T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, *Member, IEEE*, Tao Lin, Min Deng, and Haifeng Li, *Member, IEEE*

***Abstract*— Accurate and real-time traffic forecasting plays an important role in the intelligent traffic system and is of great significance for urban traffic planning, traffic management, and traffic control. However, traffic forecasting has always been considered an "open" scientific issue, owing to the constraints of urban road network topological structure and the law of dynamic change with time. To capture the spatial and temporal dependences simultaneously, we propose a novel neural network-based traffic forecasting method, the temporal graph convolutional network (T-GCN) model, which is combined with the graph convolutional network (GCN) and the gated recurrent unit (GRU). Specifically, the GCN is used to learn complex topological structures for capturing spatial dependence and the gated recurrent unit is used to learn dynamic changes of traffic data for capturing temporal dependence. Then, the T-GCN model is employed to traffic forecasting based on the urban road network. Experiments demonstrate that our T-GCN model can obtain the spatio-temporal correlation from traffic data and the predictions outperform state-of-art baselines on real-world traffic datasets. Our tensorflow implementation of the T-GCN is available at https://github.com/lehaifeng/T-GCN.**

***Index Terms*— Traffic forecasting, temporal graph convolutional network (T-GCN), spatial dependence, temporal dependence.**

I. Introduction

With the development of Intelligent Traffic Systems, traffic forecasting has received more and more attention. It is a key part of an advanced traffic management system and is an important part of realizing traffic planning, traffic management, and traffic control. Traffic forecasting is a process of analyzing traffic conditions on urban roads, including flow, speed, and density, mining traffic patterns, and predicting the trends of traffic on roads. Traffic forecasting can provide not only a scientific basis for traffic managers to sense traffic congestion and limit vehicles in advance but also security for urban travelers to choose appropriate travel routes and improve travel efficiency [1]–[3]. However, traffic forecasting has always been a challenge task due to its complex spatial and temporal dependences:

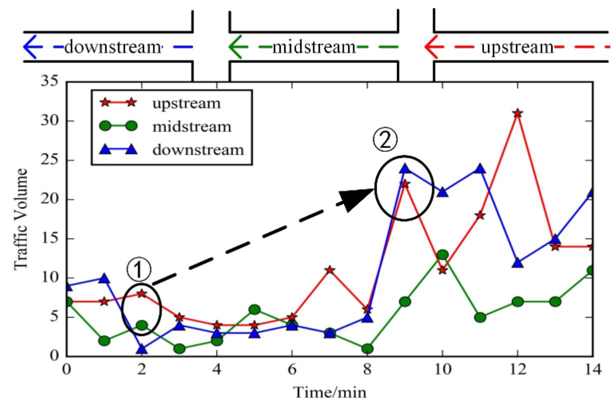


Fig. 1. Spatial dependence is restricted by the topological structure of the road network. Due to the strong influence between adjacent roads, the short-term traffic flow similarity is changed from state ① to state ②.

(1) Spatial dependence. The change in traffic volume is dominated by the topological structure of the urban road network. The traffic status at upstream roads impact traffic status at downstream roads through the transfer effect, and the traffic status at downstream roads impact traffic status at upstream through the feedback effect [4]. As shown in Figure 1, due to the strong influence among adjacent roads, the short-term similarity is changed from state ① (the upstream road is similar to the midstream road) to state ② (the upstream road is similar to the downstream road).

(2) Temporal dependence. The traffic volume changes dynamically over time and is mainly reflected in periodicity and trend. As shown in Figure 2(a), the traffic volume on the road shows a periodic change over a week. As shown in Figure 2(b), the traffic volume in one day changes over time; for example, the present traffic volume is affected by the traffic condition of the previous moment or even longer.

There are many existing traffic forecasting methods, some of which consider temporal dependence, including the

Manuscript received December 29, 2018; revised March 27, 2019 and June 21, 2019; accepted August 7, 2019. Date of publication August 22, 2019; date of current version August 28, 2020. This work was supported by the National Natural Science Foundation of China under Grant 41571397, Grant 41501442, Grant 41871364, Grant 51678077, Grant 71871224, and Grant 41771492. The Associate Editor for this article was J. Sanchez-Medina. *(Corresponding author: Haifeng Li.)*

L. Zhao, Y. Song, M. Deng, and H. Li are with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China (e-mail: lihaifeng@csu.edu.cn).

