affects the traffic condition of each node. Traffic conditions may fluctuate sharply and suddenly due to some emergencies, thus affecting the correlation between different time steps. For example, the congestion in the first ten minutes will affect the current traffic situation [5]. In addition, since the running state of vehicles is controlled by humans, human movement behavior has multimodality and randomness [11], [12]. When at a crossroads, there are multiple possibilities for the vehicle's driving route, which will affect the traffic conditions of neighboring nodes. Therefore, the randomness of the traffic state also needs to be considered.

In recent years, researchers have carried out a lot of work on traffic prediction. The traditional statistical method classifies the traffic prediction problem as a time series prediction problem, and captures the regularity of the past time series data to predict the future traffic situation. Autoregressive integrated moving average (ARIMA) [13] and its variants [14] are classic methods. However, this kind of linear model requires high stability of time series data, and it is difficult to deal with nonlinear traffic data flow. In order to solve the above problems, the machine learning method with the ability to capture complex nonlinear relations is applied to traffic prediction. Support vector regression (SVR) [15], Hidden Markov Model (HMM) [16] and BP neural network [17] are used for short-term traffic prediction. Although these methods improve the prediction effect to some extent, they need to rely on artificial design of feature extraction rules. The data-driven deep learning method has achieved great success in many challenging learning tasks [18], which makes the deep learning technology be applied to solve the traffic prediction problem. Deep belief networks (DBN) [19] and stacked auto-encoders [20] can effectively learn high-dimensional nonlinear features. Recurrent neural networks (RNN) and their variants also show great advantages in solving traffic prediction problems [21]–[23]. However, these methods focus on the modeling of single-node traffic time series data, but ignore the impact of spatial dependence on the traffic conditions of each node.

To extract the spatial feature of traffic data, researchers introduce CNN into the traffic prediction tasks. Ma *et al.* proposed an image-based method that treats the traffic networks as images and use CNN to learn the spatial features [24]. On this basis, some researches [25]–[27] have tried to combine CNN with RNN and its variants for traffic prediction and achieved good results. However, the traditional CNN is mainly good at extracting feature information in images or Euclidean space that can be converted into a two-dimensional matrix. The traffic network presents a non-Euclidean correlation, which reduces the efficiency of the convolution operation [28]. Compared with CNN, GCN [29], [30] extends convolution operator to more general graph structure data. A new research trend is to abstract traffic timing information into graph structure information for processing [31]. Diffusion convolutional recurrent neural network (DCRNN) [32], Attention based spatial-temporal graph convolutional network (ASTGCN) [33], and multi-range attentive bicomponent GCN (MRA-BGCN) [34] are all used for modeling the dynamic spatial-temporal relationship in the traffic road network. In these models, the Laplacian matrix of the graph is defined as strictly invariant in GCN, namely, the adjacency matrix of the input graph is constant. However, there are great differences in traffic patterns in different time periods, and sudden traffic accidents will also affect the relationship between road networks [35]. At present, most of the prediction methods only consider the static spatial dependence of the traffic network, so it is difficult to expand them to predict the traffic situation of the whole urban road network. Because the dependence of traffic conditions between road networks is actually dynamic [36]. In addition, the ''black box'' nature of the deep learning model also makes the traffic prediction problem lack of interpretability [37].

In order to solve the challenges of traffic prediction, we propose a global spatial-temporal graph attention network (GST-GAT) to predict the traffic situation on the topological graph of the traffic network in the future. The traffic conditions on the traffic network topology at each time can form a static ''map'' (such as traffic flow heat map). With the change of time, these static maps become a dynamic map with temporal information. It is worth noting that the interaction influence of traffic conditions is not limited to the connected roads, and considering the global interaction continuous influence of traffic network is more in line with the actual situation [38].

We regard each monitoring point on the traffic topology as a node and model the spatiotemporal representation of the road network. In the spatiotemporal diagram of dynamic road network, the influence of global traffic interaction on a single node, the time characteristic dependence of the node itself and the randomness of traffic conditions are all considered as influence factors. Structured RNN [39] is considered to be a universal tool that can transform any spatial-temporal graph into a rich, scalable, and jointly trained feed-forward mixture. The GST-GAT proposed in this article is a variant of structured RNN, an encoder and decoder architecture based on novel attention mechanism, and adopts the mode of ''global interaction + node query'' to model the traffic dynamic spatial-temporal correlation. Firstly, the dynamic spatial-temporal map of traffic is flexibly transformed into feedforward and differentiable feature coding by using LSTM [40] component. Different from the graph attention network (GAT) [41] used in reference [42], [43] to calculate the interaction effect in pairs to extract spatial dependence, the traffic global interaction is proposed to summarize the contextual changes of the traffic network, and to integrate all node features through a forward calculation at each time. Then, each node computes the influence of traffic global interaction on a single node in parallel, and the spatial-temporal interaction information is adaptive fused by gating fusion mechanism. Finally, the end-to-end encoder decoder structure is used to train the rich mixed feature coding to generate the traffic prediction status of each node. The main contributions of this article are as follows:

1) In this article, GST-GAT is proposed for traffic prediction. The network is flexible and scalable, and can capture the complex space-time dependence in the dynamic traffic topology.
2) When modeling the spatial interaction of traffic nodes, this article adopts the method of ''global interaction + node query''. Compared with the pairwise interactive computing mode, the proposed method can integrate the global traffic interaction characteristics with only one forward operation, which greatly reduces the computational cost.
3) We conduct extensive experiments on two real-world traffic datasets to evaluate the GST-GAT. Experimental results show that GST-GAT has better prediction effect and faster calculation efficiency.

The rest of this article is organized as follows. First, we analyze the work related to traffic prediction. Next, we specify the principle of GST-GAT traffic prediction model. Then, we discuss the evaluation results of the GST-GAT traffic prediction model on public data sets. Finally, we summarize the work of this article.

II. RELATED WORK

In this section, we will gradually review the work progress of traffic prediction research, the application of structured RNN on spatial-temporal graph and deep learning with attention mechanism. The following work helps readers to understand what is traffic prediction problem, why structured RNN is used as the basis to design traffic prediction model, how to improve the performance of deep learning network with attention mechanism, and the defects of existing models.

A. TRAFFIC PREDICTION

Traffic prediction is the foundation and key of intelligent transportation system. The early time series prediction model [13], [14] is only applicable to the relatively stable and smooth traffic data, traffic prediction. Although the subsequently developed machine learning methods [15]–[17] can deal with nonlinear data, the prediction effect of such methods largely depends on artificially designed feature engineering, which often requires expert experience and a lot of experiment. Moreover, they do not perform well in long-term forecasting, especially when in the face of complex road network relationships. Compared with machine learning, deep learning can more effectively extract and integrate features to deal with some more complex problems.

In recent years, many deep learning methods have been applied to solve traffic prediction problems. Due to the great success achieved in processing time series data, RNN and its variants [21]–[23] are used for traffic prediction. However, these methods only focus on the time-series changes of the traffic situation of a single node, ignoring the interaction between traffic networks. In order to model the spatial impact between traffic networks, CNN [24]–[27] was creatively used in traffic prediction. However, the non-Euclidean correlation of road network space reduces the modeling effect of convolution operator. Compared with CNN, Graph Convolutional Network (GCN) [29], [30], [33], [34], [44], [45] extends the convolution operator to more general graph structure data and is more suitable for handling traffic prediction problems. In these models, the Laplacian matrix is defined as strictly unchanged in GCN. However, the traffic situation on the road network topology is actually a dynamic graph with time series information, which affects the prediction effect of the model.

