Attention Based Spatial-Temporal Graph Convolutional Networks for RSU Communication Load Forecasting

Hang Zheng[1], Xu Ding[2], Yang Wang[1], and Chong Zhao[3](✉)

[1] School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China
[2] Institute of Industry and Equipment Technology, Hefei University of Technology, Hefei 230009, China
[3] Engineering Quality Education Center of Undergraduate School, Hefei University of Technology, Hefei 230009, China
zhaochong@hfut.edu.cn

Abstract. As a special type of base station, Road Side Units (RSU) can be deployed at low cost and effectively alleviate the communication burden of regional Vehicular Ad-hoc Networks (VANETs). However, because of peak hour communication demands in VANETs and limited energy storage, it is necessary for RSU to adjust their participation in communication according to the requirements and allocate energy reasonably to balance the workload. Firstly, tidal traffic flow is generated according to the information of morning and evening peak in the city, so as to simulate the vehicle distribution around RSU on urban roads. Secondly, by inputting the historical information around RSU and the topological relationship between RSU, a network load prediction model is established by using the Attention based Spatial-Temporal Graph Convolutional Networks (ASTGCN) to predict the future communication load around RSU. Finally, according to the forecast of the future communication load, a RSU working mode alteration scheme is proposed with respect to the safety range amongst vehicles in order to control the corresponding area communication load. Compared with other models, our model has better accuracy and performance.

Keywords: VANETs · RSU · Communication load forecasting · ASTGCN · Energy schedule

1 Introduction

Due to the ability of extending the horizon of drivers which will improve road traffic safety, VANETs are attracting an extensive attention from both academia

Supported by the Fundamental Research Funds for the Central Universities (GRANT NO. PA2021GDSK0095).

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 99–114, 2021.
https://doi.org/10.1007/978-3-030-92635-9_7

and industry [1,2]. Vehicles in VANETs act as communication nodes, exchanging safety-related information such as location, speed, and brake status with surrounding ones, base stations and infrastructures by means of on-board devices. As an important part of VANETs, Road Side Unit (RSU) provide significant support for regional vehicle communication [3]. TRSUs are with fixed locations and could improve communication speed and reduce the delay in VANETs [4,5].

These studies show that RSU can play an important role in improving the communication quality of VANET. However, it is not necessary to keep the RSU at full power. Under the circumstance of ensuring low communication delay in this area, the online algorithm is used to dynamically adjust the working state of the RSU to arrange the energy consumption wisely. Since the communication delay is closely related to the communication load, we predict the communication load of the RSU for a period of time in the future to assess whether the VANET in the corresponding area is in a safe state and decide whether to continue using the RSU.

With RSU usually being deployed at the intersection of urban roads, the communication load of RSU is closely related to the vehicle information of road interchanges that can be obtained from the urban intelligent transportation system [6], such as the number of vehicles in the area, the occupancy rate of lanes and other factors. Vehicle data are sampled every fixed time interval at certain locations along the road successivley. Apparently, the observed data is not independent but dynamically related to surrounding and historical data. And the correlation of vehicle data on urban roads possesses dynamic features in both spatial and temporal dimensions. It becomes a very challenging problem that exploring nonlinear and complex spatiotemporal data to discover inherent spatiotemporal patterns and make accurate communication loads prediction.

With the development of the VANETs, many cameras, sensors and other information collection devices have been deployed on urban roads. Each device is placed at a unique geospatial location, constantly generating time series data about traffic. These devices store rich traffic time series data with geographic information, providing a solid data foundation for communication load forecasting.

On this basis, we use a deep learning model: Attention based Spatial-Temporal Graph Convolution Network (ASTGCN) to predict communication load at every location on the traffic network. This model can process the traffic data directly on the original graph-based traffic network and effectively capture the dynamic spatial-temporal features. Through the effective evaluation of the future network load, a certain safety threshold is set to determine the working status of the RSU to reduce energy consumption. The main contributions of this paper are listed as follows:

- Attention based Spatial-Temporal Graph Convolution Network has been improved so that it can predict the RSU communication load on each node through the vehicle data on the graph-based transportation network.
- An RSU energy decision algorithm based on future communication load evaluation is proposed to enable RSU to reduce energy consumption.

The remainder of this paper is organized as follows. The related works are introduced in Sect. 2. The system model is presented in Sect. 3. The communication load evaluation is introduced in Sect. 4. The communication load prediction model are introduced in Sect. 5. Simulation and analysis are introduced in Section 6 and a conclusion is made in Section 7.

2 Related Works

At present, in many studies, researchers altered the working modes of the RSU to reduce the global energy consumption. In [7], Wen *et al.* proposed the RSU could determine to work or not depending on the current communication situation. If there was any vehicle in the network communicating with others, the RSU kept service time, otherwise stopped working. However, the non-differentiated service of RSU without considering network load made the performance of RSU unstable. In [8], Patra *et al.* proposed an algorithm taking into account the coverage of RSU and the overall energy consumption. The goal of the algorithm was to minimize the energy consumption rate.