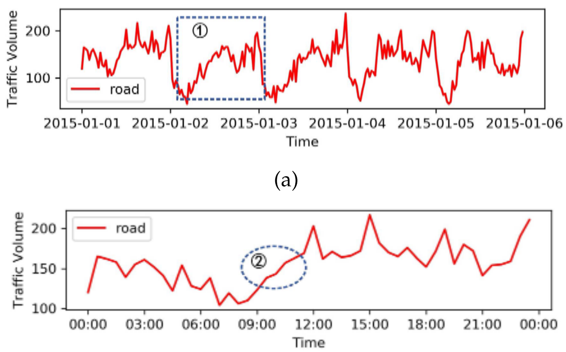
C. Zhang is with the School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Y. Liu is with the School of Earth and Space Sciences, Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China.

P. Wang is with the School of Traffic and Transportation Engineering, Central South University, Changsha 410083, China.

T. Lin is with the College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.

Digital Object Identifier 10.1109/TITS.2019.2935152



(a)

(b)

Fig. 2. (a) Periodicity. The traffic volume in the road changes periodically within one week. (b) Trend. The traffic volume in the road has tendency change within one day.

Autoregressive Integrated Moving Average (ARIMA) model [5], [6], the Kalman filtering model [7], the support vector regression machine model [8], [9], the k-nearest neighbor model [10], the Bayesian model [11], and partial neural network model [12], [13]. The above methods consider the dynamic change of traffic condition but ignore the spatial dependence, so that the change of traffic condition is not restricted by the road network and we cannot predict the state of traffic data accurately. To characterize the spatial features better, some studies [14]–[16] introduce a convolution neural network for spatial modeling; however, the convolutional neural network is commonly used for Euclidean data [17] such as images, regular grids, and so on. Such models cannot work under the context of an urban road network with complex topological structures so in essence they cannot describe the spatial dependence.

To solve the above problems, we propose a new traffic forecasting method called the temporal graph convolutional network (T-GCN) for traffic forecasting task based on urban road network. Our contributions are three-fold:

(1) The T-GCN model integrates the graph convolutional network and gated recurrent units. The graph convolutional network is used to capture the topological structure of the road network for modeling spatial dependence. The gated recurrent unit is used to capture the dynamic change of traffic data on the roads for modeling temporal dependence. The T-GCN model can also be applied to other spatio-temporal forecasting tasks.

(2) The forecasting results of the T-GCN model show a steady state under different prediction horizons, which indicates that the T-GCN model can not only achieve short-term prediction but also be used for long-term traffic prediction tasks.

(3) We evaluate our approach using two real-world traffic datasets. The results show that our approach reduces the prediction error by approximately 1.5%-57.8% compared to all baseline methods and demonstrats that the T-GCN model has superiority in traffic forecasting.

The rest of the paper is organized as follows. Section II reviews relevant research about traffic forecasting. Section III introduces the details of our method. In section IV, we evaluate the predictive performance of the T-GCN by real-world traffic dataset, including design of model parameters, prediction results analysis, perturbation analysis, and model interpretation. We conclude the paper in Section V.

II. Related Work

Existing traffic forecasting methods can be divided into two categories: the model-driven approaches and the data-driven approaches. First, the model-driven approaches mainly explain the instantaneous and steady-state relationships among traffic volume, speed, and density. Such methods require comprehensive and detailed system modeling based on prior knowledge. The representative methods contain the queuing theory model [18], the cell transmission model [19], the "traffic velocity" model [20], the microscopic fundamental diagram model [21], and so on. In reality, traffic data is influenced by many factors and it is difficult to obtain an accurate traffic model. The existing models cannot accurately describe the variations of traffic data in complex real-world environments. In addition, the construction of these models requires significant computing capability [22] and is easily influenced by traffic disturbances and sampling point spacing, etc.

Second, data-driven approaches infer the variation tendency based on statistical regularity of the data and are eventually used to predict and evaluate the traffic state [23], [24]. This type of methods does not analyze the physical properties and dynamic behavior of the traffic system and has high flexibility. One of the methods from this category is the historical average model [2], in which the average value of the traffic volume in historical periods is used as the prediction value. This method does not require any assumptions and the calculation is simple and fast but it does not consider temporal features and the prediction precision is low. With the continuous advancement of research on traffic forecasting, a large number of methods with higher prediction precision have emerged, which can be mainly divided into two categories based on the model: parametric or nonparametric [25], [26].