For traffic prediction, except for the highly correlated traffic conditions between two sections which are geographically close to each other. As traffic conditions will spread between road networks, there is also a correlation between the traffic conditions of the two sections away from each other [46]. In addition, the future traffic condition of a road segment depends not only on the previous traffic condition, but also the influence of adjacent roads and the uncertainty facing the future should be considered. The core of the traffic prediction problem is how to effectively model the complex dynamic spatial-temporal relationship between traffic road networks. Structured RNN is regarded as a method that can successfully combine the functions of advanced spatial-temporal graph and sequence learning of RNN, and has been successfully applied to solve problems such as human motion modeling, trajectory prediction, and driving motivation prediction [39]. Traffic prediction is a typical spatial-temporal data prediction problem, which is very suitable for modeling with structured RNN after transforming into dynamic spatial-temporal graph.

B. STRUCTURAL RECURRENT NEURAL NETWORK ON SPATIAL-TEMPORAL GRAPHS

For spatial-temporal data problems, spatial-temporal graph structure has been used as a basic modeling idea. Because the spatial-temporal graph naturally represents the spatial and temporal information of the modeling subject, and acts as the general middleware representation, which facilitates the abstract extraction of element features during modeling. In addition, since multiple subjects in the space-time graph may use the same basic model, this brings benefits such as super-linear parameter scaling and parameter sharing [47]. The emergence of structured RNN, a pioneering deep learning method, makes people no longer limited to using methods like Markov chain [48] to process graph structure data. Because the method based on deep learning has better generalization ability and faster processing speed.

When using structured RNN and its variants [39], [47], [49], [50] to process dynamic spatial-temporal data, these models convert the physical world spatial-temporal problems into dynamic spatial-temporal graphs for processing. In spatial-temporal graph, multiple subjects are regarded as nodes, which are connected by the spatial-temporal edge. RNN components are used to model these spatial-temporal edge relationships, and the probability graph model is implemented through the structured RNN architecture to predict the future status of nodes. Inspired by this, Kim *et al.* [51] convert the road network topology into a spatial-temporal graph, and use structured RNN to model the interaction and time dependence between adjacent road segments. Experiments show that structured RNN has good scalability, and the prediction effect on traffic data sets is higher than that of convolutional neural networks [51]. However, the above model only considers the equal influence of various factors, and does not consider the relative importance of interactive information. In the human trajectory prediction problem similar to traffic prediction, Vemula *et al.* [49] used Social Attention to capture the relative importance of each person in crowd navigation. The application of attention mechanism in structured RNN improves the accuracy of model prediction and the interpretability of deep learning models [42].

C. DEEP LEARNING WITH ATTENTION MECHANISM

The structured RNN deep learning framework is essentially an encoder - decoder network. With the increase of input length, information loss occurs, and the prediction performance of the model will decline rapidly [52]. The attention mechanism significantly improves the accuracy and interpretability of the model by assigning attention weights related to its importance to each input element, and has been widely applied in various fields [53]. Recently, researchers began to apply the attention mechanism to traffic prediction. Liang *et al.* [38] proposed a general sensor time series data prediction framework, namely multistage attention neural network, to predict future data of geographic sensors (such as air quality, water quality, and traffic conditions). Both the spatial-temporal attentional neural network (STANN) [53] and the attention-based periodic temporal neural network (APTN) [34] have been used in traffic prediction. These models capture the spatial dependence and time step dependence between road networks respectively through the attention mechanism. Furthermore, the latest work combines attention mechanisms with flexible graph-structure networks for traffic prediction. Graph multi-attention network (GMAN) [42] and spatial-temporal graph attention network (ST-GAT) [43] both use graph attention mechanism to capture the spatial-temporal dependence between road networks. However, these models only model the pairwise interaction between road network nodes, which oversimplifies the influence mode between traffic networks. The pooling operation in the model is used to summarize the interaction effects of the surrounding nodes. As the number of nodes increases, the amount of calculation increases squarely, so it is more time-consuming in model deployment [54].

After sorting out the relevant work of traffic prediction, we find that the dynamic temporal and spatial correlation of traffic network has been generally considered. The modeling method of dynamic spatial-temporal graph and graph attention mechanism are also applied in the latest work. Fig. 2 shows the attention network paradigm of the classic traffic prediction spatial-temporal graph. The physical structure of the road network is abstracted as a dynamic spatial-temporal graph for modeling. The attention mechanism and structured RNN are used to extract the spatial-temporal effects of surrounding nodes. However, the existing work uses pair-wise calculations when modeling the interaction between nodes, which simplifies the impact between traffic networks and makes calculations cumbersome. In this article, the pooling scope is extended from the local scope to all the global nodes, which is more in line with the actual traffic dynamic interaction. Next, we will introduce the GST-GAT model principle in detail.

III. PROBLEM REPRESENTATION AND MODEL PRINCIPLES

First, we define the traffic prediction problem. Then, the global spatial-temporal graph structure will be introduced. Finally, we explain in detail the working principle of the GST-GAT encoder-decoder deep learning framework.

A. PROBLEM DEFINITION

We assume that there are N sensor nodes on the traffic road network, and each sensor uploads the traffic monitoring status at consecutive intervals. We use $v_i^t \epsilon \mathbf{R}^{1\times C}$ to denote the traffic condition of node i at time t, where C is the number of related traffic conditions (such as traffic flow, traffic speed). In the traffic prediction problem, the observable time step is T_{obs},

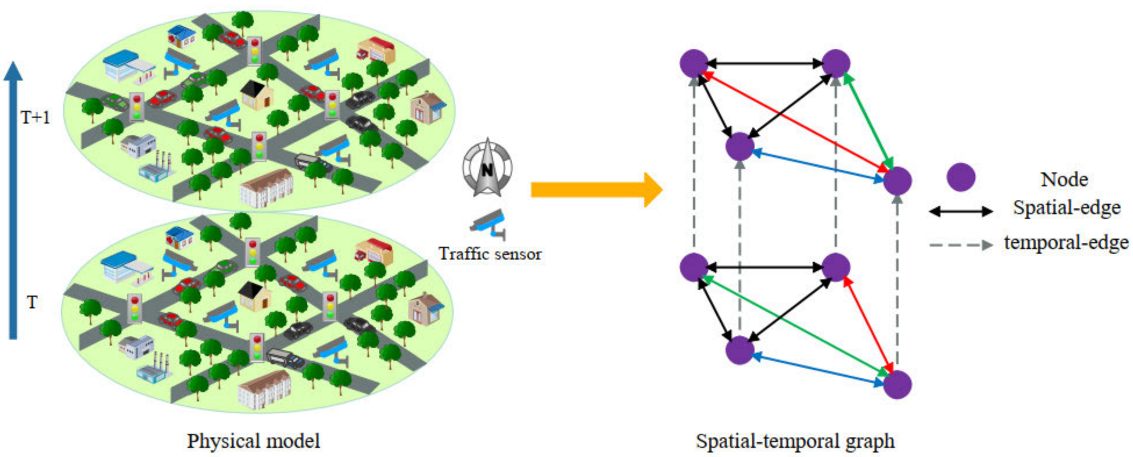


FIGURE 2. The attention network paradigm of the classic traffic prediction spatial-temporal graph.

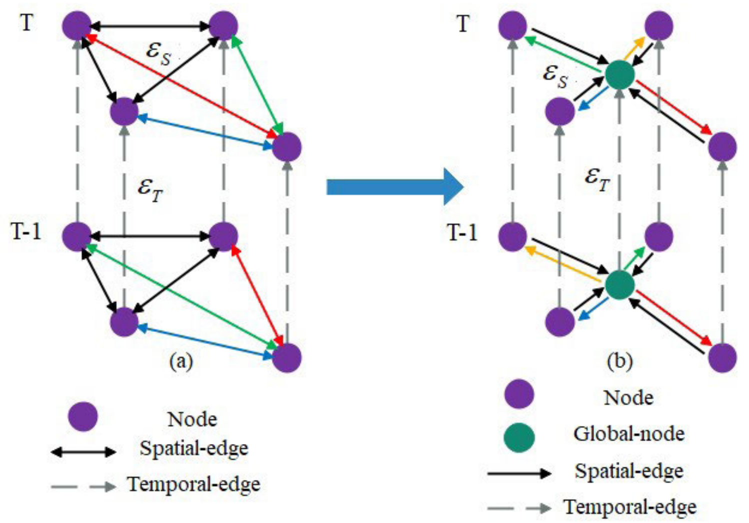


FIGURE 3. Classic spatial-temporal graph and global spatial-temporal graph.

and the predicted time step is T_{pred}. The observable traffic time series data of the traffic network at time t is defined as $O_t = \{v_i^t \mid i = 1, 2 \ldots N\} \epsilon \mathbf{R}^{N \times C}$. Similarly, the future traffic time series data of the traffic road network is defined as $\hat{F} = \{v_i^t \mid i = 1, 2 \ldots N\} \epsilon \mathbf{R}^{N \times C}$.