Reyhanian *et al.* proposed two novel online approaches for enabling energy trading in multitier cellular networks with non-cooperative energy-harvesting base stations (BSs) to minimize the nonrenewable energy consumption in a multitier cellular network [9]. Zhang *et al.* proposed a resource-allocation solution for the heterogeneous CR sensor networks to achieve the sustainability of spectrum sensors and conserve the energy of data sensors. Extensive simulation results demonstrate that the energy consumption of the data sensors can be significantly reduced, while maintaining the sustainability of the spectrum sensors [10].

These studies focused on the global energy consumption allocation problem of RSU. However, the main reason of short-term and inconsistent service provided by RSU is the imbalanced network communication demand brought by the tidal fluctuation of traffic [11]. If the communication load can be predicted, the RSU would be able to adjust energy allocation in advance and be adapt to the network's tidal communication and the stability of the RSU service would be guaranteed.

The prediction of communication load is a spatial-temporal data prediction problem that many researchers have used deep learning to solve in recent years. Especially in the field of traffic flow, many deep learning models have been proposed to predict traffic flow. In [12], Liang et al. designed a ST-ResNet model based on the residual convolution unit to predict crowd flows. In [13], Yao et al. proposed a method to predict traffic by integrating CNN and long-short term memory (LSTM) to jointly model both spatial and temporal dependencies. Although the spatial-temporal features of the traffic data can be extracted by these model, their limitation is that the input must be standard 2D or 3D grid data. In [14], Guo *et al.* designed a novel spatial-temporal convolution module for modeling spatial-temporal dependencies of traffic data. It consists of graph convolutions for capturing spatial features from the original graph-based traffic network structure and convolutions in the temporal dimension for describing dependencies from nearby time slices.

Motivated by the studies mentioned above, considering the graph structure of the traffic network and the dynamic spatial-temporal patterns of the traffic data, we employ Attention Based Spatial-Temporal Graph Convolutional Networks to predict the communication load. Then specify the energy plan based on the network load forecast so that the RSU can obtain a longer effective working time.

3 System Model

In our research scenario, RSU is arranged at each intersections, as shown in Fig. 1. RSU is equipped with multiple antennas, which can receive signals sent by multiple vehicles at the same time.

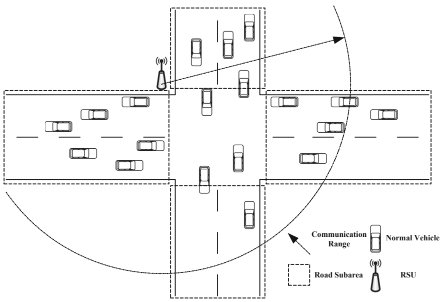

Fig. 1. Typical RSU scenario in VANETs

As the uneven distribution of vehicles have an unbalanced impact on communication requirements, the coverage of each RSU area is divided into several subareas, such as road section, connection points and entrances. The scheduling time of RSU is divided into K time frames, and the traffic flow, lane occupancy rate and number of vehicles in each area are counted every other time frame. D_i^t represents the vehicle information collected by the i^{th} RSU in the t^{th} time range. The vehicle information around RSU includes the number of vehicles in various surrounding areas, traffic flow, and traffic occupancy rate. Limited by length of the road, there is an upper limit for the number of vehicles in each subarea.

The vehicles and RSU can only initiate communication during the time slot assigned to it. Since the length of time frame is very short, during the same time frame, the relative positions among vehicles can be considered unchanged. The RSU is equipped with multiple antennas that can forward messages for multiple vehicles at the same time. But the vehicle can only communicate with only one vehicle in a time slot. Both vehicles and RSU use half-duplex mode which means they cannot transmit or receive messages simultaneously in one time slot.

In this study, RSU is deployed at each node of the entire road network. The collected information of the surrounding vehicles of each RSU is continuous in time, and the information corresponding to the time is used as a node to form a graph, as shown in Fig. 2.

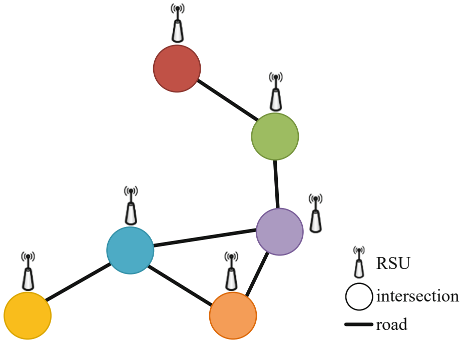

Fig. 2. The spatial structure of traffic data

The straight lines in the figure represent the roads in the city, the circles represent the nodes where the roads intersect, and the RSU is deployed around each intersection. We define a traffic network as an undirected graph $G = (V, E, A)$, where V is a finite set of $|V| = N$ nodes; E is a set of edges, indicating the connectivity between the nodes; $A \in R^{N \times N}$ denotes the adjacency matrix of graph G.