The parametric models presuppose the regression function, the parameters are determined through processing the original data, and then the traffic forecasting realization is based on the regression function. The time series model, the linear regression model [27], [28], and the Kalman filter model are common methods. The time series model fits a parametric model to the observed time series to predict future data. As early as 1976, Box and Jenkins [5] proposed the Autoregressive Integrate Moving Average Model (ARIMA), which is the most widely used time series model. In 1995, Hamed *et al.* [6] used the ARIMA model to predict the traffic volume in urban arterials. To improve the prediction precision of the model, different variants were produced, including Kohonen ARIMA [29], subset ARIMA [30], seasonal ARIMA [31], and so on. Lippi *et al.* [32] compared the support vector regression model with the seasonal ARIMA model and found that the former (SARIMA) model has better results in traffic congestion. The linear regression model builds a regression function based on historical traffic data to predict traffic flow. In 2004, Sun *et al.* [27] solved the problem of interval forecasting using the local linear model, and obtained better result on the real-world traffic dataset. The Kalman filter model predicts future traffic information based on the traffic state of the previous moment and the current moment. In 1984, Okutani and Stephanedes [7] used the Kalman filter to establish the traffic flow state prediction model. Subsequently, some studies [33], [34] used the Kalman filter to realize traffic prediction tasks.

The traditional parametric models have simple algorithms and convenient calculation. However, these models depend on the assumption that the system model is static, cannot reflect the nonlinearities and uncertainties of traffic data, and cannot overcome the interference of random events such as traffic accidents. The nonparametric models solve these problems well and only require enough historical data to learn the statistical regularity from traffic data automatically. The common nonparametric models include: the k-nearest neighbor model [10], the support vector regression model [8], [9], [35], the Fuzzy Logic model [36], the Bayesian network model [11], the neural network model, and so on.

In recent years, with the rapid development of deep learning [37]–[39], the deep neural network models have received attention because they can capture the dynamic characteristics of traffic data well and achieve the best results at present. According to whether or not spatial dependence is considered, models can be divided into two categories. Some methods consider temporal dependence only, e.g., Rilett *et al.* [40] used Feed Forward NN to implement traffic flow prediction tasks. Huang *et al.* [12] proposed a network architecture consisting of a deep belief network (DBN) and a regression model and verified that the network could capture random features from traffic data on multiple datasets and this model improved prediction accuracy in traffic forecasting. Qi *et al.* [41] presented the importance of congestion, proposed the unified definition of traffic congestion, and used the locality constraint distance metric learning method for traffic congestion detection tasks on the road. Subsequently, Qi *et al.* [42] proposed a robust hierarchical deep learning method for deep semantic feature extraction, which was used to better express the attributes of traffic congestion. In addition, since the recurrent neural network (RNN) and its variants long short-term memory (LSTM) and the gated recurrent unit (GRU) can effectively use the self-circulation mechanism, they can learn temporal dependence well and achieve better prediction results [13], [43].

Models mentioned above only take the temporal features into account but ignore the spatial dependence, so that the change of traffic data is not constrained by the urban road network and thus they cannot accurately predict the traffic information on the road. Making full use of the spatial and temporal dependences is the key to solve traffic forecasting problems. To better characterize spatial features, many studies had made improvements in this area. Lv *et al.* [44] proposed a SAE model to capture the spatio-temporal features from traffic data and realize short-term traffic flow prediction. Zhang *et al.* [14] proposed a deep learning model called ST-ResNet, which designed residual convolutional networks for each attribute based on the temporal closeness, period, and trend of crowd flows, and then three networks and external factors were dynamically aggregated to predict the inflow and outflow of crowds in each region of a city. Wu and Tan [15] designed a feature fusion architecture for short-term prediction by combining CNN and LSTM. A 1-dimensional CNN was used to capture spatial dependence and two LSTMs were used to mine the short-term variability and periodicity of traffic flow. Cao *et al.* [16] proposed an end-to-end model called ITRCN, which converted interactive network traffic into images and used CNN to capture interactive functions of traffic, used GRU to extract temporal features, and proved that the prediction error of this model is 14.3% and 13.0% higher than that of GRU and CNN, respectively. Ke *et al.* [45] proposed a new deep learning method called the fusion convolutional long short-term memory network (FCL-Net), taking into spatial dependence, temporal dependence, and exogenous dependence account for short-term passenger demand forecasting. Yu *et al.* [46] combined Deep Convolutional Neural Network (DCNN) and LSTM to propose a model named SRCN; in this model, DCNN was used to capture spatial dependence, while LSTM was used to capture temporal dynamics. SRCN has been proved effective and superior via experiments on Beijing traffic network data.

