RTS-GAT: Spatial Graph Attention-Based Spatio-Temporal Flow Prediction for Big Data Retailing

SHIYANG LUO, (Member, IEEE)
Shanghai Infortech Software Development Company Ltd., Shanghai 200233, China
e-mail: shayne_luo@foxmail.com

ABSTRACT Intelligent logistics is a crucial element in the era of innovative retailing. The retail industry may benefit from more sophisticated logistics forecasting, management, and collaboration to lower costs, boost efficiency, and improve service standards. It is vital to intelligent retail’s optimization of logistics and retail product transportation. Decoupling the retail road network traffic flow big data from the temporal or spatial dimension and creating strong correlations are the keys to effective modeling since it is a high-dimensional, Spatio-temporal sequence. While the spatial dimension does not correlate with the temporal dimension of the retail road network traffic flow, the temporal dimension varies continuously. This feature allows us to separate the spatial-temporal sequence of retail road network traffic flow into spatial graph data and model it primarily using the spatial correlation paradigm. In this study, we examine the variables that affect the retail road network’s traffic flow based on the demand for human travel. And we suggest a unique deep learning-based representation and fusion method of temporal, spatial, and traffic flow features as features of the spatial graph nodes of the shared spatial graph neural network framework RTS-GAT. Individual primitive features are restricted to be simultaneously learned and optimized in the feature space associated with each other in the modeling context of shared model parameters, and efficient eigenfeature representations are learned. After rigorous and comprehensive experimental validation, the RTS-GAT model achieves the best performance to date on multiple datasets.

INDEX TERMS Attention mechanism, big data retailing, graph neural network, intelligent retail logistics.

I. INTRODUCTION
A. RESEARCH BACKGROUND
This paper’s research is focused on how intelligent logistics will change the retail sector. A neural network technique is used as a research tool to predict traffic flow on the road. Using neural networks [1] to model spatiotemporal sequence data [2], [3] from road network traffic flow has become popular in recent years. However, when it comes to the related issue of road network traffic, the majority of study [4], [5], [6], [7], [8] studies fall short due to the association structure’s complexity, multiplicity, and even unknowability. In this section, we go through how traffic flow data from roads is organized and how deep learning is used in this context.

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

We discover that graph neural network [4], [9], [10] and attention mechanism [11] are appropriate for simulating the relationship between traffic flow on a road network.

The logistics road network serves as the primary traffic container in this study. In the case of the retail sector, the primary traffic flow object acting as a traffic container [12] on the road network is simply the traffic flow, and the traffic flow that the traffic flow carries is the flow of people. The road network is, therefore, the carrier of the human flow, and the human flow represents the need for mobility among humans as well as the underlying behavioral patterns. Road network traffic data can be viewed from the standpoint of human behavior as data samples on human traffic trip behavior that were gathered utilizing technology. It is possible to represent the data collection space for traffic flow and velocity as a set of points in space.
In practice, toll booths, magnetic coils placed throughout the road network, and cameras are some of the technical tools used to gather data on traffic movement. The overall traffic flow status on the road network can be approximated by combining the data from each space. We can obtain a regular tabular representation of the traffic flow information [13] of the road network by adding the traffic flow information in each time slice or averaging the traffic speed information. The sole variation between the data types for traffic flow and traffic speed is that the traffic speed data contains a large number of null values. When the traffic flow data contains null values, no cars traverse that road network area, and the traffic flow is 0. Road sections or stations can be used to segment the data-gathering area for the flow of the road network. The graphic’s edges representing the traffic network [14], [15] are the road segments, and the points connecting the stations are the stations. As they arrive and the toll booths record exit, vehicles entering or exiting the network from the stations. Vehicles passing by the edges are collected by the coil detectors scattered throughout them. As can be observed from this definition, traffic flows on road segments and stations have very distinct characteristics. Road segment traffic and station traffic flow data are both Spatio-temporal sequence data [16], [17] and have comparable data structure formats. However, these two types of Spatio-temporal series data have different traffic logical meanings and display highly different data features due to the difference in their collection spaces.

**B. RESEARCH MOTIVATION**

In the preceding section of the research background, the introduction of road traffic flow data describes the data display format. It can be distinguished depending on how the data is collected (road section, station) and presented (flow rate, flow speed). Despite their different methods of collection or contents, the geographical and temporal series data of the traffic flow on the road network have comparable mathematical structures [18]. The flow and velocity statistics should not be confused because the traffic flow on road segments and stations have different characteristics. We need to understand the correlation [19] underlying the emergence of the traffic since the spatiotemporal series data of road network traffic flow that was collected is merely the outward manifestation of traffic. There are two components [20] to the phrase “road network traffic flow”: a static road network and a dynamic traffic flow. While traffic flows may compete with one another for resources in the road network, resulting in complicated changes in traffic flow, road networks also lead to mutual influence between networks while achieving geographical interoperability in the form of networks. Our study argues that it is crucial to adequately characterize the spatial correlation between these numerous time series to model road traffic flow’s geographical and temporal data successfully. How can the association between traffic flow on roads be understood, furthermore? Intrinsic and interactive relevance are the two sections of this essay.

Humans travel, and the movement of traffic reflects their transportation demands. We begin at a starting place and travel to an endpoint, indicating that this starting point and end point have particular significance or utility. What should be done to model the intrinsic value of traffic road networks [21]? One suggestion is to gather more data, such as the distribution of points of interest nearby each location in the spatial dimension. Most traffic flow databases on road networks do not contain this information, and it may not be possible to depict the detailed information found in the actual world accurately. Intrinsic relevance is frequently not immediately accessible, and even the comparatively meaningful relationships of commuter demand behavior require adequate data analysis. By using substantial data-driven eigen-feature representation, deep models can automatically learn the inherent relevance. Assume that the intrinsic correlation results from how each area outside the road network serves humans. In that situation, traffic flow interactions on the road network cause the interactive correlation [22], [23]. The usual spatial limitations hamper the road network’s traffic flows. The road network is a complicated system of interconnected roads where traffic flows from different locations converge, diverge, cross, and unavoidably impact one another.

The previous section discussed two types of correlation: interactive correlation within the road network [24], which is dynamic in time and relative to other roads, and intrinsic correlation outside the road network [25], which is static in time and relative to human requirements. What kind of association is it, intrinsic or interactive? Eigenrelations are concerned with the motivations of human actions in traffic. However, we are unaware of any information besides the traffic flow data on the road network. Even if specific qualitative descriptions can be created using human a priori knowledge, human cognition is limited, and it is not easy to quantify. In other words, there is a correlation between traffic flow on the road network, but it is complex, challenging to explain, and sometimes even unknown.