Symbols used in the following discussion are listed in Table 1.

Table 1. Notation table

Symbol	Definition
N	The total number of vehicles on the node
D_i^t	Vehicle information of the t^{th} time segment around the i^{th} RSU
$c_{i,j}^t$	$c_{i,j}^t$ is binary value indicating that vehicle i and vehicle j need to complete an information exchange during t^{th} time frame
$r_{i,j}^t$	$r_{i,j}^t$ represents the distance between vehicle i and vehicle j
R_d, R_v	Communication radius of RSU and vehicles
$D_{i,j}$	Routing path of v_i sends message to v_j
$D_{i,j}(k,l)$	Indicates that v_k sends message to v_l as routing nodes for v_i and v_j information exchange
$C_{i,j}$	The amount of data sent by v_i to v_j during the t^{th} time frame
$R_{i,j}^t$	$R_{i,j}^t$ is the communication speed between v_i and v_j
Y_v^t	Communication load at t^{th} time without RSU
Y_s^t	Communication load at t^{th} time with RSU

4 Communication Load Evaluation

In this section, we will introduce how to evaluate the network load and provide a method for obtaining training data sets based on historical data.

In VANETs, the most important communication content is safety-related information, which has higher requirements for network delay. Therefore, the length of the time slot required for all vehicles to complete this type of message interaction is used as the communication load standard. According to the requirements of VANETs, vehicles must initiate an information interaction with others within safety distance during every time frame. C is the set of communication sessions that element $c_{i,j}^t$ is binary value indicating that v_i and v_j need to complete an information exchange during t^{th} time frame,

$$C = \left\{c_{i,j}^t\right\}, (r_{i,j}^t \leq R_v) \tag{1}$$

Vehicle communicates with each other in a multi-hop fashion. Let $D_{i,j}$ denote the routing path and $D_{i,j}(k,l) \in D_{i,j}$ denote that v_k and v_l take part in the data forwarding task sent from v_i to v_j. Therefore, the amount of data sent by v_i is equal to the amount of data received by v_j, i.e.,

$$\sum_{k=1}^{N} D_{i,j}(i,k) = \sum_{k=1}^{N} D_{i,j}(k,j), (r_{i,k}^t \leq R_v, r_{j,k}^t \leq R_v, c_{i,j}^t = 1) \tag{2}$$

Let $F_{i,j}^t$ denote the amount of data sent by v_i to v_j during the t^{th} time frame. $F_{i,j}^t$ could be divided into relayed data and directly transmitted data, and we have:

$$\begin{aligned} C_{i,j}^t = & \sum_{k=1(k\neq i)}^{N} \sum_{l=1(l\neq k)}^{N} D_{k,l}^t(i,j) + \sum_{l=1}^{N} D_{i,j}^t(i,l), \\ & (r_{k,l}^t \leq R_v, r_{i,l}^t \leq R_v, r_{i,j}^t \leq R_v) \end{aligned} \tag{3}$$

If v_i could communicate with v_j directly during t^{th} time frame, the communicate rate can be obtained as following

$$R_{i,j}^t = W \cdot \log_2(1 + \frac{P_i}{N_0}\delta(r_{i,j}^t)^{-\gamma}\left|h_{i,j}^t\right|^2), (r_{i,j}^t \leq R_v) \tag{4}$$

$R_{i,j}^t$ is the communication speed between v_i and v_j, W is bandwidth and $h_{i,j}^t$ is the Rayleigh distributed fading magnitude with $E\left[\left|h_{i,j}^t\right|^2\right] = 1$. N_0 is the power of additive white Gaussian noise (AWGN). δ is the log-normal shadowing component, with a mean of 0 dB. γ is typically chosen as the path fading exponent for VANETs.

In order to have a unified evaluation standard for the network load carried by the regional VANETs, the required time slots for communication is used as a metric for the communication load represented by Y_v^t, then we have

$$Y_v^t = \min \frac{\sum_{i=1}^{N} \sum_{j=1(i\neq j)}^{N} \frac{C_{i,j}^t}{R_{i,j}^t}}{N} \tag{5}$$

The RSU covers the entire network and the main job it participates in V2V communication is forwarding data for vehicles. Y_s^t is the number of time slots required when RSU participates in communication during t^{th} time frame, which can be calculated as

$$Y_s^t = \kappa(N_a) \cdot Y_v^t \tag{6}$$

κ is the EH-RSH improvement function, related to the number of antennas in RSU denoted by N_a, meaning the RSU can provide service for N_a vehicles simultaneous.