Although the above methods introduced the CNN to model spatial dependence and made great progress in traffic forecasting tasks, the CNN is essentially suitable for Euclidean space, such as images, regular grids, etc., and has limitations on traffic networks with complex topological structures, and thus cannot essentially characterize the spatial dependence. Therefore, this type of method also has certain defects. In recent years, with the development of the graph convolutional network model [47], which can be used to capture structural features of graph network, provides a good solution for the above problem. Li *et al.* [48] proposed a DCRNN model, which captured the spatial features through random walks on graphs, and the temporal features through encoder-decoder architecture.

Based on this background, in this research we propose a new neural network approach that can capture the complex temporal and spatial features from traffic data, and can then be used for traffic forecasting tasks based on an urban road network.

III. METHODOLOGY

A. Problem Definition

In this research, the goal of the traffic forecasting is to predict the traffic information in a certain period of time based on the historical traffic information on the roads. In our method, the traffic information is a general concept which can be traffic speed, traffic flow, or traffic density. Without loss of generality, we use traffic speed as an example of traffic information in experiment section.

Definition 1: road network G. We use an unweighted graph $G = (V, E)$ to describe the topological structure of the road network, and we treat each road as a node, where V is a set of road nodes, $V = \{v_1, v_2, \cdots, v_N\}$, N is the number of the nodes, and E is a set of edges. The adjacency matrix A is used to represent the connection between roads, $A \in R^{N \times N}$. The adjacency matrix contains only elements of 0 and 1.

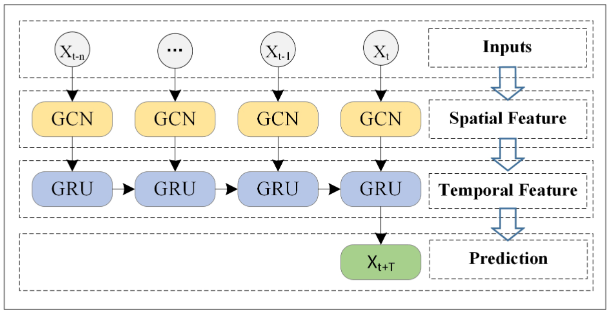


Fig. 3. Overview. We take the historical traffic information as input and obtain the finally prediction result through the Graph Convolution Network and the Gated Recurrent Unit model.

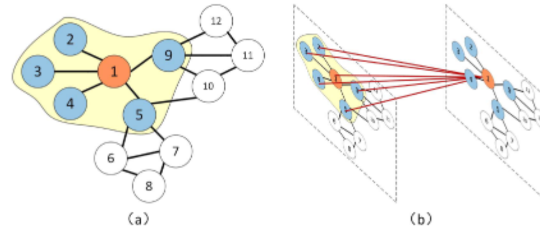


Fig. 4. Assuming that node 1 is a central road. (a) The blue nodes indicate the roads connected to the central road. (b) We obtain the spatial features by obtaining the topological relationship among the road 1 and its surrounding roads.

The element is 0 if there is no link between two roads and 1 denotes there is a link.

Definition 2: feature matrix $X^{N \times P}$. We regard the traffic information on the road network as the attribute features of the node in the network, expressed as $X \in R^{N \times P}$, where P represents the number of node attribute features (the length of the historical time series) and $X_t \in R^{N \times i}$ is used to represent the speed on each road at time i. Again, the node attribute features can be any traffic information such as traffic speed, traffic flow, and traffic density.

Thus, the problem of spatio-temporal traffic forecasting can be considered as learning the mapping function f on the premise of road network topology G and feature matrix X and then calculating the traffic information in the next T moments, as shown in equation 1:

$$\left[X_{t+1}, \cdots, X_{t+T}\right] = f\left(G; \left(X_{t-n}, \cdots, X_{t-1}, X_t\right)\right) \quad (1)$$

where n is the length of historical time series and T is the length of the time series needed to be predicted.

B. Overview

In this section, we describe how to use the T-GCN model to realize the traffic forecasting task based on the urban roads. Specifically, the T-GCN model consists of two parts: the graph convolutional network and the gated recurrent unit. As shown in Figure 3, we first use the historical n time series data as input and the graph convolution network is used to capture topological structure of urban road network for obtaining the spatial features. Second, the obtained time series with spatial features are input into the gated recurrent unit model and the dynamic change is obtained by information transmission between the units, to capture temporal features. Finally, we get results through the fully connected layer.