**C. INNOVATIVE CONTRIBUTIONS**

The following three elements make up this paper’s contribution to academic innovation.

- In order to organize the data in the structure of spatial graphs based on the temporal, spatial, and traffic flow characteristics of spatial graph nodes, a big data-driven edge representation modeling scheme is proposed. This analysis looks at the Spatio-temporal correlation structure of traffic flow on retail road networks as well as the characteristics of the short-time traffic flow prediction problem.
- We propose the concepts of intrinsic association and interactive association and develop the spatial graph neural network framework TSGCN to implicitly implement the intrinsic association modeling of traffic flows in retail road networks by sharing model channels after analyzing the association problem in modeling the traffic logistics of retail road networks.
We introduce a multi-headed dot product attention mechanism to model interactive correlations and builds RTS-GAT in which edge relations and interactive correlations are learned together in a complementary manner. Both of these techniques are based on the spatial graph neural network framework. A universal approach that can be used for many retail traffic and logistics datasets has produced excellent prediction results on two genuine retail traffic and logistics datasets.

This paper’s research focuses on Spatio-temporal sequence modeling for road network traffic flow, a crucial area of intelligent logistics systems in the retail sector. The article explicitly chooses the phenomena of “association” of retail road network traffic flow as the starting point for the retail logistics road network. It incorporates the most recent deep learning technique to model the association and enhance the effects of road network modeling.

II. RELATED WORK
Road traffic flow modeling has been an important problem in transportation for a long time, and the problem of short-time traffic flow prediction has been the benchmark for measuring the effectiveness of road traffic flow modeling. In this paper, we consider the road network traffic flow as Spatio-temporal series data and analyze its characteristics in terms of Spatio-temporal correlation based on the decoupling method and correlation method of the data in the time dimension and spatial dimension.

The first modeling paradigm is the early and most straightforward one when simple time series models such as Informer [26] can also be modeled. In this paradigm, the time series of each space are isolated from each other, and the Spatio-temporal series data [27], [28] are reduced to multiple mutually independent one-dimensional time series data; only the correlation of spatial dimensions is ignored. This type of modeling paradigm escapes the complexity of the Spatio-temporal sequence problem [29] and has been abandoned by the current researchers.

The second type of modeling paradigm is the temporally correlated spatial paradigm, which can cleverly combine the natural temporal characteristics of traffic flow Spatio-temporal sequences with recurrent neural networks that process sequential data, thus reducing the number of model parameters for modeling traffic flow Spatio-temporal sequences by letting the model process the states of each time slice in steps. Recurrent neural networks have the excellent property of being iteratively scalable due to their natural temporal sequence modeling capabilities and can be used to predict traffic flows of arbitrary length. However, these RNN [30] class models, such as LSTM [31], LSTM-CRF [32], Deep-RNN [33], and EESEN [34], have some problems, such as scaling up the error progressively in the traffic flow prediction task. The approach used in modeling takes the predicted value output at the last moment as the input at the current moment, which means that its input traffic flow signal is in error in the subsequent multi-step prediction [35], [36] and this error is also gradually amplified in the subsequent prediction. At the same time, the stepwise iterative approach has the problem of slow model training because of its serial connection style.

The third type of modeling paradigm is the spatially correlated temporal paradigm, and the neural network of this paradigm is implemented as a graph neural network approach [37]. The key to modeling graph neural networks is to represent graph nodes efficiently and to model associations between graph nodes reasonably to pass messages. A variant of graph convolutional neural networks based on diffusion effects, DCRNN [38], adopts this paradigm.

The fourth type of modeling paradigm is the Spatio-temporal association paradigm, which requires modeling the Spatio-temporal dimension of the traffic flow Spatio-temporal sequences [39] separately according to their correlation. Researchers have recently heavily adopted the Spatio-temporal correlation paradigm in traffic [40]. The basic idea is to model the correlation characteristics of traffic flow’s temporal and spatial dimensions separately and then integrate the two types of correlation modeling by graph convolution [41], attention mechanism [11], recurrent neural network, and other methods. One of the classic papers is the diffusion convolution recurrent neural network DCRNN which adopts the third and fourth type of modeling paradigm at the same time and is often cited as the classic model. In terms of temporal correlation, the authors employ recurrent neural networks to model the temporal dynamics of traffic flow on road networks. Specifically, the authors used the gated recurrent unit GRU [42], which allows each Spatio-temporal graph node to determine to what extent historical information is retained through a gating mechanism, thus establishing a correlation in the temporal dimension. In the spatial dimension, the authors obtain a diffusion convolution kernel based on the phenomenon of spatial diffusion effect in traffic science, which is calculated based on the distance between individual nodes. Some parallel second-order and third-order positive and negative convolution kernels are obtained by self-multiplication [43] and inversion of this convolution kernel to represent the static correlation matrix of road network traffic flow [44] in the spatial dimension. By fusing the graph convolutional neural network [39] and recurrent neural network, the original authors combined the temporal and spatial correlation modeling to obtain the diffusion convolutional [38], [45] recurrent neural network DCRNN.

III. METHOD AND THEORETICAL ANALYSIS
The spatiotemporal data representing traffic flow on a road network is dispersed in the time and space dimensions for a typical traffic flow modeling task. It is typically high-dimensional data. High-dimensional data processing is frequently too computationally expensive, making some form of division and rule imperative. On the one hand, the spatial correlation time modeling paradigm is a superior option for modeling typical road network traffic flow because of the
significant temporal continuity and sparse dynamic spatial correlation of geographic and temporal series data of road network traffic flow. On the other hand, we select to complete the modeling of complex correlations through a more appropriate graph neural network when we need to undertake complicated correlation modeling to identify the changing patterns of road network traffic flow from the data. In summary, this study processes the spatiotemporal sequence data of traffic movement on a road network and models it using the graph neural network technique. Specifically, we model the temporal and spatial aspects of the road network’s traffic flow characteristics using spatial graph data modeling. These models are then input to a shared perceptron to build a model called TSGCN which is described in technical detail in the following subsection. Then, tests are created to determine the importance of various feature types for modeling road traffic flow and to confirm the efficiency of graph neural networks in handling traffic flow.