5 The Communication Load Prediction Model

In this section, we will introduce how to make a prediction of the communication load and explain the various modules of the Attention Based Spatial-Temporal Graph Convolutional Networks(ASTGCN) model in detail.

Suppose the f^{th} time series recorded on each node in the traffic network G is the traffic flow sequence, and $f \in (1, \ldots, F)$. We use $x_t^{c,i} \in R$ to denote the value of the c^{th} feature of node i at time t, and $x_t^i \in R^F$ denotes the values of all the faatures of node i at time t. $X_t = (x_t^1, x_t^2, \ldots, x_t^N)^T \in R^{N \times F}$ denotes the values of all the features of all the nodes at time t. $\chi \in (X_1, X_2, \ldots, X_\tau) \in R^{N \times F \times \tau}$ denotes the value of all the features of all the nodes over τ time slices.In addition, we set $y_t^i = x_t^{f,j} \in R$ to represent the communication load of RSU in the node i at time t in the future.

Given χ, all kinds of the historical measurements of all the nodes on the traffic network over past τ time slices, predict future communication load of RSU sequences $Y = (y^1, y^2, \ldots y^N)^T \in R^{N \times T_p}$ of all the nodes on the whole traffic network over the next T_p time slices, where $y^i = (y^i_{\tau+1}, y^i_{\tau+2}, \ldots, y^i_{\tau+T_p})$ denotes the future communication load of node i from $\tau + 1$. Each node on the traffic network G detects F measurements with the same sampling frequency, that is, each node generates a feature vector of length F at each time slice.

In order to transform this information into our predicted results, we improved the Attention Based Spatial-Temporal Graph Convolutional Networks model. Fig. 3 shows the overall framework of the ASTGCN model proposed in this paper. It is composed of five parts: time attention module, spatial attention module, spatial dimension graph convolution module, time dimension convolution module, and fully connected layer module.

Suppose the sampling frequency is q times per day. Assume that the current time is t_0 and the size of predicting window is T_p. We intercept the time series segment of length T_r along the time axis and use it as the most recent time slice input, where T_r is integer multiples of T_p.

$$\chi_r = (X_{t_0 - T_h + 1}, X_{t_0 - T_h + 2}, \ldots, X_{t_0}) \in R^{N \times F \times T_r} \tag{7}$$

A spatial-temporal attention mechanism is used in our model to capture the dynamic spatial and temporal correlations on the traffic network. It contains two kinds of attentions, spatial attention and temporal attention.

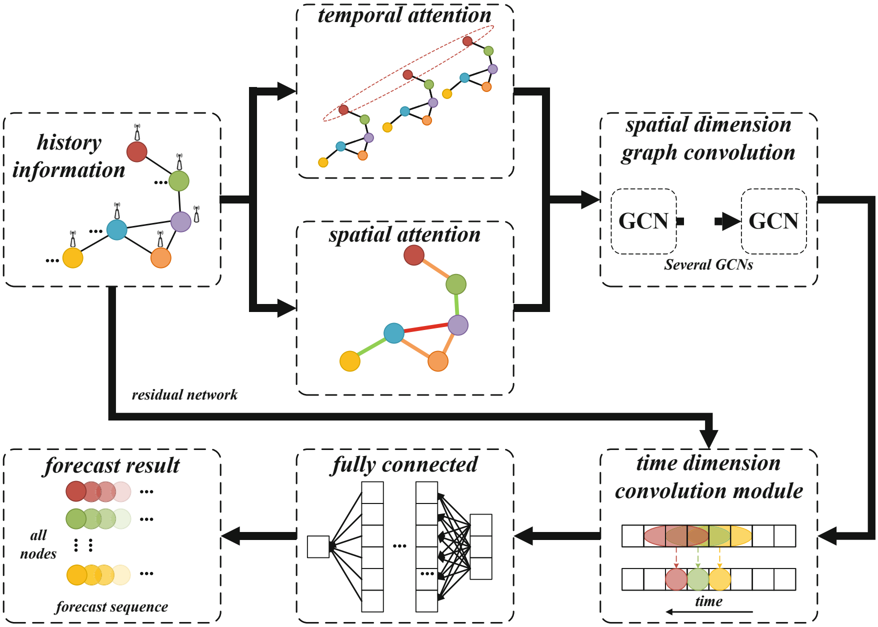

Fig. 3. The framework of ASTGCN

5.1 Temporal Attention

In the temporal dimension, there exist correlations between the traffic conditions in different time slices, and the correlations are also varying under different situations. Thus, we use an attention mechanism [15] to adaptively attach different importance to data:

$$E = V_e \cdot \sigma(((\chi_r^{(r-1)})^T U_1) U_2 (U_3 \chi_r^{(r-1)}) + b_e) \tag{8}$$