C. Methodology

1) Spatial Dependence Modeling: Acquiring the complex spatial dependence is a key problem in traffic forecasting. The traditional convolutional neural network (CNN) can obtain local spatial features, but it can only be used in Euclidean space, such as images, a regular grid, etc. An urban road network is in the form of graph rather than two-dimensional grid, which means the CNN model cannot reflect the complex topological structure of the urban road network and thus cannot accurately capture spatial dependence. Recently, generalizing the CNN to the graph convolutional network (GCN), which can handle arbitrary graph-structured data, has received widespread attention. The GCN model has been successfully used in many applications, including document classification [17], unsupervised learning [47] and image classification [49]. Given an adjacency matrix A and the feature matrix X, the GCN model constructs a filter in the Fourier domain. The filter, acting on the nodes of a graph, captures spatial features between the nodes by its first-order neighborhood, then the GCN model can be built by stacking multiple convolutional layers, which can be expressed as:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\, \widehat{A}\, \widetilde{D}^{-\frac{1}{2}}\, H^{(l)}\, \theta^{(l)}\right) \quad (2)$$

where $\widetilde{A} = A + I_N$ is the matrix with added self-connections, I_N is the identity matrix, $\widetilde{D}$ is the degree matrix, $\widetilde{D} = \sum_j \widetilde{A}_{ij}$, $H^{(l)}$ is the output of l layer, $\theta^{(l)}$ contains the parameters of that layer, and $\sigma(\cdot)$ represents the sigmoid function for a nonlinear model.

In this research, the 2-layer GCN model [47] is chosen to obtain spatial dependence, which can be expressed as:

$$f(X, A) = \sigma\left(\widehat{A}\, ReLU\left(\widehat{A}\, X\, W_0\right) W_1\right) \quad (3)$$

where $\widehat{A} = \widetilde{D}^{-\frac{1}{2}}\, \widetilde{A}\, \widetilde{D}^{-\frac{1}{2}}$ denotes pre-processing step, $W_0 \in R^{P \times H}$ represents the weight matrix from input to hidden layer, P is the length of feature matrix, and H is the number of hidden unit, $W_1 \in R^{H \times T}$ represents the weight matrix from hidden to output layer. $f(X, A) \in R^{N \times T}$ represents the output with the prediction length T, and $ReLU()$, standing for REctified Linear Unit, which is a frequently used activation layer in modern deep neural networks.

In summary, we use the GCN model [47] to learn spatial features from traffic data. As shown in Figure 4, assuming that node 1 is a central road, the GCN model can obtain the topological relationship between the central road and its surrounding roads, encode the topological structure of the road network and the attributes on the roads, and then obtain spatial dependence.

2) Temporal Dependence Modeling: Acquiring the temporal dependence is another key problem in traffic forecasting. At present, the most widely used neural network model for processing sequence data is the recurrent neural network (RNN). However, due to defects such as gradient disappearance and gradient explosion, the traditional recurrent neural

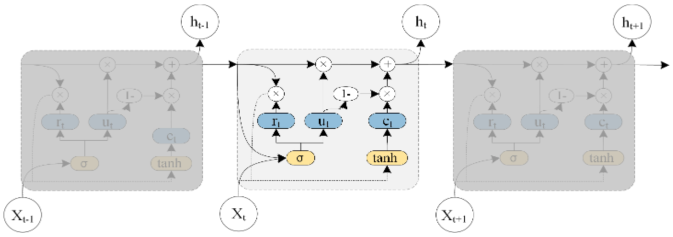


Fig. 5. The architecture of the Gated Recurrent Unit model.

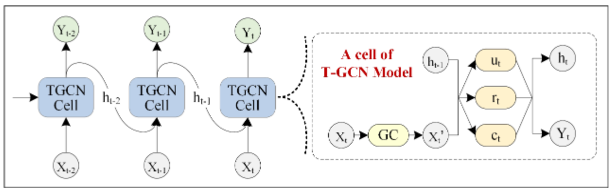


Fig. 6. The overall process of spatio-temporal prediction. The right part represents the specific architecture of a T-GCN unit, and GC represents graph convolution.

network has limitations for long-term prediction [50]. The LSTM model [51] and the GRU model [52] are variants of the recurrent neural network and have been proven to solve the above problems. The basic principles of the LSTM and GRU are roughly the same [53]. They all use gated mechanism to memorize as much long-term information as possible and are equally effective for various tasks. However, due to its complex structure, LSTM has a longer training time while the GRU model has a relatively simple structure, fewer parameters, and faster training ability. Therefore, we chose the GRU model to obtain temporal dependence from the traffic data. As shown in Figure 5, h_{t-1} denotes the hidden state at time t-1; x_t denotes the traffic information at time t; r_t is the reset gate, which is used to control the degree of ignoring the status information at the previous moment; u_t is the update gate, which is used to control the degree of to which the status information at the previous time is brought into the current status; c_t is the memory content stored at time t; and h_t is output state at time t. The GRU obtains the traffic information at time t by taking the hidden status at time t-1 and the current traffic information as inputs. While capturing the traffic information at the current moment, the model still retains the changing trend of historical traffic information and has the ability to capture temporal dependence.