A. SPATIAL GRAPH NEURAL NETWORK
This chapter aims to suggest a modeling approach of feature representation and fusion for the numerous influencing factors that contribute to the phenomenon of road traffic flow. It does this by organizing the Spatio-temporal sequences of road network traffic flow in the form of spatial graph data. Furthermore, in order to realize road traffic correlation modeling and create a solid foundation for the subsequent interactive correlation modeling, learn the eigenfeatures of each spatial graph node collectively through a shared spatial graph neural network. The short-time traffic flow prediction problem is used as the modeling aim in this chapter’s spatial graph neural network framework in order to simplify the presentation. It combines spatial, temporal, and historical traffic flow variables to forecast future traffic flows.

$$[G^{(t-T'}, \ldots, G^{(t')}; D] \xrightarrow{f} [G^{(t+1)}, \ldots, G^{(t+T')}]$$

This chapter’s content focuses on structuring the spatiotemporal sequence data of road network traffic flow according to the structure of spatial graph data and creating the framework for a spatial graph neural network. The following section explains each module of this framework.

1) INPUT AND OUTPUT
The model’s inputs and outputs, which primarily deal with the issue of predicting short-term traffic flow, must be specified. The spatial graph neural network will take the attributes of each graph node as input and output the predicted future traffic flow for that graph node when we utilize the spatial graph as the data structure. Future traffic flow features will be the features used as input. Historical traffic flow, spatial variables, and temporal factors make up the bulk of the information for the data set in this study. As the importance of short-term historical traffic flows is based on traffic flow continuity, the relevance of human travel patterns to temporal and geographical aspects has been examined in the previous section. This paper’s subsequent subsections describe the individual modules of the graphical neural network framework.

2) SPATIAL REPRESENTATION
The embedding method is a popular deep-learning technique for teaching new features. Embedding, initializing embedded features into learnable parameter vectors at random and adapting the optimization during training the model. Word2vec [46] in natural language processing and graph neural networks in the Network Representation Learning (NRL)
branch [47] of graphical neural networks are just two examples of embedding techniques with a wide range of applications. Both deftly exploit the full feature learning capabilities of deep learning to acquire useful embedded feature representations. We concentrate on model task-oriented learning of eigenfeatures in this study.

\[
\text{Emb}(I) = \begin{pmatrix} I_1^T \\ \vdots \\ I_n^T \end{pmatrix}, \quad s_i = \begin{pmatrix} I_i^T \\ \vdots \\ I_{n_{space}}^T \end{pmatrix}
\]

(2)

\[
l_i = \begin{bmatrix} I_i; \text{ longitude, latitude} \end{bmatrix}
\]

(3)

Among them, \( S_d \) represents the dimension of spatial representation. Other spatial data points are also accessible with the known geographical information, including geographic coordinates (latitude and longitude), nearby areas (neighboring nodes), and the road network’s structure. We can also incorporate valuable portions of this data into the spatial representation when building models. Inferred information regarding the relative position of space on geographic space is one of them, and it is contained in the latitude and longitude information of geographic coordinates. We can quickly compute the geographic distance between them based on the latitude and longitude characteristics of the two places. Traffic flow on a road network might also follow some spatial resemblance in the immediate vicinity, and either congestion or smoothness will extend to nearby areas. As a result, adding geographic position information to the spatial feature representation makes sense. In order to aid in modeling, the longitude and latitude data are standardized in this study to constants with a mean \((0,0)\) and variance \((1,1)\) and spliced behind the joint entry vector parameters. The addition of latitude and longitude data can aid the model in understanding the relative proximity of two spatial graph nodes on the feature space, speeding up training and possibly boosting accuracy.

### 3) TIME REPRESENTATION

The link between human travel patterns and time is the root cause of the strong correlation between traffic flow dynamics and time. The addition of temporal variables unquestionably helps in modeling traffic flow. Similar to how spatial characteristics are described, temporal features are represented as a learnable parameter vector using a join-in technique. Only the temporal labels must be set without considering the specific feature structure. The daily and weekly cycles are significantly connected aspects in predicting short-term traffic flow, which is also consistent with human common sense and social production mechanisms. As a result, the temporal aspects are also split into two categories: the intra-day period, which represents the daily cycle, and the day date, which represents the weekly cycle.

\[
\text{Emb}(d) = \begin{pmatrix} d_1^T \\ \vdots \\ d_{n_{space}}^T \end{pmatrix}, \quad \text{Emb}(w) = \begin{pmatrix} w_1^T \\ \vdots \\ w_{n_{space}}^T \end{pmatrix}
\]

(4)

\( s_i \) is the number of time slices in a day, determined by the total duration and the granularity of the time slice division. Furthermore, the weekly period from 1 to 7 represents the seven days of the week.

Another aspect of time must be taken into account in addition to the embedded representation: the relative location of the various times, which remain perpetually next to one another on the time axis. The traffic flow time series in traffic science also follow the principle of temporal adjacency similarity, where the traffic flows of roughly consecutive periods have a strong correlation. On the other hand, temporal locations are represented to take advantage of the continuity of temporal characteristics. A straightforward solution is to use a number between \([-1,1]\) to represent time, for example, \(-1\) for when time begins and 1 for when it ends. This solution, however, ignores a crucial aspect of time, namely the periodicity of the end as the beginning.

\[
\text{location}_{r_i} = \begin{bmatrix} \sin \frac{2\pi}{m}\iota \cos \frac{2\pi}{m}\iota \end{bmatrix}
\]

(5)

\[
r_i = \begin{bmatrix} r_i; \text{ location}_{r_i} \end{bmatrix}
\]

(6)

Clock location embedding refers to connecting the location attributes of each minute of the day to create the shape of a clock. The daily cycle moment features are combined with the clock location embedding features to create a time feature representation that includes positional data. The position representation is not necessary due to the length of the weekly cycle and the minimal continuity between Sunday and Monday. Thus, the clock location and time embedding features can be stitched together to obtain the time feature representation \( r_i \).