$$E'_{i,j} = \frac{exp(E_{i,j})}{\sum_{j=1}^{T_{r-1}} exp(E_{i,j})} \tag{9}$$

where $V_e, b_e \in R^{T_{r-1} \times T_{r-1}}, U_1 \in R^N, U_2 \in R^{C_{r-1} \times N}, U_3 \in R^{C_{r-1}}$ are learnable parameters. The temporal correlation matrix E is determined by the varying inputs. The value of an element $E_{i,j}$ in E semantically indicates the strength of dependencies between time i and j. At last, E is normalized by the softmax function. We directly apply the normalized temporal attention matrix to the input and get $\hat{\chi}_h^{(r-1)} = (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{T_r-1}) = (X_1, X_2, \ldots, X_{T_{r-1}})E' \in R^{N \times C_{r-1} \times T_{r-1}}$ to dynamically adjust the input by merging relevant information.

In the time dimension, as shown in the Fig. 4, the historical traffic data of different locations has different effects on the communication load of other nodes at different times in the future. We train this weight through the attention mechanism and convert it into a matrix $R^{N \times C_{r-1} \times T_{r-1}}$.

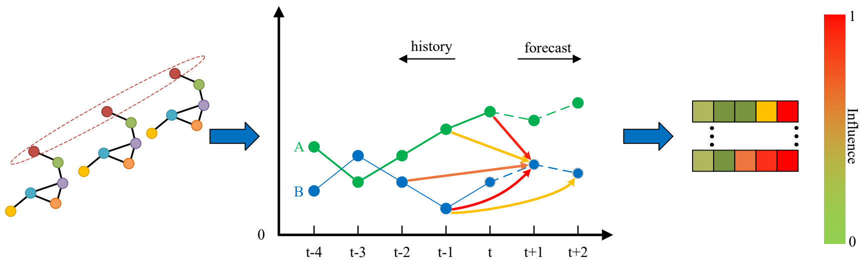

Fig. 4. Schematic diagram of temporal attention mechanism

5.2 Spatial Attention

In the spatial dimension, the traffic conditions of different locations have influence among each other and the mutual influence is highly dynamic. Here, we use an attention mechanism to adaptively capture the dynamic correlations between nodes in the spatial dimension.

$$S = V_s \cdot \sigma((\chi_r^{(r-1)} W_1) W_2 (W_3 \chi_r^{(r-1)})^T + b_s) \tag{10}$$

$$S'_{i,j} = \frac{exp(S_{i,j})}{\sum_{j=1}^{N} exp(S_{i,j})} \tag{11}$$

where $\chi_h^{(r-1)} = (X_1, X_2, \ldots, X_{T_r-1}) \in R^{N \times C_{r-1} \times T_{r-1}}$ denotes the input of the r^{th} spatial module. C_{r-1} is the number of channels of the input data in the r^{th} layer. T_{r-1} is the length of the sequence for forecasting, When $r = 1$, $C_0 = F$, $T_0 = T_r$. $V_s, b_S \in R^{N \times N}, W_1 \in R^{T_r - 1}, W_2 \in R^{C_{r-1} \times T_{r-1}}, W_3 \in R^{C_{r-1}}$ are learnable parameters and sigmoid σ is used as the activation function. The attention matric S is dynamically computed according to the current input of this layer.The value of an elements $S_{i,j}$ in S semantically represents the correlation strength between node i and node j. Then a softmax function is used to ensure the attention weights of a node sum to one.

As shown in Fig. 5, when performing the graph convolutions, we will accompany the adjacency matrix A with the spatial attention matrix $S' \in R^{N \times N}$ to dynamic adjust the impacting weights between nodes.

5.3 Graph Convolution in Spatial Dimension

In order to make full use of the topological properties of the traffic network, at each time slice we adopt graph convolutions based on the spectral graph theory to directly process the signals, exploiting signal correlations on the traffic network in the spatial dimension. The spectral method transforms a graph into an algebraic form to analyze the topological attributes of graph, such as

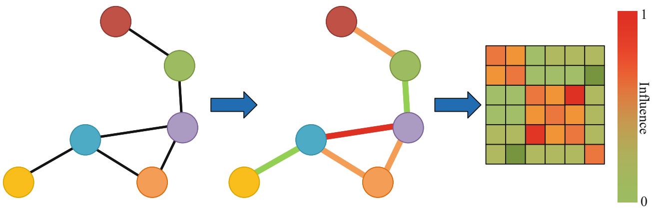

Fig. 5. Schematic diagram of spatial attention mechanism

the connectivity in the graph structure. In spectral graph analysis, a graph is represented by its corresponding Laplacian matrix, and its properties can be obtained by analyzing the Laplacian matrix and its eigenvalues. Therefore, we use Chebyshev convolution polynomial [16] to solve this problem.

$$g_\theta * Gx = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x \tag{12}$$

where $*G$ denotes a graph convolution operation, the parameter $\theta \in R^K$ is a vector of polynomial coefficients. $\tilde{L} = \frac{2}{\lambda_{max}} L - I_N$, λ_{max} is the maximum eigenvalue of the Laplacian matric, I_N is a unit matrix. The recursive definition of the Chebyshev polynomial is $T_k(x) = 2xT_{k*-1}(x) - T_{k-2}(x)$, where $T_0(x) = 1, T_1(x) = x$. Using approximate expansion of Chebyshev polynomial to solve this formulation corresponds to extracting information of the surrounding 0 to $(K-1)^{th} - order$ neighbors centered on each node in the graph by the convolution kernel g_θ.