3) Temporal Graph Convolutional Network: To capture the spatial and temporal dependences from traffic data at the same time, we propose a temporal graph convolutional network model (T-GCN) based on the graph convolutional network and gated recurrent unit. As shown in Figure 6, the left side is the process of spatio-temporal traffic prediction, the right side shows the specific structure of a T-GCN cell, h_{t-1} denotes the output at time t-1, GC is graph convolution process, and u_t, r_t are update gate and reset gate at time t, and h_t denotes the output at time t.

The specific calculation process is shown below. $f(A, X_t)$ represents the graph convolution process and is defined in equation 3. represents the point-wise multiplication. W and b represent the weights and biases in the training process.

$$u_t = \sigma(W_u [f(A, X_t), h_{t-1}] + b_u) \tag{4}$$
$$r_t = \sigma(W_r [f(A, X_t), h_{t-1}] + b_r) \tag{5}$$
$$c_t = tanh(W_c [f(A, X_t), (r_t * h_{t-1})] + b_c) \tag{6}$$
$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t \tag{7}$$

In summary, the T-GCN model can deal with the complex spatial dependence and temporal dynamics. On one hand, the graph convolutional network is used to capture the topological structure of the urban road network for obtaining the spatial dependence. On the other hand, the gated recurrent unit is used to capture the dynamic variation of traffic information on the roads for obtaining the temporal dependence and eventually for realizing traffic prediction tasks.

4) Loss Function: In the training process, the goal is to minimize the error between the real traffic speed on the roads and the predicted value. We use Y_t and $\widehat{Y_t}$ to denote the real traffic speed and the predicted speed, respectively. The loss function of the T-GCN model is shown in equation 8. The first term is used to minimize the error between the real traffic speed and the prediction. The second term L_{reg} is the L2 regularization term that helps to avoid an overfitting problem and λ is a hyperparameter.

$$loss = \| Y_t - \widehat{Y_t} \| + \lambda L_{reg} \tag{8}$$

IV. Experiments

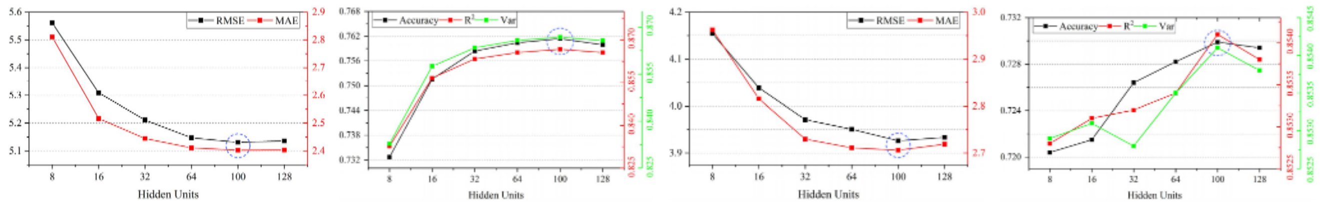
A. Data Description

In this section, we evaluate the prediction performance of the T-GCN model on two real-world datasets: SZ-taxi dataset and Los-loop dataset. Since these two datesets are all related to traffic speed. Without loss of generality, we use traffic speed as traffic information in the experiment section.

(1) SZ-taxi. This dataset consists of the taxi trajectory of Shenzhen from Jan. 1 to Jan. 31, 2015. We selected 156 major roads of Luohu District as the study area. The experimental data mainly includes two parts. One is an 156*156 adjacency matrix, which describes the spatial relationship between roads. Each row represents one road and the values in the matrix represent the connectivity between the roads. Another one is a feature matrix, which describes the speed changes over time on each road. Each row represents one road; each column is the traffic speed on the roads in different time periods. We aggregated the traffic speed on each road every 15 minutes.

(2) Los-loop. This dataset was collected from the highway of Los Angeles County in real time by loop detectors. We selected 207 sensors and its traffic speed from Mar. 1 to Mar. 7, 2012. We aggregated the traffic speed every 5 minutes. Similarity, the data consists of an adjacency matrix and a feature matrix. The adjacency matrix is calculated by the distance between sensors in the traffic networks. Since the Los-loop dataset contained some missing data, we used the linear interpolation method to fill missing values.