Traffic prediction is defined as a time series data prediction problem. Given the observable time series data $O = (O_1, O_2 \ldots O_{T_{obs}}) \epsilon \mathbf{R}^{N \times T_{obs} \times C}$, we aim to predict the future traffic situation $\hat{F} = (\hat{F}_{T_{obs}+1}, \hat{F}_{T_{obs}+2} \ldots \hat{F}_{T_{obs}+T_{pred}}) \epsilon \mathbf{R}^{N \times T_{pred} \times C}$,where N is the number of traffic nodes.

B. GLOBAL SPATIAL-TEMPORAL GRAPH STRUCTURE

As a popular modeling representation of dynamic spatial-temporal interaction, spatial-temporal graph has been widely used. In literature [39], [47], [49]–[51], researchers have used the classical spatial-temporal graph described in Fig. 3(a) to represent the spatial-temporal interaction between nodes. In a classic spatial-temporal graph $G_1 = (v, \varepsilon_S, \varepsilon_T)$, v is a set of traffic nodes, ε_S is a set of spatial edges, and ε_T is a set of temporal edges. Specifically, the spatial edges are limited to adjacent nodes, and the colored connecting lines represent different influence weights between nodes. Classical spatial-temporal graphs are applied to many structured data tasks by modeling the graph neural network structure of hidden relationships and interactions around nodes. In order to better infer hidden relationships and interactions between nodes to achieve link prediction, some variants of classic spatial-temporal graphs are proposed. For example, dynamic spatial-temporal graphs [47], [54] were proposed for pedestrian trajectory prediction, and the interaction between pedestrians was extended to the global situation. Haddad *et al.* [50] used the instance layer to learn the movement and interaction of instance nodes, and the category layer to learn the similarity of instance nodes belonging to the same type to improve traffic trajectory prediction. Obstacle nodes are added to the crowd space-time graph to model the interaction between people and the environment [55].

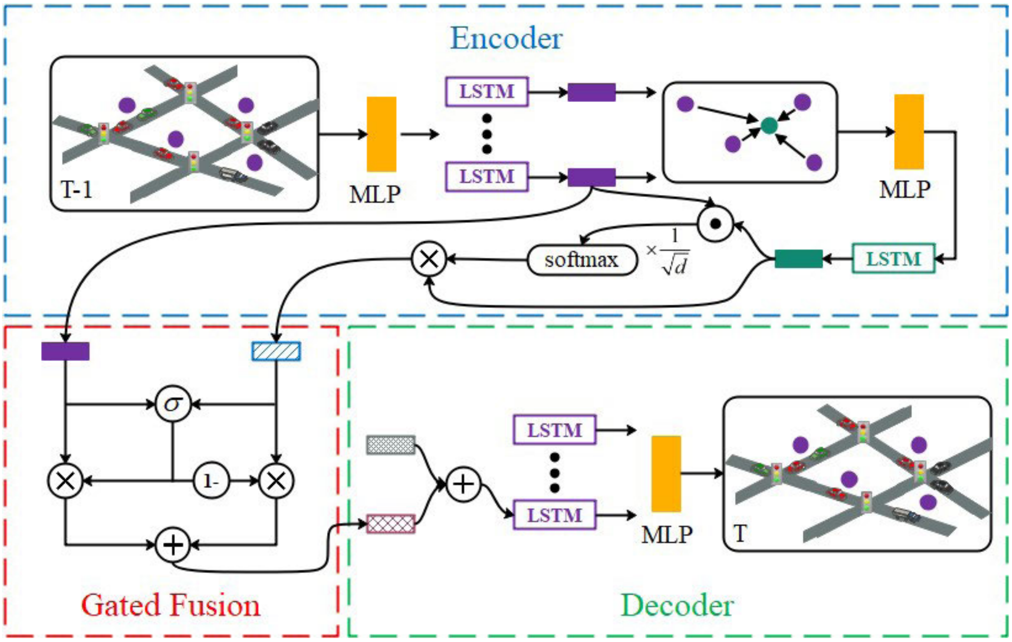


FIGURE 4. Global spatial-temporal graph attention network.

Inspired by the above research, in order to better capture the dynamic interaction between road networks and optimize the calculation efficiency, we improved the structure of the classic spatial-temporal graph. The paired computing mode between nodes was abandoned, which caused a lot of repeated operations. The global spatial-temporal graph structure of "global interaction + node query" is proposed. Fig. 3(b) shows the abstract representation of the global spatiotemporal graph $G_2 = (v, \varepsilon_S, \varepsilon_T, g)$. Specifically, v is a traffic node set, ε_S is a set of spatial edges, ε_T is a set of temporal edges, and g is a newly added global interactive node. In this article, the vertices are the sensor nodes all over the traffic network, and the size of the vertices can be flexibly changed according to the actual situation. Spatial edges are divided into two types to model spatial influence, one connects all vertices to global interaction nodes, and the other feeds back the relative influence of global interaction nodes to each vertex. At the same time, the temporal edge is connected to the same vertex with adjacent time steps to construct the time dependence of the vertex itself. Based on the representation of global spatial-temporal graph, we propose a global spatial-temporal graph attention network to model the complex dynamic spatial-temporal dependence between traffic networks.

C. GLOBAL SPATIAL TEMPORAL GRAPH ATTENTION NETWORK

In urban traffic network, the traffic data collected by distributed sensor nodes can form a dynamic graph with timing information. In previous studies, the spatial interaction between adjacent sections and the temporal dependence of traffic conditions have been widely considered. However, the interaction effect of traffic conditions is not limited to the connected roads. It is more realistic to consider the overall interaction effect of traffic network. We regard each monitoring point on the traffic topology as a node and model the global spatial-temporal graph representation of the traffic. In the global spatiotemporal graph, the influence of traffic global interaction on a single node, the time dependence of the node itself and the randomness of traffic conditions are all considered as influence factors. In particular, we propose a "global interaction + node query" graph neural network modeling method to calculate the status update of each node in parallel. The overall frame structure of the global spatial-temporal graph attention network is shown in Fig. 4. The graph neural network prediction model adopts encoder and decoder structure, which is mainly divided into feature encoder module, gate fusion module and decoder module.

In the feature encoder module, first, we use the global spatial-temporal graph structure to model the spatial-temporal interaction of the traffic road network, and calculate the feature codes of each node on the traffic road network in parallel through the LSTM component. Then, the maximum pooling operation is used to embed the feature information of all nodes into the global interactive node. At this time, the global interactive node contains rich traffic context information. Since the traffic condition is affected by the surrounding and distant road conditions, we use scaled dot product [56] to query the specific interaction effects of global interactive nodes on a single node. Through the feature extraction method of "global interaction + node query", it is possible to calculate the global interaction and reduce the calculation cost.

In the gate fusion module, spatial-temporal feature coding is input into sigmoid activation function to generate influence coefficient. Then, the spatiotemporal feature codes that affect the traffic conditions of the nodes are fused and calculated according to the different weights. In particular, in order to enhance the generalization ability of the model, we add random noise to simulate the randomness of traffic conditions.