In order to dynamically adjust the correlations between nodes, for each term of Chebyshev polynomial, we accompany $T_k(\tilde{L})$ with the spatial attention matrix $S' \in R^{N\times N}$ [14], then obtain $T_k(\tilde{L}) \odot S'$, where $\odot$ is the Hadamard product. Therefore, the above graph convolution formula changes to:Q

$$g_\theta * Gx = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) \odot S'x \tag{13}$$

We can generalize this definition to the graph signal with multiple channels. For example, in the recent component, the input is $\hat{\chi}_h^{r-1} = (\hat{X}_1, \hat{X}_2, \ldots, \hat{X_{T_{r-1}}}) \in R^{N\times C_{r-1}\times T_{r-1}}$, where the feature of each node has C_{r-1} channels. For each time slice t, performing C_r filters on the graph $\hat{X}_t$, we get $g_\theta * G\hat{X}_t$, where $\ominus = \{\ominus_1, \ominus_2, \ldots, \ominus_{C_r}\} \in R^{K\times C_{r-1}\times C_r}$ is the convolution kernel parameter. Therefore, each nodes is updated by the information of the $0 \sim K-1$ neighbors of the node, as shown in Fig. 6.

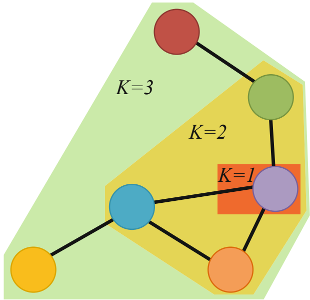

Fig. 6. The receptive field of different K values in Chebyshev convolution

5.4 Convolution in Temporal Dimension

After the graph convolution operations having captured neighboring information for each node on the graph in the spatial dimension, a standard convolution layer in the temporal dimension is further stacked to update the signal of a node by merging the information at the neighboring time slice. Take the operation on the r^{th} layer in the recent component as an example:

$$\chi_h^{(r)} = \Phi * (ReLU(g_\theta * G\hat{X}_h^{(r-1)})) \in R^{C_r \times N \times T_r} \tag{14}$$

where $*$ denotes a standard convolution operation, Φ is the parameter of the time dimension convolution kernel. The activation function uses ReLU to avoid gradient disappearance and gradient explosion problems in back propagation.

5.5 Fully Connected Layer

By maximizing the interaction between the extracted features and the input data, we can preserve the potential correlation between the extracted features and the input data. And in order to establish the relationship between the extracted features and the final output, we need to map them to the same dimension, for which we use a fully connected layer. At the same time, due to the large difference between the dimensions of the input data and the data that needs to be predicted, the introduction of the fully connected layer can accelerate the convergence of our model.

6 Simulation and Analysis

6.1 Dataset

In order to evaluate the performance of our model, we conducted comparative experiments on the data set obtained from the SUMO simulation. The road network simulated by SUMO uses the real road network in a certain area of Hefei, China, as shown in Fig. 7.

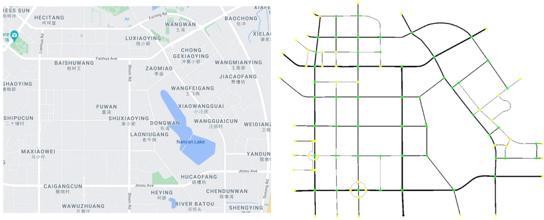

Fig. 7. Schematic diagram of simulated road network

The relevant information is collected every second and aggregated every 5 min. The traffic flow of the road nodes, the vehicle occupancy rate, the number of vehicles in the area and the communication load within the jurisdiction of each RSU. The entire road network contains 105 nodes and 282 edges, and 28 days of vehicle data are simulated (the daily traffic flow data is simulated by morning and evening peak hours). We extract 24-h related information of a certain RSU in the obtained information and draw it, as shown in Fig. 8.