### 4) TRAFFIC FLOW REPRESENTATION

Depending on the data set, traffic flow can be represented mathematically as either flow rate or flow rate. Each geographic node’s historical short-term traffic flow is input to each spatial graph node as a traffic flow characteristic in the problem of short-term traffic flow prediction. To ensure that the input and output of the neural network obey normal distribution, the traffic flow is normalized to a numerical vector \( X_s \) with mean 0 and variance 1, which can be represented as:

\[
x_i = \begin{pmatrix} x_{i_s}^T \\ \vdots \\ x_{i_{n_{space}}}^T \end{pmatrix}, \quad X_s^i = (x_{i_s}^{T-1}, \ldots, x_{i_s}^{T})
\]

(7)

A nonlinear transformation is then applied to the numerical vector representing the traffic flow to obtain the state representation of the traffic flow.

\[
h_{\text{flow}} = W_{\text{flow}} X_s + b_{\text{flow}}
\]

(8)

\[
\text{Emb}(m) = \begin{pmatrix} m_1, \ldots, m_T \end{pmatrix}
\]

(9)

where \( m_i \) represents the feature representation of the missing traffic flow data, since the missing values at different locations have different meanings, they need to be represented...
separately according to the location. This can cause some trouble in the calculation. We can represent the effect of the missing value of location on the traffic flow state as:

\[
\mathbf{h}_s^{\text{flow}} = \mathbf{W}^{\text{flow}} \mathbf{X}_s + b^{\text{flow}} + \sum_{i, X_i^{\text{null}} \in N} \text{null}_i
\]  

(10)

How can it be efficiently vectorized for matrix operations? We can fill the traffic flow data at the default values with zeros and use a 0-1 vector to depict the distribution of default values. As a result, the following representation is the most ideal for computing traffic flow characteristics:

\[
\mathbf{h}_s^{\text{flow}} = \mathbf{W}^{\text{flow}} \mathbf{X}_s + \mathbf{W}^{\text{null}} \text{Null}_s + b^{\text{flow}}
\]  

(11)

In doing so, we maintain the use of an embedding representation for the missing data and, in actuality, use vectorized computation to enhance the computational performance.

**B. SPATIAL GRAPH ATTENTION NETWORK FRAMEWORK**

The spatial graph attention network presents an attention mechanism to explain the interaction correlation between spatial graph nodes using a spatial graph neural network as its fundamental building block. Because of this, we concentrate on the multi-headed dot product attention mechanism in this section and suggest a form of the spatial graph attention recurrent network based on the Spatio-temporal association modeling paradigm.

1) MULTI-POINT ACCUMULATION ATTENTION MECHANISM

How to build a multi-headed dot product attention mechanism? We need to map the features \( f \) of graph nodes to different subspaces. The multi-headed dot product attention mechanism can be obtained by stitching together several different dot product attention mechanisms.

\[
\text{MultiHead}(f) = \text{Concat}(\text{head}_1, \ldots, \text{head}_{n_{\text{head}}}) W^O
\]

where \( \text{head}_i = \text{Attn}(f W^q_i \cdot f W^k_i, f W^v_i) \)

(12)

where \( W^q_i \in \mathbb{R}^{d_{\text{model}} \times d_k}, W^k_i \in \mathbb{R}^{d_{\text{model}} \times d_k}, W^v_i \in \mathbb{R}^{d_{\text{model}} \times d_v} \) map the feature \( f \) of spatial graph nodes into the subspaces of query, key, and value. The sum of each subspace’s dimensions is equal to the model dimension \( W^O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}} \). Further, the attention results of each subspace of multi-headed attention are linearly mapped to integrate the output and summed with residual connections and \( f \). \( n_{\text{head}} \) is the number of heads of a multiple, usually set to 1,2,4,8,16,... The number of heads further determines the number of dimensions \( d_{\text{model}} \) of each head.

2) MODEL FUSION

The model can learn the underlying correlations of road network traffic flows based on the spatial graph neural network architecture, which can fuse diverse information sources from arbitrary data sets into the properties of spatial graph nodes. The attention layer will then do self-attentive computing among graph nodes using a multi-headed attention mechanism to facilitate message forwarding based on graph node attributes, and the interactive correlation is modeled. We are able to simulate correlations and interactive correlations using a general graph neural network-based deep learning model called RTS-GAT which is shown in Figure 1 for short-term traffic flow prediction challenges.

3) SPATIAL GRAPH ATTENTION RECURRENT NETWORK

We may create a spatial graph attention recurrent network modeled in the Spatio-temporal association paradigm to compare the modeling impacts of the two paradigms and to examine the potential of the multi-headed attention mechanism. As in this paper, we integrate the gated recurrent unit GRU with the multi-headed attention mechanism to update the input and output gates of the GRU based on spatial interaction correlation. This is similar to the diffusion convolutional recurrent neural network DCRNN.

\[
G_i = (G_i, G_i; D)
\]

(13)

\[
G_i = (G_i \oplus N_{i-1}, N_{i-1}; D)
\]

(14)

\[
U_t = \sigma(\text{Attn}(G_i) + \text{Attn}(G_n))
\]

(15)

\[
R_t = \sigma(\text{Attn}(G_i) + \text{Attn}(G_n))
\]

(16)

\[
N'_i = n(\text{Attn}(G_i) + R_i \circ \text{Attn}(G_n))
\]

(17)

\[
N_i = (1 - U_t) \circ N'_i + U_t \circ N_{i-1}
\]

(18)

As is shown in Figure 1, RTS-GAT can be built in this manner. This approach’s input and output data are distinct, and the traffic flow is no longer extended to the input and output as numerical vectors. Instead, a stepwise prediction method is used to train a recurrent neural network. Each time slice’s spatial and temporal data are fed into the model along with the traffic flow signal to anticipate the next slice. Furthermore, the anticipated value serves as the input for the following action. Recurrent neural networks will use the spatial graph attention recurrent network to capture the dynamic trend data of traffic flow in the temporal dimension.

4) ADVERSARIAL TRAINING ALGORITHM

Countermeasure training involves adding a minor disruption to the original input to generate adversarial samples that could be utilized for training [65]. It can be represented by the model below:

\[
\text{Loss} = -\log p(y|r_d dv + e; \tilde{\xi})
\]

\[
\text{wherer}_d dv = \arg\min_{r, |r| < \epsilon} \log p(y|r + e; \tilde{\xi})
\]

\[
= -\arg\max_{r, |r| < \epsilon} \log p(y|r + e; \tilde{\xi})
\]

(19)

In particular, given the model \( f_c(m; \xi) \) and \( k \) data points of the target task indicated by \( (e_i, y_i)_{i=1}^{k} \), where \( e_i \)’s signify the embedding of the input sentences derived from the first embedding layer of the language model and \( y_i \)’s are the associated labels, our technique performs the following
optimization for fine-tuning:

$$\min_{\zeta} = T_p R_p(\zeta) + \frac{1}{k} \sum_{i=1}^{k} l(fcn(\zeta; e_i), y_i)$$

(20)

where \(\text{Loss}(\theta)\) is the function of loss defined as:

$$\text{Loss}(\theta) = \frac{1}{n} \sum_{i=1}^{n} l(f(x_i; \theta), y_i)$$

(21)

and where \(fcn(\cdot)\) is the loss function that changes based on the goal task, \(T_p\) is a tuning parameter, and \(R_p(\zeta)\) is the adversarial regularizer that promotes smoothness. Here is how we define \(R_p(\zeta)\):

$$R_p(\zeta) = \frac{1}{k} \sum_{j=1}^{k} \max_{\|e_i - e_j\|_p \leq \epsilon} l_s(\zeta; fcn(\tilde{e}_i), fcn(\zeta; e_i))$$

(22)

where \(\epsilon\) is an adjustment factor, be aware that \(fcn(\zeta; \cdot)\) produces a probability simplex for classification tasks, whereas \(l_s\) is selected as the symmetrized KL-divergence.

$$ls(M, N) = D_{KL}(M\|N) + D_{KL}(N\|M)$$

(23)

\(fcn(\cdot)\) outputs a scalar for regression tasks, whereas \(l_s\) is chosen as the squared loss. \(l_s(m, n) = (m - n)^2\) notes that the computation of \(R_s(\zeta)\) includes a problem involving maximizing, which can be effectively handled via projected gradient ascent.