In the decoder module, through forward propagation, we integrate node feature coding, interactive feature coding and random noise as the input of the decoder. Based on the rich mixed feature coding and back propagation mechanism, the parameters in the neural network are continuously optimized and updated to get a more accurate model. In the GST-GAT model, the parameters in the LSTM component are Shared, which enables the training network to flexibly adapt to different number of node inputs without increasing the number of parameters. In the same spatiotemporal structure, parameter sharing can reduce model complexity and improve model generalization ability. Next, we will introduce the specific principle of GST-GAT model in detail.

1) TRAFFIC NODE

In order to transform the global spatial-temporal graph into a graph neural network model for calculation. First, we use multi-layer perceptron (MLP) to embed the traffic condition attributes of each traffic node to obtain a fixed-length vector. We embed the traffic node v_i^t into a vector e_i^t, which is used as the input to LSTM to get the traffic node coding h_{ei}^t (i.e. the purple square in Fig. 4).

$$e_i^t = \phi(v_i^t; W_{tem}^e) \tag{1}$$

$$h_{ei}^t = LSTM(h_{ei}^{t-1}, e_i^t; W_{tem}^T) \tag{2}$$

where $\phi(\cdot)$ is a nonlinear embedding function, W_{tem}^e is embedding weights. h_{ei}^t is the hidden layer and output of the LSTM component at time t, and W_{tem}^T is the cell weight of the LSTM at the temporal edge. In particular, the weights W_{tem}^e and W_{tem}^T are shared parameters of all traffic nodes.

2) GLOBAL INTERACTION NODE

In the literature [34], [42], [43], [51], [53], researchers capture the paired interaction between traffic nodes based on the classical spatial-temporal graph model, only considering the interaction between the surrounding limited nodes on the central node, which has a lot of repeated calculation and fails to consider the global context information of traffic. In order to improve the performance of existing models, StarNet [57] is proposed to efficiently aggregate global interaction information in the human trajectory prediction problem. This mechanism can calculate global interaction and compress computational overhead. Because the traffic prediction problem is similar to the human trajectory prediction problem. In GST-GAT, we extract the feature coding of all traffic nodes through parallel forward calculation, and embed the feature information of these nodes into the global interactive node through the max-pooling operation, which is the key index to describe the global traffic context.

Drawing lessons from the method used by researchers to construct a feed-forward neural network through structured LSTM in StarNet [57], we first use the max-pooling component to integrate the traffic condition attribute vectors after all traffic nodes are embedded to obtain the traffic global interaction g^t. Then, we embed the traffic global interaction g^t into a vector e_g^t, which is used as the input to LSTM to get the traffic global feature coding h_g^t (i.e. the green square in Fig. 4).

$$g^t = MaxPooling(e_1^t, e_2^t \dots e_i^t) \tag{3}$$

$$e_g^t = \phi(g^t; W_g^e) \tag{4}$$

$$h_g^t = LSTM(h_g^{t-1}, e_g^t; W_{spa}^g) \tag{5}$$

The operation of max-pooling can reduce the parameters and calculation amount while retaining the main features, which helps to improve the model generalization ability and prevent overfitting. W_g^e is the embedding weight, and W_{spa}^g is the cell weight of the spatial edge LSTM.

3) NODE QUERY

According to the definition in the global spatial-temporal graph above, the traffic context features contained in the global interaction node are shared among all traffic nodes. However, the global interaction nodes have different effects on different traffic nodes. We use the method of "global interaction + node query" to obtain the specific impact of global interaction nodes on a single traffic node. We use scaled dot product attention to assign a global influence weight $w(h_{g\to i}^t)$ to each traffic node, and then multiply the global influence weight $w(h_{g\to i}^t)$ with the traffic global feature coding h_g^t to obtain the interactive feature coding $S_{g\to i}^t$ (i.e. the blue square with stripes in Fig. 4).

$$w(h_{g\to i}^t = softmax(\frac{1}{\sqrt{d}} Dot(W_1 h_g^t, W_2 h_{ei}^t)) \tag{6}$$

$$S_{g\to i}^t = w(h_{g\to i}^t) \otimes h_g^t \tag{7}$$

where W_1 and W_2 are weights to linearly scale and project the hidden states into d_e dimensional vectors. $Dot(\cdot)$ is the dot product [55], $softmax(\cdot)$ is the activation function, and $\frac{1}{\sqrt{d}}$ is the scale factor.

So far, on the basis of modeling traffic problems with global spatial-temporal map, we use structure LSTM components to extract the spatial-temporal interaction features between traffic nodes, and use the idea of "global interaction + node query" to capture the specific interaction effects of global interaction nodes on each traffic node. Through the above operations, we extract the spatial-temporal effects in the dynamic road network, namely the traffic node coding h_i^t and interactive feature coding $S_{g\to i}^t$.

4) GATED FUSION

The traffic condition of a single node in the road network is influenced by its previous state and the interaction of other nodes. In the feature encoder module, we have extracted the spatial-temporal correlation features that affect the traffic conditions of the road section, namely the traffic node coding h_i^t and interactive feature coding $S_{g\to i}^t$. It is a common practice to directly connect spatial-temporal feature coding as the input of decoder [34], [38], [53]. However, how to make the neural network adaptively distribute the influence weight of spatial-temporal features is very important to improve the prediction accuracy. Inspired by the structure of the LSTM, the gate fusion mechanism is used to adjust the state of the information flow vector in the neural network. The spatial-temporal feature vector is input into the activation function, and the sigmoid layer outputs a number between 0 anda 1 to describe the passing degree of spatial-temporal information flow. The output value of the activation function is 0, which means ''no information is allowed to pass'', while the output value of 1 means ''all information is allowed to pass''.

$$z = \sigma(W_T h_i^t + W_S S_{g\rightarrow i}^t + b) \tag{8}$$

$$H_i^t = z \otimes h_{ei}^t + (1 - z) \otimes S_{g\rightarrow i}^t \tag{9}$$

where $\sigma(\cdot)$ is the sigmoid activation function, W_T and W_S are learnable parameters, and z is the gate. H_i^t is the fusion spatial-temporal feature coding (i.e. the pink grid square in Fig. 4). The gate fusion mechanism adaptively controls the spatial-temporal information flow of traffic nodes.

5) DECODER AND MODEL TRAINING

According to the previous analysis, similar to the problem of crowd trajectory prediction, the randomness of traffic conditions is also worth considering. We splice the fusion spatial-temporal feature coding H_i^t and random noise coding z together and input it into the decoder LSTM component to obtain the feature vector h_{di}^t. Then, the MLP converts this feature vector into the predicted traffic condition.

$$h_{di}^t = LSTM(h_{di}^{t-1}, concat[H_i^t, z]; W_d) \tag{10}$$

$$\hat{F}_i^{t+1} = \hat{v}_i^{t+1} = MLP(h_{di}^t; W_{de}) \tag{11}$$

where z is the noise vector satisfying the standard normal distribution, $MLP(\cdot)$ is the multi-layer perceptron, W_d and W_{de} are the embedded weights.

Since the structured neural network is smooth and differentiable, after the above forward propagation computation, we train the model through back propagation. During the training phase, we use the Adam optimizer to train GST-GAT by minimizing the mean absolute error (MAE) between the predicted and true values.

$$L(\theta) = \frac{1}{T_{pred}} \sum\nolimits_{t=T_{obs}+1}^{T_{obs}+T_{pred}} |\hat{F}_t - F_t| \tag{12}$$

where θ represents all learnable parameters in GST-GAT, and $\hat{F}_t$ and F_t respectively represent the predicted and true values of all traffic nodes on the road network at time t.

IV. EXPERIMENTS

In this section, we introduce the public data set used in the experiment, various baseline methods, evaluation indicators and model deployment details. In order to evaluate the performance of the proposed model, we did a lot of experiments to compare the performance of GST-GAT and other baseline methods in accuracy and inference speed. In addition, in order to better study the performance of the structured graph neural network, we also set up some ablation experiments.

TABLE 1. Overview of the experimental data set.