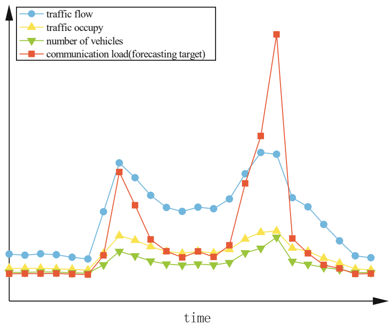

Fig. 8. Three measurements are detected on RSU and the future communication load is the forecasting target

We choose data on the first 23 days as the training set, and the remains as the test set. The communication radius of RSU is selected as 100m, and the safety distance between cars is selected as 50 m. The vehicle needs to transmit information to all vehicles within the safe distance for each calculation.

6.2 Model Parameters

We implemented the ASTGCN model based on the MXNet framework. We tested the number of terms of the Chebyshev polynomial K to changed the size of the receptive field of the graph convolution in the graph. In our model, all the graph convolution layers use 64 convolution kernels. All the temporal convolution layers use 64 convolution kernels and the time span of the data is ad justed by controlling the step size of the temporal convolutions. For the input historical information length, we set it as: the input recent vehicle information $T_h = 18$, the prediction window size $T_p = 6$. The input sequence is 3 times the length of the output sequence, which means that our goal is to predict the next half an hour communication load within. The mean square error is used to calculate the predicted value and the true value as the loss function, and minimized by the back propagation algorithm. In the training process, the batch size is selected as 16, the optimizer selects ADAM optimization, and the learning rate is 0.001.

6.3 Comparison and Result Analysis

Through our model, we predict the communication load of all RSUs in the road network during the next period of time. As shown in the Fig. 9, this is a 24-h communication load forecast formed by splicing the communication load predicted by a certain node.

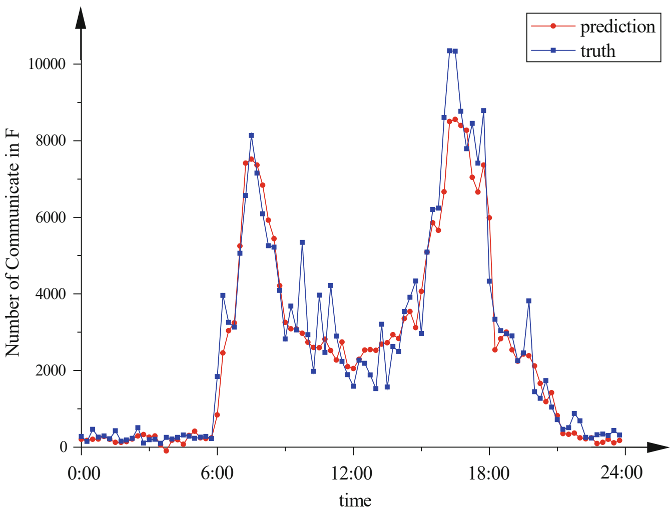

Fig. 9. Communication load estimation simulation

We compare our models with the following four baseline methods on our dataset:

- LSTM [17]: Long Short Term Memory network, a special RNN model.
- GRU [18]: Gated Recurrent Unit network, a special RNN model.
- STGCN [19]:A spatial-temporal graph convolution model based on the spatial method.
- MSTGCN [14]: A multi-layer spatial-temporal graph convolutional neural network

Table 2. Average performance comparison of different approaches on our dataset.

Model	MAE			RMSE			Validation time
	1	3	6	1	3	6	
LSTM	475.9	478.5	485.7	2176	2178	2185	3.128s
GRU	469.4	470.6	473.7	2174	2175	2177	3.070s
STGCN	286.8	295.1	301.8	1755	1763	1779	1.787s
MSTGCN	242.42	255.8	274.7	1499	1560	1627	6.719s
ASTGCN (ours)	**230.7**	**245.3**	**260.9**	**1366**	**1495**	**1589**	**6.632s**

Table 2 shows the average results of communication load prediction performance over the next half one hour. We respectively evaluate the next 1, 3, and 6 prediction sequences. Root mean square error (RMSE) and mean absolute error (MAE) are used as the evaluation metrics. At the same time, we also enumerate the validation time of each model to indicate the time consumed by different models when they are used. The results show that our improved ASTGCN model has better prediction accuracy for communication load.

We also did an ablation experiment to verify the role of each module, as shown in the Table 3 to reflect the role of each module. **Experiment 1** removes the attention mechanism module, **Experiment 2** removes the graph convolution module, and then also removes spatial attention, **Experiment 3** removes the final fully connected layer module. **Experiment 4** changed the convolution step size in the time dimension, **Experiment 5** has all modules,

Table 3. Comparison of average performance of ablation experiments

Experiment	MSE	RMSE
1	260.9	1514.0
2	251.19	1477.5
3	248.9	1505.1
4	242.4	1495.6
5	230.7	1366.0

After removing the attention mechanism module, the performance of the model decreases. In order to more intuitively reflect the role of the attention module, we extract the four points in the road network separately as a sub-image, and we output the trained attention matrix and convert it into a heat map to reflect the difference the correlation between nodes is strong or weak.