IV. EXPERIMENTAL SETUP

The following research questions are addressed in this part by extensive experimentation.

RQ1: How do standard one-dimensional time series forecasting models fare compared to spatial graph neural networks?

RQ2: How effectively do various tuned structural derivative models learn inherent correlations relative to spatial graph neural networks?

RQ3: With the addition of the attention mechanism, the initial spatial graph neural network TSGCN can be molded through a message-passing path. Then, how much benefit can the attention mechanism provide when modeling the interactive correlation of spatial and temporal sequences of traffic flows on a road network?

A. DATA PREPROCESSING

The tests employ data from the Beijing Expressway and the Los Angeles Expressway for predicting the short-term traffic flow rate. Similar to other investigations, we used traffic flow data with 5-minute temporal granularity. It is analogous to utilizing the dynamic information from the first 12 time slices to predict the dynamic information from the following 12-time slices to use the historical traffic flow rate of the first hour to anticipate the traffic flow rate of the following hour. For data on traffic flow, we employ a time granularity of 15 minutes and use the first two hours to estimate the following two hours, which entails forecasting the subsequent 12-time slices based on the first 12 time slices. For the final performance comparison, we compare the MAE, RMSE, and MAPE forecast accuracy using 15 minutes, 30 minutes, and 1 hour of traffic flow rate prediction.

B. METRICS

This paper uses the evaluation metrics commonly used in short-term traffic flow forecasting, and the mean absolute error (MAE) [48] and root mean square error (RMSE) [49] for flow and velocity forecasting tasks. However, both MAE and RMSE are not intuitive numerically, so we usually use percentage error when evaluating the forecasting accuracy. However, in terms of percentage error, the calculation is different for traffic flow and velocity. For traffic flow rate, the percentage error is the Mean Average Percentage Error (MAPE) [48]. However, for traffic flow forecasting, it is not appropriate to use Mean Average Percentage Error. When the traffic flow is small, the traffic flow as the denominator will magnify its percentage error. For example, in the typical case where only one vehicle or no vehicle passes, the difference of just one vehicle will result in a 100% error. Therefore, for the traffic flow prediction task, the percentage error needs to be weighted according to the traffic flow and evaluated using Weighted Average Percentage Error (WAPE) [50].

C. BASELINES

1) HA [51] & ARIMA [52] & SVR [53]

The historical average (HA) model used in the prior study is inaccurate and unable to capture the characteristics of passenger flow that change over time. Over the past three decades, one of the most widely used linear models for time series forecasting has been the autoregressive integrated moving average (ARIMA). To predict short-term freeway traffic flow under both usual and atypical circumstances, SVR presents an application of a supervised statistical learning technique.

2) FNN [54] & AGNN [55]

Fully connected neural networks, or FNNs, are also referred to as multilayer perceptrons (although they essentially are made up of several layers of logistic regression). In order to provide valuable recommendations, a novel supervised Adaptive Genetic Neural Network (AGNN) method is proposed. It finds the cluster’s most favored data points.

3) GCN [56] & EAGCN [57]

GCN presents a scalable method for semi-supervised learning on graph-structured data, and it is based on an effective convolutional neural network version that operates directly on graphs. A space-adaptive graph convolutional module is presented by EAGCN and has the potential to investigate the transmission of user interest and social impact simultaneously. A user-specific gating mechanism is additionally created to combine user representations from both areas.
### TABLE 1. TSGCN model performance comparison experiment.

| Dataset     | Model     | 15 minutes | 1 hour/30 minutes | 2 hour/1 hour |
|-------------|-----------|------------|-------------------|--------------|
|             | MAE       | RMSE       | (M/W)/APE         | MAE          | RMSE       | (M/W)/APE |
| Being Expressway | HA  | 19.03 | 35.21 | 23.48% | 19.24 | 35.26 | 23.67% | 19.65 | 35.91 | 24.55% |
|             | ARIMA     | 14.98 | 26.71 | 19.43% | 19.07 | 34.47 | 23.95% | 27.12 | 47.33 | 32.13% |
|             | SVR       | 14.32 | 21.89 | 17.53% | 17.11 | 27.62 | 20.09% | 21.38 | 36.79 | 27.76% |
|             | PNN       | 15.12 | 22.34 | 18.16% | 17.27 | 28.47 | 21.19% | 21.97 | 37.36 | 28.11% |
|             | GCN       | 13.28 | 20.65 | 16.47% | 15.85 | 25.26 | 19.35% | 19.63 | 34.77 | 25.45% |
|             | GAT       | 12.76 | 20.33 | 16.18% | 15.24 | 24.79 | 19.04% | 19.35 | 34.33 | 25.26% |
|             | GraphSAGE | 12.14 | 19.87 | 15.69% | 14.95 | 24.01 | 18.54% | 18.68 | 33.59 | 24.77% |
| Los Angeles Expressway | TSGCN     | 11.64 | 18.52 | 14.77% | 14.29 | 23.56 | 17.23% | 17.47 | 32.83 | 23.23% |
|             | HA  | 4.04 | 7.76 | 12.87% | 4.03 | 7.37 | 12.56% | 4.01 | 7.59 | 12.64% |
|             | ARIMA     | 3.96 | 8.17 | 9.49% | 5.03 | 10.33 | 12.45% | 6.63 | 13.09 | 17.19% |
|             | SVR       | 3.91 | 8.38 | 9.26% | 4.93 | 10.76 | 12.01% | 6.54 | 13.58 | 16.55% |
|             | PNN       | 3.87 | 7.83 | 9.75% | 4.17 | 8.02 | 12.74% | 4.35 | 8.59 | 13.73% |
|             | GCN       | 3.76 | 7.74 | 9.63% | 4.01 | 7.98 | 12.65% | 4.31 | 8.44 | 13.68% |
|             | GAT       | 3.65 | 7.46 | 9.61% | 3.97 | 7.83 | 12.11% | 4.23 | 8.35 | 13.19% |
|             | GraphSAGE | 3.61 | 7.25 | 9.53% | 3.77 | 7.68 | 11.34% | 3.96 | 7.84 | 12.36% |
|             | TSGCN     | 3.57 | 7.16 | 9.47% | 3.67 | 7.56 | 10.45% | 3.84 | 7.47 | 11.73% |

4) GAT [58] & MODIFIED-GAT [59]
To overcome the drawbacks of earlier techniques based on graph convolutions or their approximations, GAT is a unique neural network architecture that operates on graph-structured data. Modified-GAT can create new graph topologies by finding valuable connections between disconnected nodes on the original graph and learning efficient node representation on the new graphs end-to-end.