In the experiments, the input data is normalized to the interval [0, 1]. In addition, 80% of the data is used as the training set and the remaining 20% is used as the test set.



(a) (b) (c) (d)

Fig. 7. Comparison of predicted performance under different hidden units in the training and test set based on SZ-taxi dataset. (a) Changes in RMSE and MAE in the training set. (b) Changes in Accuracy, R^2 and var in the training set. (c) Changes in RMSE and MAE in the test set. (d) Changes in Accuracy, R^2 and var based in the test set.

We predict the traffic speed of the next 15 minutes, 30 minutes, 45 minutes and 60 minutes.

B. Evaluation Metrics

We use five metrics to evaluate the prediction performance of the T-GCN model:

(1) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \widehat{y_i^j})^2} \tag{9}$$

(2) Mean Absolute Error (MAE):

$$MAE = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left|y_i^j - \widehat{y_i^j}\right| \tag{10}$$

(3) Accuracy:

$$Accuracy = 1 - \frac{\| Y - \widehat{Y} \|_F}{\| Y \|_F} \tag{11}$$

(4) Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \widehat{y_i^j})^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \bar{Y})^2} \tag{12}$$

(5) Explained Variance Score (var):

$$var = 1 - \frac{Var\left\{Y - \widehat{Y}\right\}}{Var\left\{Y\right\}} \tag{13}$$

where y_i^j and $\widehat{y_i^j}$ represent the real traffic information and predicted one of the jth time sample in the ith road. M is the number of time samples; N is the number of roads; Y and $\widehat{Y}$ represent the set of y_i^j and $\widehat{y_i^j}$ respectively, and $\bar{Y}$ is the average of Y.

Specifically, RMSE and MAE are used to measure the prediction error: the smaller the value is, the better the prediction effect is. Accuracy is used to detect the prediction precision: the lager the value is, the better the prediction effect is. R^2 and var calculate the correlation coefficient, which measures the ability of the predicted result to represent the actual data: the larger the value is, the better the prediction effect is.

C. Choosing Model Parameters

In this section, we choose the relevant parameters of the T-GCN model.

(1) Hyperparameters

The hyperparameters of the T-GCN model mainly include: learning rate, batch size, training epoch, and the number of hidden unit. In the experiment, we manually adjust and set the learning rate to 0.001, the batch size to 32, and the training epoch to 5000.

The number of hidden units is a very important parameter of the T-GCN model, as different number of hidden units may greatly affect the prediction precision. To choose the best value, we experiment with different hidden units and select the optimal value by comparing the predictions.

In our experiment, for the SZ-taxi dataset, we choose the number of hidden units from [8, 16, 32, 64, 100, 128] and analyze the change of prediction precision. As shown in 7, the horizontal axis represents the number of hidden units and the vertical axis represents the change of different metrics. Figure 7(a) shows the results of RMSE and MAE for different hidden units in the training set. It can be seen that the error is the smallest when the number is 100. Figure 7(b) shows the variation of Accuracy, R^2, and var for different hidden units. Figure 7(c) and 7(d) show the results in the test set. Similarly, the results reach a maximum when the number is 100. In summary, the prediction results are better when the number is set to 100. When increasing the number of hidden units, the prediction precision firstly increases and then decreases. This is mainly because when the hidden unit is larger than a certain degree, the model complexity and the computational difficulty are greatly increased and as a result, overfitting on the training data occurs. Therefore, we set the number of hidden units to 100 in our experiments on the SZ-taxi dataset.

In the same way, the results of Los-loop are shown in Figure 8(a), 8(b), 8(c), 8(d). it can be seen that when the number of hidden units is 64, the prediction precision is the highest, and the prediction error is the lowest.

(2) Training

For input layer, the training dataset (80% of the overall data) is taken as input in the training process and the remaining data is used as input in the test process. The T-GCN model is trained using the Adam optimizer.

D. Experimental Results

We compare the performance of the T-GCN model with the following baseline methods:

(1) History Average model (HA) [2], which uses the average traffic information in the historical periods as the prediction.