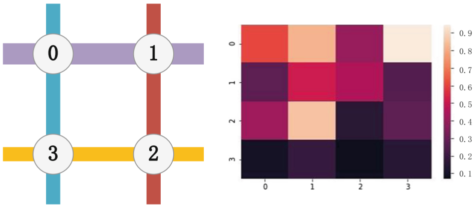

Fig. 10. The attention matrix obtained from the spatial attention mechanism

As shown in the Fig. 10, each row represents the strength of the correlation between the corresponding node and the rest of the nodes. It can be seen that each node has a relatively stronger correlation with itself. And it can be seen that there is a stronger focus on node 0, which is very reasonable, because from the traffic flow point of view, node 0 has more traffic and at the same time bears a greater communication load, which is easily congested. In addition, the relative degree of correlation between node 1 and node 3 is also low, because from the route point of view, there is no direct route connection between the two nodes, but is established through multi-hop Chebyshev graph convolution.

7 Conclusion

In this paper, we construct a traffic flow data set and use the Attention Based Spatial-Temporal Graph Convolutional network communication load estimation model to predict the network load state at the next moment, which is used to guide RSU to make appropriate actions for VANETs to improve the global network communication quality. Experimental simulation shows that model can accurately predict RSU communication load, dynamically adjust working modes, balance tasks between high-load and low-load networks and reduce RSU's energy consumption.

References

1. Hartenstein, H., Laberteaux, K.P.: A tutorial survey on vehicular ad hoc networks. IEEE Commun. Mag. **46**(6), 164–171 (2008)
2. Lee, E., Lee, E.K., Gerla, M., Oh, S.Y.: Vehicular cloud networking: architecture and design principles. IEEE Commun. Mag. **52**(2), 148–155 (2014)
3. Dragicevic, T., Lu, X., Vasquez, J.C., Guerrero, J.M.: Dc microgrids-part ii: a review of power architectures, applications and standardization issues. IEEE Trans. Power Electron. **31**(5), 1–1 (2015)
4. Barrachina, J., Garrido, P., Fogue, M., Martinez, F.: Road side unit deployment: a density-based approach. IEEE Intell. Transp. Syst. Mag. **5**(3), 30–39 (2013)
5. Gao, Z., Chen, D., Cai, S., Wu, H.C.: Optdynlim: an optimal algorithm for the one-dimensional RSU deployment problem with nonuniform profit density. IEEE Trans. Industr. Inf. **15**(2), 1052–1061 (2019)

6. Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C.: Data-driven intelligent transportation systems: a survey. IEEE Trans. Intell. Transp. Syst. **12**(4), 1624–1639 (2011)
7. Wen, C., Zheng, J.: An RSU on/off scheduling mechanism for energy efficiency in sparse vehicular networks. In: 2015 International Conference on Wireless Communications & Signal Processing (WCSP), pp. 1–5. IEEE (2015)
8. Patra, M., Murthy, C.S.R.: Performance evaluation of joint placement and sleep scheduling of grid-connected solar powered road side units in vehicular networks. IEEE Trans. Green Commun. Networking **2**(4), 1197–1209 (2018)
9. Reyhanian, N., Maham, B., Shah-Mansouri, V., Tushar, W., Yuen, C.: Game-theoretic approaches for energy cooperation in energy harvesting small cell networks. IEEE Trans. Veh. Technol. **66**(8), 7178–7194 (2017)
10. Zhang, D., et al.: Energy-harvesting-aided spectrum sensing and data transmission in heterogeneous cognitive radio sensor network. IEEE Trans. Veh. Technol. **66**(1), 831–843 (2017)
11. Gupta, L., Jain, R., Vaszkun, G.: Survey of important issues in UAV communication networks. IEEE Commun. Surv. Tutorials **18**(2), 1123–1152 (2016)
12. Liang, Y., Ke, S., Zhang, J., Yi, X., Zheng, Y.: Geoman: multi-level attention networks for geo-sensory time series prediction. In: IJCAI, vol. 2018, pp. 3428–3434 (2018)
13. Yao, H., Tang, X., Wei, H., Zheng, G., Yu, Y., Li, Z.: Modeling spatial-temporal dynamics for traffic prediction. arXiv preprint arXiv:1803.01254, pp. 922–929 (2018)
14. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proc. AAAI Conf. Artif. Intell. **33**, 922–929 (2019)
15. Feng, X., Guo, J., Qin, B., Liu, T., Liu, Y.: Effective deep memory networks for distant supervised relation extraction. In: IJCAI, vol. 17, pp. 1–8 (2017)
16. Simonovsky, M., Komodakis, N.: Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–10 (2017)
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning, December 2014, pp. 1–9 (2014)
19. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 1–8 (2018)