Data	METR-LA	PEMS-BAY
Nodes	207	325
Edges	1515	2369
Time Steps	34272	52116
Time Windows	5min	5min

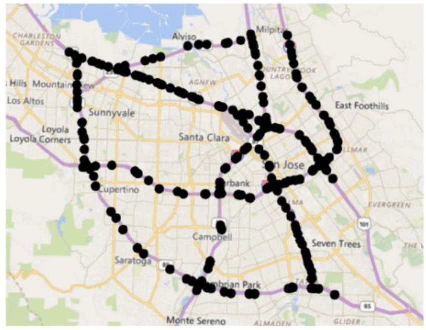


FIGURE 5. Sensor distribution of PEMS-BAY dataset.

A. DATASETS

We evaluate the performance of GST-GAT on two public traffic datasets of different road network scales: METR-LA and PEMS-BAY [32]. The attributes of the two data sets are shown in Table 1.

METR-LA: Traffic data are collected from observation sensors in the highway of Los Angeles County. In the experiment, we select 207 sensors and 4 months of data, with a date range from March 1, 2012 to June 30, 2012.

PEMS-BAY: Traffic data are collected by California Transportation Agencies Performance Measurement System (PeMS). In the experiment, we select 325 sensors and 6 months of data, with the date range from January 1, 2017 to May 31, 2017. The distribution of road network sensors is shown in Fig. 5.

In order to construct a global spatial-temporal graph of traffic, we treat each sensor as a traffic node, and the traffic conditions of all nodes are used as input and output of GST-GAT. The above two traffic data sets cover a wide range of space, and the data indicators are relatively comprehensive. The data missing rate is low and relatively stable. We have done abnormal data repair and data standardization in the data preprocessing part. Furthermore, 70% of the data is used for training, 10% for verification and the remaining 20% for testing.

The repair of abnormal traffic flow data mainly addresses the problems of missing data and data errors. In the data set, data with obvious errors can be deleted directly. For the missing historical data, the historical trend method is used to repair, and the method is to fill in the weighted value of the flow data $q(t-1)$ at the previous time and the flow data $q(t)$ at the current time. The abnormal data repair formula is as follows, where α is the smoothing coefficient.

$$q(t) = \alpha q(t) + (1 - \alpha)q(t - 1) \tag{13}$$

In order to avoid some unnecessary numerical calculation problems, and to have a better convergence speed when using the gradient descent method to solve. According to the practice in reference [28], [32], [38], we use Z-score to normalize the traffic data.

$$X = \frac{x - \mu}{\sigma} \tag{14}$$

In the formula, μ and σ are the sample mean and standard deviation, respectively.

B. BASELINES

We compare the GST-GAT model proposed in this article with the following 7 baselines: Auto-Regressive Integrated Moving Average (ARIMA) [13], Support vector regression (SVR) [15], Long Short-Term Memory Network (LSTM) [21], Spatial-Temporal Graph Convolution Network (STGCN) [31], Attention Based Spatial-Temporal Graph Convolutional Networks(ASTGCN) [33], Spatial-temporal attentive neural network (STANN) [53] and Graph Multi-Attention Network(GMAN) [42].

ARIMA: The classic time series forecasting method uses the correlation between time series data to predict traffic. **Factor**: Single node timing.

SVR: The regression prediction method based on machine learning can adapt to non-linear traffic data. **Factor**: Single node timing.

LSTM: The classic deep learning time series prediction method can effectively solve the problem of gradient disappearance and gradient explosion in the long-term dependence of traditional RNN processing time series, which is suitable for traffic prediction. **Factor**: single node timing.

STGCN: The traffic prediction is defined as a problem on the spatial-temporal graph, and the graph convolution structure is used to build a model to extract the spatial-temporal relationship between all links. **Factor**: spatial-temporal interaction.

ASTGCN: On the basis of STGCN, graph convolution and attention mechanism are used to model the traffic data of graph network structure. **Factor**: Spatial-temporal interaction + spatial-temporal attention.

STANN: Multiple LSTM components are used to extract spatial-temporal correlations from historical traffic sequences. Spatial attention is used to capture the relative influence of other nodes, and temporal attention is used to capture its own time-dependent influence. **Factor**: spatial-temporal interaction + spatial-temporal attention.

GMAN: Based on the traffic spatial-temporal graph, the model adopts an encoder-decoder architecture. Both the encoder and the decoder are composed of multiple spatial-temporal attention modules, and the gate fusion mechanism is used to merge the influence of spatial-temporal factors on traffic conditions. **Factor**: spatial-temporal interaction + spatial-temporal attention.

C. EVALUATION METRICS

In order to evaluate the performance of traffic prediction, we used three standard indicators [42], including mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE), which are defined as follows:

$$MAE = \frac{1}{N}\sum\nolimits_{i=1}^{N} \mid \hat{v}_i^t - v_i \mid \tag{15}$$

$$RMAE = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N} (\hat{v}_i^t - v_i)^2} \tag{16}$$

$$MAPE = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{\mid \hat{v}_i^t - v_i \mid}{v_i} \times 100\% \tag{17}$$

where N is the size of the testing set, v_i and $\hat{v}_i$ are the ground truth and the predicted value, respectively. For these three metrics, smaller values indicate better performance.

D. IMPLEMENTATION DETAILS

When modeling the global spatiotemporal graph of traffic, our focus is on the types of nodes and edges. Specifically, nodes are divided into traffic nodes and global interaction nodes, and edge types are divided into temporal edges and spatial edges. In this article, we make no assumptions about adjacency structures. The design of ''global interaction + node query'' makes the spatial-temporal feature information flow spread efficiently in the graph neural network. The same type of LSTM components share weight parameters, which can achieve more effective parameter scaling and improve the calculation efficiency. In the embedding layer, we convert the input into a 32-dimensional vector with a packet loss rate of 0.5. The hidden state dimension of all LSTM components is set to 64, and the feature representation dimension is also set to 64. In training, the batch size is set to 128, the learning rate is set to 0.001, and the Adam optimizer [56] is used to train the model. We configure a Linux server and the other configurations to run the experiment as follows: 2 Intel(R) CPU i9-10900KF @ 3.70GHZ; 32GB RAM; 1 NVIDIA RTX3090 GPU.

The deep learning model has many training parameters, and the phenomenon of model overfitting may occur. The specific performance of over-fitting: the model has a small loss function on the training data, and the prediction accuracy is high; but on the test data, the loss function is relatively large, and the prediction accuracy is low. The practicability of the model trained in this way is relatively poor. In order to solve the over-fitting problem, we introduce the dropout parameter, which achieves the regularization effect to a certain extent. Since the input data is used to adjust the weight during the training phase, dropout constraints are added to the hidden layer of the model, which enhances the learning ability of the LSTM network model in the absence of individual connection information. This mechanism avoids the situation that certain data features are only effective under certain circumstances, and greatly improves the generalization ability of the model. As shown in Fig. 6, the overall prediction model with a dropout value of 0.5 has relatively good performance.

E. COMPARATIVE ANALYSIS

In Table 2 and Table 3, we recorded the predictive evaluation effects of 7 baseline methods and GST-GAT models in the future 15, 30, 60 and 120 minutes on METR-LA and

TABLE 2. Performance comparison of different traffic prediction methods on the METR-LA dataset.

Method	15min			30min			60min			120min		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	3.99	8.21	9.60%	5.15	10.45	12.70%	6.90	13.23	17.40%	9.32	15.67	20.64%
SVR	3.99	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%	8.94	13.85	18.15%
LSTM	3.44	6.30	8.60%	3.77	7.23	10.90%	4.37	8.69	13.20%	6.53	10.61	14.45%
STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%	6.82	11.21	15.26%
ASTGCN	2.75	5.62	7.51%	3.31	6.98	9.32%	4.52	9.24	12.62%	6.12	9.76	13.65%
STANN	2.73	5.64	7.53%	3.25	6.96	9.26%	4.13	8.21	12.05%	5.89	8.81	13.21%
GMAN	2.69	5.55	7.42%	3.15	6.78	9.02%	**4.03**	**8.11**	**11.72%**	5.42	8.64	12.92%
GST-GAT	**2.51**	**5.23**	**7.25%**	**3.07**	**6.55**	**8.57%**	4.10	8.16	11.89%	**5.26**	**8.12**	**12.16%**

TABLE 3. Performance comparison of different traffic prediction methods on the PEMS-BAY dataset.