5) CCM [60] & GAM [61]
A unique open-domain conversation generation model called CCM is used to show how widespread commonsense knowledge can help with language comprehension and creation. A global attention mechanism called GAM increases the global interactive representations and reduces information reduction to improve the performance of deep neural networks.

6) ATTENTIONWALKS [62] & GRAPH2SEQ [63]
A conversational reasoning model called AttentionWalks carefully moves across a vast knowledge graph of common facts to present exciting and contextually varied entities and attributes. GraphSAGE is a generic, inductive framework that effectively generates node embeddings by utilizing node feature information. An attention-based LSTM approach decodes the target sequence from the vectors in the unique generic end-to-end graph-to-sequence neural encoder-decoder model known as Graph2Seq. This model maps an input graph to a sequence of vectors.

### V. RESULT AND ANALYSIS
#### A. COMPARED WITH BASELINES (RQ1)
Table 1 displays the experiment’s findings. The historical average model is predicting accuracy is the same for both short- and long-term forecasts since it uses historical averages, which have a relatively stable baseline. However, the ARIMA, a standard time-continuity forecasting model, outperforms the historical average for short-term pre-patch. However, the trend and one-dimensional amplification of its relatively important trend information cause its long-term forecasting to be less accurate than the historical average. Because nonlinear elements are considered, the SVR and FNN are more sophisticated than ARIMA and can produce more accurate forecasts than ARIMA. Because they account for nonlinearities, the SVR and FNN are more sophisticated than ARIMA models and can produce superior results. However, because they do not consider the time features, they are also less accurate than historical averages in terms of the implications of long-term predictions. The historical averaging model, which solely takes into account the static Spatio-temporal features without the dynamic short-term traffic flow characteristics, is, in essence, the antithesis of the other models.

In comparison, the other graphical neural network models merely take into account the traffic flow without considering the spatiotemporal properties. However, information regarding the need for traffic flow is contained in the static Spatio-temporal features. By combining these two, we may create a joint distribution based on static and dynamic traffic flow aspects. These models, which include GCN, GAT, and GraphSAGE, can be outperformed when combined with a graph neural network. The last row of the table demonstrates how the TSGCN spatial graph neural network, which effectively utilizes the node properties of the spatial graph, may beat the earlier models without message forwarding.

#### B. INTRINSIC CORRELATION ABLATION EXPERIMENTS (RQ2)
How well does TSGCN utilize the attributes of the spatial graph nodes? Through model sharing, this study’s spatial graph neural network aspires to eigencorrelation learning of upstream data from distinct spatial graph nodes in a common feature space. How can this be proven to learn...
In this paper’s spatial graph neural network, three original features of time, space, and traffic flow are input to the graph nodes, and the correlation is precisely about the association of these original data with each other. Suppose the three features are to be retained while avoiding feature information sharing. In that case, it is necessary to decouple the sharing modules for each graph node’s temporal, spatial, and traffic flow features separately, thus avoiding sharing of feature space. For temporal features, each spatial graph node will no longer share the same temporal features. Each graph node will enjoy an independent temporal embedding representation: for spatial features, since we are using a spatial graph neural network, the spatial features are initially shared by each graph node, so what we need to do is to make each graph node no longer share the model, which means that without using the graph neural network, we can remove the spatial feature representation and let the features obtained by fusing time and traffic flow pass through a perceptron, respectively, for traffic flow prediction. For the traffic features, since the historical traffic features are deterministic data constants, we let the traffic feature fusion perceptrons of each model be independent of each other, thus allowing each space to learn based on the characteristics of its respective traffic flow features. As a result, we can suggest the following three TSGCN derivative models.

- **TSGCN-ns**: Delete spatial features.
- **TSGCN-nt**: Delete both moment and day characteristics.
- **TSGCN-nf**: Delete traffic flow characteristics.
- **TSGCN-nn**: Delete the null representation and fill the null with the mean value.
- **TSGCN-time**: Exclusive time embedding for each space.
- **TSGCN-flow**: The traffic characteristics of each space are fused and replaced by independent perceptrons.
- **FNN-time**: The spatial features are removed, the FNN is exclusive to each space, and the input is a fusion of temporal and traffic flow features.

We obtained various TSGCN models for prediction variations in this method. Figure 2 depicts the model’s consequences. We start by examining the usefulness of each feature. Due to the loss of each space’s spatial embedding representation, which makes it impossible to match the short-term traffic flow of all spaces with the same model, the influence of the TSGCN-ns with spatial features is considerably reduced when they are eliminated. In order to accurately predict the traffic flow on road networks, spatial features are essential. The prediction accuracy of the TSGCN-nt without temporal features does not considerably decline. The correlation between temporal attributes and traffic flow trend information and the fact that some information about time is contained in the traffic trend information account for this. For instance, when traffic is heavy, it usually happens during the morning rush hour, and when it is light, it usually happens late at night or early in the morning.

Traffic flow and temporal variables can be represented together if available, which naturally aids the model in producing more accurate forecasts. The effect of TSGCN-nt in the long-term prediction drops dramatically, which is missing information in the current traffic flow trend. The TSGCN-nn model can demonstrate the importance of null embedding, and its accuracy is increased when compared to a representation without null values. We now proceed to the eigencorrelation validation results. It is preferable to provide model parameters than not to. The model used to calculate the eigenvalue of spatial graph nodes is one that was trained by learning alongside other graph nodes. As a result, the eigenvalue information of other graph nodes is indirectly received, and the eigenrelevance is built.