Method	15min			30min			60min			120min		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	1.62	3.30	3.50%	2.33	4.76	5.40%	3.38	6.50	8.30%	5.13	8.04	11.87%
SVR	1.85	3.59	3.80%	2.48	5.18	5.50%	3.28	7.08	8.12%	4.97	7.76	11.45%
LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%	4.63	7.12	10.38%
STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%	4.72	7.31	10.62%
ASTGCN	1.32	2.78	2.75%	1.75	3.98	3.95%	2.32	5.41	5.51%	4.42	7.02	9.76%
STANN	1.38	2.84	2.86%	1.68	3.81	3.75%	1.97	4.54	4.57%	4.21	6.82	9.43%
GMAN	1.34	2.82	2.81%	1.62	3.72	3.63%	**1.86**	**4.32**	**4.31%**	3.95	6.61	8.89%
GST-GAT	**1.28**	**2.65**	**2.64%**	**1.48**	**3.52**	**3.41%**	1.92	4.41	4.39%	**3.58**	**6.54**	**8.54%**

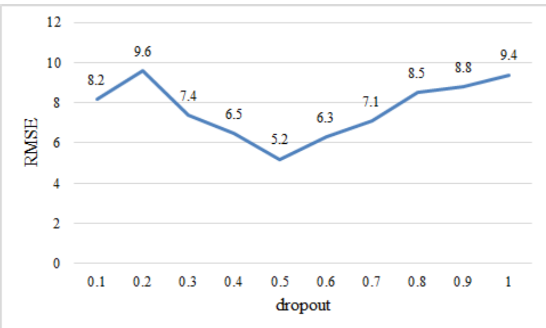


FIGURE 6. Tuning dropout parameters.

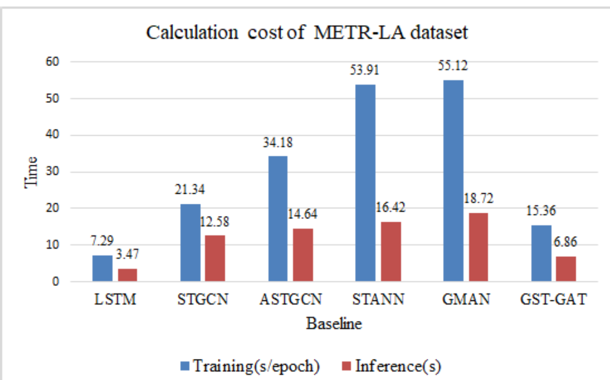
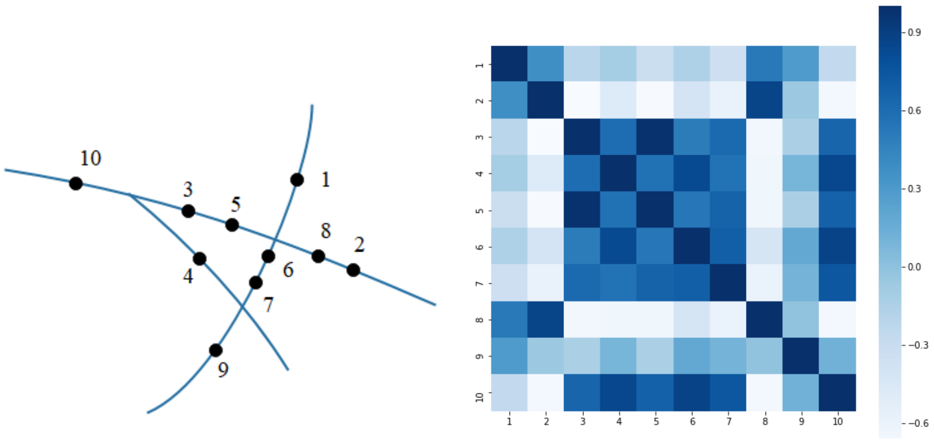


FIGURE 7. Calculation cost of METR-LA dataset.

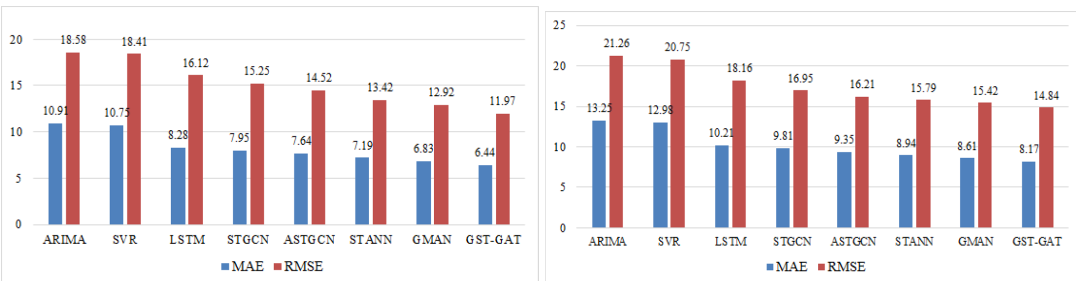
PMS-BAY data sets. The experimental results show that the GST-GAT model has excellent prediction performance on the two data sets, which shows that the model can better capture the spatial-temporal correlation between traffic road networks and has good generalization ability. Specifically, because of the poor generalization ability of the traditional time series prediction method (ARIMA) and the machine learning based prediction method (SVR) model. As the prediction step size becomes larger, the model prediction error increases significantly. The method based on LSTM improves the traffic prediction effect to some extent, but the model does not consider the spatial correlation between the road networks. In deep learning methods, graph-based models including STGCN, ASTGCN, STANN, GMAN and GST-GAT generally perform better than simple LSTM methods, indicating that spatial-temporal interaction between road networks is crucial for traffic prediction. Since the spatial relationship of the road network remains unchanged in GCN, the long-term traffic prediction effect based on the GCN model (STGCN, ASTGCN) is relatively poor, which is more obvious on the PEMS-BAY dataset. In order to capture the complex nonlinear time dependence and dynamic spatial relationship between road networks, STAN and GMAN both adopt structured networks for modeling, respectively using spatial attention and temporal attention modules to capture the dynamic temporal and spatial dependence between nodes. From the evaluation effect of the two data sets, GMAN performs best in the 60min traffic forecast. The reason may be that GMAN adds a gate fusion mechanism on the basis of STANN to adaptively control the flow of space-time interactive information, which helps to reduce the error propagation between prediction time steps. In particular, in GST-GAT, we use the "global interaction + node query" graph neural network modeling method to capture the specific impact of the global road network on a single traffic node. In addition, the gate fusion mechanism is used to self-adjust the mixed effects of spatial-temporal interactions. Compared with other baseline methods, GST-GAT performs best in experiments with prediction windows of 15 min, 30min and 120 min, which reflects the stable superior performance of GST-GAT in traffic forecasting.

In addition to evaluating the prediction effect of the model, we record the training time and reasoning time



(a) Sensor network distribution (b) Correlation analysis

FIGURE 8. Spatial correlation between sensors.



(a) Off-peak period 0:00 2:00 (b) Peak period 8:00 10:00

FIGURE 9. Performance comparison of the model during peak and off-peak hours.

[57] of six deep learning models on the METR-LA dataset in Fig. 7. Since the pure LSTM model does not consider the spatial-temporal interaction between road networks at all, the model has the fastest computing performance but poor prediction effect. Thanks to the use of spatial-temporal convolutional modules instead of recursive structures, graph convolutional neural networks can achieve fully parallel training. Compared with STAN and GMAN, STGCN and ASTGCN have faster running speed. It is worth noting that the introduction of the conventional attention mechanism not only improves the accuracy of traffic prediction of the model, but also causes a lot of repeated operations in the paired computation mode, which seriously slows the model's operation efficiency. Different from the above model, GST-GAT summarizes the traffic road network context information through global nodes, and calculates the specific impact of the global road network interaction on a single traffic node in parallel. Compared with GMAN, the calculation method of "global interaction + node query" reduces the time of GST-GAT training and reasoning by 72% and 63% respectively.

In order to intuitively study the role of the attention mechanism in the model, we conducted a case study: select 10 detectors from PEMS-BAY, and display the average spatial attention matrix between each detector in the training set. As shown on the right side of Fig. 8, the heat map represents the correlation between nodes in the spatial attention matrix. For example, focusing on the middle area, we can know that the traffic conditions between the 3rd, 4th, 5th, 6th, and 7th detectors are closely related. This is reasonable because the five detectors are very close in the actual traffic network, as shown on the left side of Fig. 8. Therefore, our model not only achieves the best predictive performance, but also shows the advantage of interpretability.

In addition, we tested the performance of the model during peak and off-peak hours on the METR-LA dataset. As shown in Fig. 9, MAE and RMSE during peak hours are higher than during off-peak hours because the traffic during peak hours is much higher than during off-peak hours. We also tested the performance of all benchmarks during peak and off-peak periods. Compared with other benchmarks, GST-GAT showed excellent predictive performance during peak and off-peak periods.

F. ABLATION ANALYSIS

In order to further investigate the performance of the GST-GAT traffic prediction model, especially whether the "global interaction + node query" model of traffic dynamic spatial-temporal correlation is really effective, we set up a set of ablation experiments for verification.

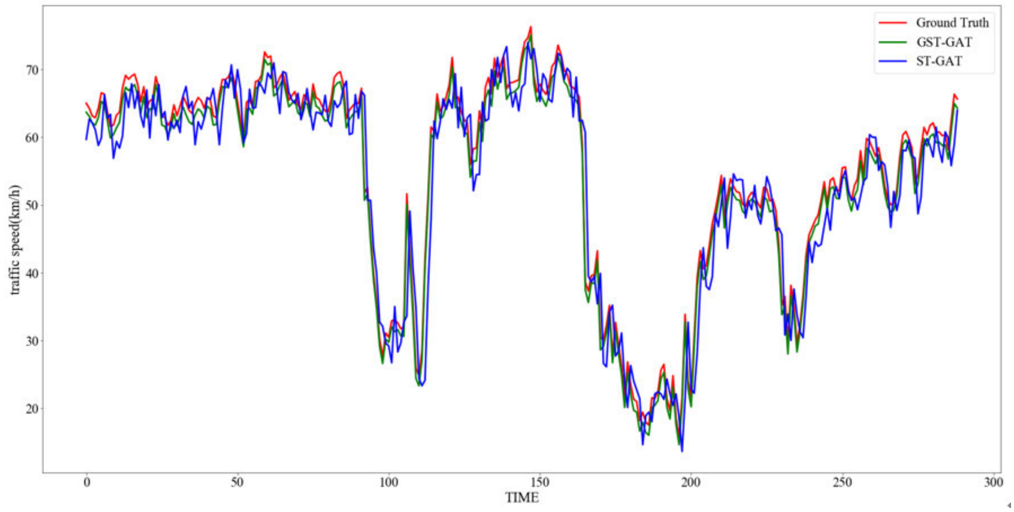


FIGURE 10. The ablation experiment on the METR-LA dataset(1).

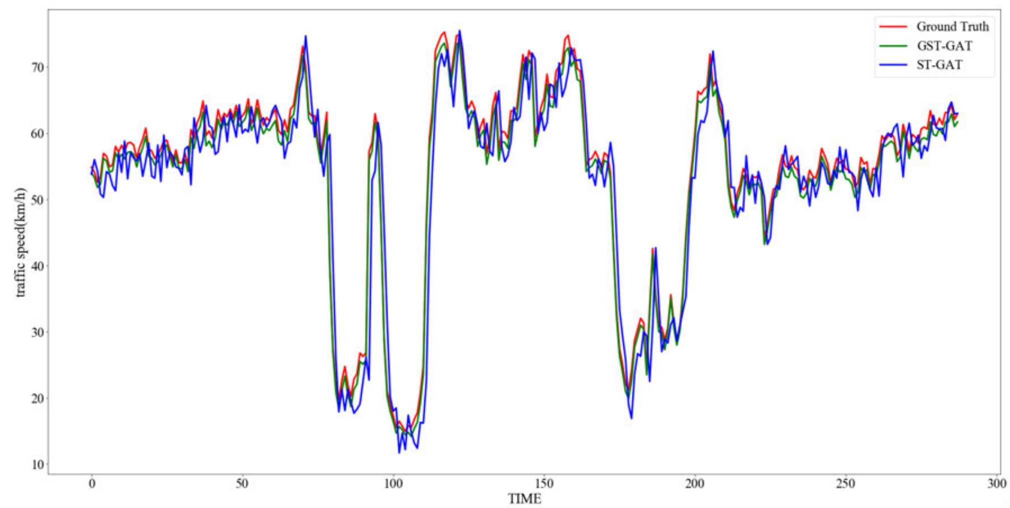


FIGURE 11. The ablation experiment on the METR-LA dataset(2).

ST-GAT: Referring to Fig. 3(a), this model adopts the classical spatial-temporal graph structure to model the spatial-temporal interaction relationship of traffic network, and extracts the spatial dependency relationship between sections by means of the graph attention mechanism.

GST-GAT: Referring to Fig. 3(b), the model adopts a global spatial-temporal graph structure to model the spatial-temporal interaction relationship of the traffic road network, and constructs a graph neural network model by the way of ''global interaction + node query''.

We test the above two graph neural network prediction models on the METR-LA data set, and randomly select two days of traffic speed prediction results for visual comparison. As shown in Fig. 10 and Fig. 11, the red line is the ground truth, the green line is GST-GAT, and the blue line is ST-GAT. Experimental results show that GST-GAT can capture the arrival of traffic peak more accurately. In addition, GST-GAT can quickly respond to dynamic changes between traffic networks. On the whole, the GST-GAT traffic prediction effect is closer to the real data, and there are some fluctuations in the ST-GAT traffic prediction.

V. CONCLUSION

In this article, we propose a novel graph neural network deep learning framework GST-GAT for traffic prediction. Based on the global spatial-temporal graph of traffic, we adopt the ''global interaction + node query'', an efficient way of information flow between nodes, to capture the impact of spatial-temporal interaction between traffic road networks. Specifically, the traffic global interaction is proposed to summarize the traffic network context changes and integrate all node features at each moment through a forward calculation. Then, each node computes the influence of traffic global interaction on a single node in parallel, and the spatial-temporal interaction information is adaptive fused by gating fusion mechanism. Experiments on two real data sets show that GST-GAT has great potential in exploring the spatial-temporal interaction of traffic networks. In particular, the structured GST-GAT has good flexibility and scalability. Parameter sharing of the same functional modules improves the generalization ability of the model, and efficient interactive computing makes the model respond faster. In the future work, we plan to introduce external factors (such as weather, POI, etc.) to further improve the accuracy of traffic prediction.

REFERENCES

[1] Z. Zheng, Y. Yang, J. Liu, H. N. Dai, and Y. Zhang, "Deep and embedded learning approach for traffic flow prediction in urban informatics," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3927–3939, Oct. 2019.

[2] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[3] G. Yan and Q. Qin, "The application of edge computing technology in the collaborative optimization of intelligent transportation system based on information physical fusion," *IEEE Access*, vol. 8, pp. 153264–153272, 2020.

[4] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, and J. Ye, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 1–5.

[5] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *IEEE Trans. Knowl. Data Eng.*, early access, Jul. 13, 2020, doi: 10.1109/TKDE.2020.3008774.

[6] Y. Han, S. Wang, Y. Ren, C. Wang, P. Gao, and G. Chen, "Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 6, p. 243, May 2019.