### C. COMPARISON OF ATTENTION-RELATED MODELS (RQ3)

This subsection offers a cross-sectional comparison of several models, with the primary objective being to confirm the viability of graph attention networks for simulating the flow of vehicular traffic. The correlation issue of road network traffic flows can be resolved, as indicated in the preceding part, using graph attention networks to represent road traffic flows. In particular, it is the learning of attention mechanisms for interactive correlations and the learning of graph neural networks for intrinsic correlations. The secret to modeling such data is effective association modeling.

We start by examining the outcomes of the Los Angeles Expressway traffic rate prediction, as indicated in the
TABLE 2. Results of comparison of attention-related models.

| Dataset          | Model    | MAE       | RMSE     | [M/W]/APE | MAE       | RMSE     | [M/W]/APE | MAE       | RMSE     | [M/W]/APE |
|------------------|----------|-----------|----------|------------|-----------|----------|------------|-----------|----------|------------|
| Beijing Traffic Flow | AGNN     | 11.86     | 20.38    | 15.77%     | 13.46     | 20.28    | 15.34%     | 12.53     | 23.73    | 15.43%     |
|                  | EAGCN    | 12.49     | 18.45    | 14.68%     | 12.17     | 19.47    | 14.28%     | 13.66     | 22.51    | 14.58%     |
|                  | Modified-GAT | 10.55   | 17.41    | 13.93%     | 11.23     | 19.29    | 13.64%     | 11.27     | 21.74    | 13.77%     |
|                  | CCM      | 13.47     | 19.26    | 14.84%     | 12.79     | 21.56    | 14.75%     | 11.96     | 22.76    | 15.11%     |
|                  | GAM      | 11.23     | 17.85    | 13.62%     | 12.26     | 20.61    | 15.10%     | 12.43     | 22.35    | 14.39%     |
|                  | AttentionWalks | 10.72   | 16.97    | 13.97%     | 11.58     | 19.66    | 13.67%     | 11.42     | 21.28    | 13.46%     |
|                  | Graph2Seq | 10.72     | 17.17    | 13.37%     | 10.75     | 18.87    | 13.59%     | 11.28     | 20.98    | 13.20%     |
| Los Angeles Traffic Flow | AGNN     | 9.65      | 16.38    | 12.57%     | 10.03     | 18.42    | 12.80%     | 10.91     | 20.36    | 12.17%     |
|                  | EAGCN    | 4.12      | 6.63     | 6.73%      | 4.07      | 6.74     | 8.94%      | 3.77      | 8.04     | 9.87%      |
|                  | Modified-GAT | 3.54   | 5.81     | 5.19%      | 3.69      | 6.39     | 8.71%      | 3.86      | 7.84     | 9.64%      |
|                  | CCM      | 3.21      | 5.69     | 6.45%      | 3.15      | 6.15     | 7.93%      | 3.60      | 7.65     | 9.93%      |
|                  | GAM      | 2.76      | 6.14     | 6.27%      | 2.90      | 6.03     | 7.58%      | 3.42      | 7.49     | 9.87%      |
|                  | AttentionWalks | 2.54     | 5.37     | 5.85%      | 3.13      | 5.94     | 7.36%      | 3.25      | 7.44     | 9.26%      |
|                  | Graph2Seq | 2.78      | 5.24     | 5.96%      | 2.97      | 5.66     | 7.41%      | 3.26      | 7.05     | 8.97%      |
|                  | RTS-GAT  | 2.37      | 4.97     | 5.49%      | 2.68      | 5.34     | 7.03%      | 2.95      | 6.66     | 8.75%      |

Tables 1 and 2. First, we observe that RTS-GAT has a higher overall prediction accuracy than TSGCN, demonstrating the benefit of the attention mechanism for correlation modeling. Furthermore, TSGCN performs admirably in terms of long-term prediction while being relatively ineffective in terms of short-term prediction. RTS-GAT continues to outperform TSGCN and other top-notch models, including AGNN, Modified-GAT, GAM, and others that use the attention mechanism. It is a better option because the computational complexity is significantly lowered simultaneously.

So far, we have shown how well graph attention networks simulate traffic flow on roads. The multi-headed dot product attention mechanism is added to the model further to simulate the interaction correlation of road traffic flow. This model is based on the spatial graph neural network for simulating intrinsic correlation. For traffic flow and traffic velocity statistics, the spatial graph attention network RTS-GAT performs exceptionally well, illuminating the potency of the attention mechanism for modeling interaction correlation. In conclusion, the graph attention network offers a novel approach to the problem of road traffic modeling by analyzing spatiotemporal sequences of traffic flows in a road network from the viewpoint of graph data.

VI. CONCLUSION

The following is a condensed summary of the work of this paper. We summarize earlier research efforts and examine this paper’s main and challenging sections on road traffic flow modeling. Then, suggest RTS-GAT, a spatial graph attention network focused on interactive relevance modeling, and TSGCN, a spatial graph neural network for eigenrelevance modeling. Depending on the complexity and problems of the issue, we offer an attention mechanism based on the TSGCN model.

From the perspective of future exploratory research efforts, we can further enhance the present research results’ interpretability, performance, and short-term prediction capabilities, primarily from three elements of interpretability research, smart city multi-source data fusion, and short-time traffic flow prediction. According to interpretability research, an accurate short-term traffic flow prediction will impact decisions made by departments responsible for traffic management, the scheduling of traffic police personnel, the frequency of traffic lights, etc. A model that offers explanations will have several advantages, including establishing confidence, identifying the reason for traffic issues, assisting in model adjustment, etc.

We will put forward better models which is flexible enough to accommodate various traffic flow datasets from the viewpoint of intelligent city multi-source data fusion. Short-term traffic flow data is high-dimensional, complex, and challenging to manage from the perspective of short-term traffic flow forecasts. Nevertheless, there is no denying the need to simulate short-term OD traffic flow. The traffic flow has a higher information content than the conventional point traffic flow, which demands the model’s prediction abilities more.

REFERENCES

[1] H. Yi, H. Jung, and S. Bae, “Deep neural networks for traffic flow prediction,” in Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp), Feb. 2017, pp. 328–331.
[2] K. Niu, H. Zhang, C. Cheng, C. Wang, and T. Zhou, “A novel spatio-temporal model for city-scale traffic speed prediction,” IEEE Access, vol. 7, pp. 30050–30057, 2019.
[3] A. B. Parsa, R. S. Chauhan, H. Taghipour, S. Derrible, and A. Mohammadian, “Applying deep learning to detect traffic accidents in real time using spatiotemporal sequential data,” 2019, arXiv:1912.06991.
[4] Z. Xie, W. Lv, S. Huang, Z. Lu, B. Du, and R. Huang, “Sequential graph neural network for urban road traffic speed prediction,” IEEE Access, vol. 8, pp. 63349–63358, 2019.
[5] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, “Improved deep hybrid networks for urban traffic flow prediction using trajectory data,” IEEE Access, vol. 6, pp. 31820–31827, 2018.
[6] T. Kim, Y. Yue, S. Taylor, and I. Matthews, “A decision tree framework for spatiotemporal sequence prediction,” in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2015, pp. 577–586.
[7] T. Xiong, J. He, H. Wang, X. Tang, Z. Shi, and Q. Zeng, “Contextual sa-attention convolutional LSTM for precipitation nowcasting: A spatiotemporal sequence forecasting view,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 12479–12491, 2021.
[8] X. Shi and D.-Y. Yeung, “Machine learning for spatiotemporal sequence forecasting: A survey,” 2018, arXiv:1808.06865.

[9] J. Skarding, B. Gabrys, and K. Musial, “Foundations and modeling of dynamic networks using dynamic graph neural networks: A survey,” IEEE Access, vol. 9, pp. 79143–79168, 2021.

[10] Q. Fu, Y. Ji, and H. H. Huang, “LPGGN: A lightweight two-level parallelism paradigm for graph neural network computation on GPU,” in Proc. 31st Int. Symp. High-Perform. Parallel Distrib. Comput., Jun. 2022, pp. 122–134.

[11] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” Neurocomputing, vol. 452, pp. 48–62, Mar. 2021.

[12] D. Roy, A. Gupta, and R. B. M. De Koster, “A non-linear traffic flow-based queuing model to estimate container terminal throughput with AGVs,” Int. J. Prod. Res., vol. 54, no. 2, pp. 472–493, Jan. 2016.

[13] L. Liu, J. Zhen, G. Li, G. Zhan, Z. He, B. Du, and L. Lin, “Dynamic spatial–temporal representation learning for traffic flow prediction,” IEEE Trans. Intell. Transp. Syst., vol. 22, no. 11, pp. 7169–7183, Nov. 2021.

[14] W. Cao, Z. Yan, Z. He, and Z. He, “A comprehensive survey on geometric deep learning,” IEEE Access, vol. 8, pp. 35929–35949, 2020.

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

[16] L. Mourad, H. Qi, Y. Shen, and B. Yin, “ASTIR: Spatio-temporal data mining for crowd flow prediction,” IEEE Access, vol. 7, pp. 175159–175165, 2019.

[17] B. Sun, T. Sun, and P. Jiao, “Spatio-temporal segmented traffic flow prediction with ANPRs data based on improved XGBoost,” J. Adv. Transp., vol. 2021, pp. 1–24, May 2021.

[18] Y. Hu, W. Luo, Z. Yu, L. Yuan, and G. Lu, “Geometric algebra-based modeling and analysis for multi-layer, multi-temporal geographic data,” Adv. Appl. Clifford Algebras, vol. 26, no. 1, pp. 151–168, Mar. 2016.

[19] B. Ran, H. Tan, Y. Wu, and P. J. Jin, “Tensor based missing traffic data completion with spatio–temporal correlation,” Phys. A, Stat. Mech. Appl., vol. 446, pp. 54–63, Mar. 2016.

[20] V. Vzyryanov, V. Kocherga, and I. Topilin, “Investigation of dependences between parameters of two-component models of the kinetic theory of traffic flow and traffic characteristics,” Transp. Res. Proc., vol. 20, pp. 746–750, Jan. 2017.

[21] A. Jayasinghe, K. Sano, C. C. Abenayake, and P. K. S. Mahanama, “A novel approach to model traffic on road segments of large-scale urban road networks,” MethodsX, vol. 6, pp. 1147–1163, Jan. 2019.

[22] D. Gai, C. Ma, and X. Xu, “Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU,” IEEE Access, vol. 7, pp. 143025–143035, 2019.

[23] Y. Zhang, X. Zheng, M. Chen, Y. Li, Y. Yan, and P. Wang, “Urban fine-grained spatial structure detection based on a new traffic flow interaction analysis framework,” ISPRS Int. J. Geo-Inf., vol. 10, no. 4, p. 227, Apr. 2021.

[24] K. Liu, S. Gao, and F. Lu, “Identifying spatial interaction patterns of vehicle movements on urban road networks by topic modelling,” Comput. Environ. Urban Syst., vol. 74, pp. 50–61, Mar. 2019.

[25] M. Serra and B. Hilier, “Angular and metric distance in road network analysis: A nationwide correlation study,” Comput. Environ. Urban Syst., vol. 74, pp. 194–207, Mar. 2019.

[26] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in Proc. AAAI Conf. Artif. Intell., 2021, vol. 35, no. 12, pp. 11106–11115.

[27] 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, 2018.

[28] G. Athuri, A. Karpatne, and V. Kumar, “Spatio-temporal data mining: A survey of problems and methods,” ACM Comput. Surveys, vol. 51, no. 4, pp. 1–41, Jul. 2018.

[29] S. Wang, J. Cao, and P. S. Yu, “Deep learning for spatio-temporal data mining: A survey,” IEEE Trans. Knowl. Data Eng., vol. 34, no. 8, pp. 3681–3700, Aug. 2022.

[30] W. Zaremba, I. Sutskever, and O. Vinyals, “Recurrent neural network regularization,” 2014, arXiv:1409.2329.

[31] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” Phys. D, Nonlinear Phenom., vol. 404, Mar. 2020, Art. no. 125306.
SHIYANG LUO (Member, IEEE) received the B.S. degree in computer science and technology from the Dalian Neusoft University of Information, China, in 2018. He is currently pursuing the Ph.D. degree.

Since 2018, he has been the Managing Director/General Manager/CEO at Shanghai Infortech Software Development Company Ltd., which is a high-tech enterprise centered on Chinese mainland, providing IT solutions and services to Southeast Asia. Founded, in 1998, the company is headquartered in Shanghai. In 2012, for the rapid development of the company, it formally introduced the investment of Japanese joint-stock company NTT DATA, one of the global 500 companies, to break into international market more quickly. He is also a Visiting Professor at the Institute of Business Innovation, China Business Economics Association. He is the author of four articles that are published on international journals, and he patented more than six inventions. His research interests include data science, network security, artificial intelligence, machine learning, retail industry big data, and enterprise management.

Mr. Luo is a member of the IEEE Computer Society and a member of the Asia-Pacific Artificial Intelligence Association. He was a recipient of the Gold Gull Award 2021 Innovative Personage. He was elected as the Director of the Seventh Council of China Computer Industry Association, in 2021.

* * *