[7] X. Song, H. Kanasugi, and R. Shibasaki, "Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 2618–2624.

[8] J. Barthélemy, N. Verstaevel, H. Forehead, and P. Perez, "Edge-computing video analytics for real-time traffic monitoring in a smart city," *Sensors*, vol. 19, no. 9, p. 2048, May 2019.

[9] C. J. Dong, C. F. Shao, C. X. Zhuge, and M. Meng, "Spatial and temporal characteristics for congested traffic on urban expressway," *Bjing Gongye Daxue Xuebao/J. Bjing Univ. Technol.*, vol. 38, no. 8, pp. 1242–1246 and 1268, 2012.

[10] S. Fang, Q. Zhang, G. Meng, S. Xiang, and C. Pan, "GSTNet: Global spatial-temporal network for traffic flow prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 2286–2293.

[11] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," in *Proc. Adv. Neural Inf. Process. Syst.* 2019, pp. 137–146.

[12] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2255–2264.

[13] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, and W. Zhang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[14] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, Sep. 2015.

[15] 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.

[16] Y. Qi and S. Ishak, "Application of hidden Markov models to short-term speed prediction during peak periods," *Transp. Res. Board Meeting*, Dec. 2011.

[17] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 21, no. 1, pp. 148–162, Apr. 2012.

[18] X. Hao, G. Zhang, and S. Ma, "Deep learning," *Int. J. Semantic Comput.*, vol. 10, no. 3, pp. 417–439, 2016.

[19] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[20] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

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

[22] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, *Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting*. 2017.

[23] F. Kong, J. Li, B. Jiang, T. Zhang, and H. Song, "Big data-driven machine learning-enabled traffic flow prediction," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 9, p. e3482, Sep. 2019.

[24] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.

[25] K. Chen, F. Chen, B. Lai, Z. Jin, Y. Liu, K. Li, L. Wei, P. Wang, Y. Tang, and J. A. Huang, "Dynamic spatio-temporal graph-based CNNs for traffic prediction," 2018, *arXiv:1812.02019*. [Online]. Available: https://arxiv.org/abs/1812.02019

[26] S. Du, T. Li, X. Gong, and S.-J. Horng, "A hybrid method for traffic flow forecasting using multimodal deep learning," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, p. 85, 2020.

[27] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, no. 7, p. 2287, Jul. 2018.

[28] C. Chen, K. Li, S. G. Teo, X. Zou, and Z. Zeng, "Gated residual recurrent graph neural networks for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 485–492.

[29] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: https://arxiv.org/abs/1609.02907

[30] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. V. Berg, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.*, 2018, pp. 593–607.

[31] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1–7.

[32] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*. [Online]. Available: https://arxiv.org/abs/1707.01926

[33] 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.

[34] W. Chen, L. Chen, Y. Xie, W. Cao, and X. Feng, "Multi-range attentive bicomponent graph convolutional network for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 3529–3536.

[35] S. Guo, Y. Lin, S. Li, Z. Chen, and H. Wan, "Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3913–3926, Oct. 2019.

[36] L. N. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, "An effective spatial-temporal attention based neural network for traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 108, pp. 12–28, Nov. 2019.

[37] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee, "Enhancing transportation systems via deep learning: A survey," *Transp. Res. C, Emerg. Technol.*, vol. 99, pp. 144–163, Feb. 2019.

[38] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3428–3434.

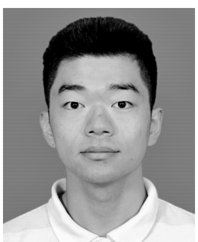
[39] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-RNN: Deep learning on spatio-temporal graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5308–5317.

[40] F. Gers, "Long short-term memory in recurrent neural networks," Ph.D. dissertation, Verlag Nicht Ermittelbar, 2001.

[41] P. Velikovi, G. Cucurull, A. Casanova, A. Romero, P. Liã, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*. [Online]. Available: https://arxiv.org/abs/1710.10903

[42] C. Zheng, X. Fan, C. Wang, and J. Qi, "Gman: A graph multi-attention network for traffic prediction," in *Proc. AAAI*, 2020, vol. 34, no. 1, pp. 1234–1241.

[43] C. Zhang, J. J. Q. Yu, and Y. Liu, "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166246–166256, 2019.
[44] Q. Zhang, J. Chang, G. Meng, S. Xiang, and C. Pan, "Spatio-temporal graph structure learning for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 1, pp. 1177–1185.
[45] J. Fu, W. Zhou, and Z. Chen, "Bayesian spatio-temporal graph convolutional network for traffic forecasting," 2020, *arXiv:2010.07498*. [Online]. Available: http://arxiv.org/abs/2010.07498
[46] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang, "TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 219–230, Jan. 2021.
[47] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2375–2384.
[48] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their application," *Biometrika*, vol. 57, no. 1, 1970.
[49] A. Vemula, K. Muelling, and J. Oh, "Social attention: Modeling attention in human crowds," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2018, pp. 1–7.
[50] S. Haddad, M. Wu, H. Wei, and S. K. Lam, "Situation-aware pedestrian trajectory prediction with spatio-temporal attention model," 2019, *arXiv:1902.05437*. [Online]. Available: https://arxiv.org/abs/1902.05437
[51] Y. Kim, P. Wang, and L. Mihaylova, "Structural recurrent neural network for traffic speed prediction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5207–5211.
[52] S. Chaudhari, G. Polatkan, R. Ramanath, and V. Mithal, "An attentive survey of attention models," 2019, *arXiv:1904.02874*. [Online]. Available: https://arxiv.org/abs/1904.02874
[53] Z. He, C.-Y. Chow, and J.-D. Zhang, "STANN: A spatio–temporal attentive neural network for traffic prediction," *IEEE Access*, vol. 7, pp. 4795–4806, 2019.
[54] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "TrafficPredict: Trajectory prediction for heterogeneous traffic-agents," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 6120–6127.
[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[57] Y. Zhu, D. Qian, D. Ren, and H. Xia, "StarNet: Pedestrian trajectory prediction using deep neural network in star topology," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Nov. 2019, pp. 1–6.



BIN SUN received the B.S. degree from the College of Electronic Information Science and Technology, Xuzhou University of Technology, in 2018. He is currently pursuing the M.S. degree with the School of Information and Control Engineering, China University of Mining and Technology. His current research interests include traffic prediction, the Internet of Things, artificial intelligence, and edge computing.

DUAN ZHAO received the B.S. and M.S. degrees from the School of Information and Electronic Engineering, China University of Mining and Technology (CUMT), Xuzhou, China, in 2006 and 2009, respectively, and the Ph.D. degree in communication engineering from CUMT. From 2011 to 2013, he was with the School of Communication Technology, University of Duisburg-Essen, Germany, as a Visiting Ph.D. Researcher. He is currently a Research Assistant with the IoT Center, CUMT. His research interests include wireless sensor networks, energy harvesting in coalmine, and machine learning.

XINGUO SHI was a Technician with Zibo Mining Bureau Engineering Company from 1994 to 1998. From 1998 to 2009, he served as the Deputy General Manager with Geting Coal Mine of Zibo Mining Group Company. From 2009 to 2015, he served as the Deputy General Manager of Xinhe Coal Mine of Zibo Mining Group Company. He is currently the Director of the Information Center of Shandong Energy Zibo Mining Group Company, Ltd. He is currently a Senior Engineer. He led the completion of 135 Zibo Mining Group information construction projects, covering all levels of information systems, such as underground business processing, safety warning, and auxiliary production. These tasks have laid a solid foundation for enterprises to promote digital and intelligent development. His research focuses on the construction and optimization of digital mines.

YONGXIN HE received the B.S. degree from the College of Electronic Information Science and Technology, Xuzhou University of Technology, in 2018. He is currently pursuing the M.S. degree with the School of Information and Control Engineering, China University of Mining and Technology. His current research interests include the Internet of Things, wireless sensor networks, and energy harvesting technology.