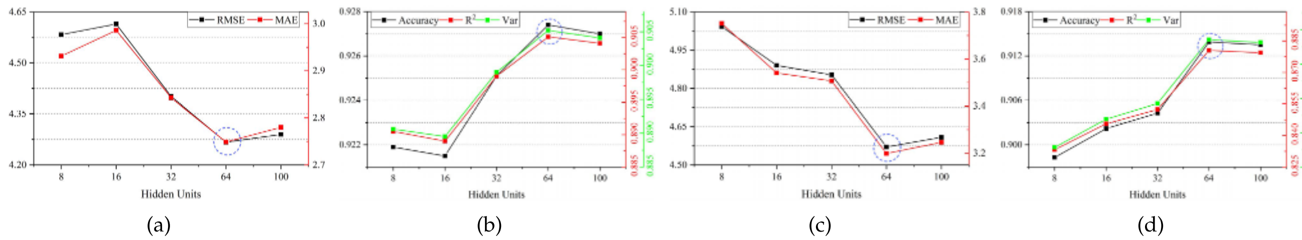


Fig. 8. Comparison of predicted performance under different hidden units in the training and test set based on Los-loop dataset. (a) Changes in RMSE and MAE in the training set. (b) Changes in Accuracy, R^2 and var in the training set. (c) Changes in RMSE and MAE in the test set. (d) Changes in Accuracy, R^2 and var in the test set.

TABLE I

THE PREDICTION RESULTS OF THE T-GCN MODEL AND OTHER BASELINE METHODS ON SZ-TAXI AND LOS-LOOP DATASETS

T	Metric	SZ-taxi						Los-loop					
		HA	ARIMA	SVR	GCN	GRU	T-GCN	HA	ARIMA	SVR	GCN	GRU	T-GCN
15min	$RMSE$	4.2951	7.2406	4.1455	5.6596	3.9994	**3.9265**	7.4427	10.0439	6.0084	7.7922	5.2182	**5.1264**
	MAE	2.7815	4.9824	2.6233	4.2367	**2.5955**	2.7117	4.0145	7.6832	3.7285	5.3525	**3.0602**	3.1802
	$Accuracy$	0.7008	0.4463	0.7112	0.6107	0.7249	**0.7299**	0.8733	0.8275	0.8977	0.8673	0.9109	**0.9127**
	R^2	0.8307	$*$	0.8423	0.6654	0.8329	**0.8541**	0.7121	$*$	0.8123	0.6843	0.8576	**0.8634**
	var	0.8307	0.0035	0.8424	0.6655	0.8329	**0.8541**	0.7121	$*$	0.8146	0.6844	0.8577	**0.8634**
30min	$RMSE$	4.2951	6.7899	4.1628	5.6918	4.0942	**3.9663**	7.4427	9.3450	6.9588	8.3353	6.2802	**6.0598**
	MAE	2.7815	4.6765	**2.6875**	4.2647	2.6906	2.7410	4.0145	7.6891	3.7248	5.6118	**3.6505**	3.7466
	$Accuracy$	0.7008	0.3845	0.7100	0.6085	0.7184	**0.7272**	0.8733	0.8275	0.8815	0.8581	0.8931	**0.8968**
	R^2	0.8307	$*$	0.8410	0.6616	0.8249	**0.8456**	0.7121	$*$	0.7492	0.6402	0.7957	**0.8098**
	var	0.8307	0.0081	0.8413	0.6617	0.8250	**0.8457**	0.7121	$*$	0.7523	0.6404	0.7958	**0.8100**
45min	$RMSE$	4.2951	6.7852	4.1885	5.7142	4.1534	**3.9859**	7.4427	10.0508	7.7504	8.8036	7.0343	**6.7065**
	MAE	2.7815	4.6734	**2.7359**	4.2844	2.7743	2.7612	4.0145	7.6924	4.1288	5.9534	**4.0915**	4.1158
	$Accuracy$	0.7008	0.3847	0.7082	0.6069	0.7143	**0.7258**	0.8733	0.8273	0.8680	0.8500	0.8801	**0.8857**
	R^2	0.8307	$*$	0.8391	0.6589	0.8198	**0.8441**	0.7121	$*$	0.6899	0.5999	0.7446	**0.7679**
	var	0.8307	0.0087	0.8397	0.6590	0.8199	**0.8441**	0.7121	$*$	0.6947	0.6001	0.7451	**0.7684**
60min	$RMSE$	4.2951	6.7708	4.2156	5.7361	4.0747	**4.0048**	7.4427	10.0538	8.4388	9.2657	7.6621	**7.2677**
	MAE	2.7815	4.6655	2.7751	4.3034	**2.7712**	2.7889	4.0145	7.6952	**4.5036**	6.2892	4.5186	4.6021
	$Accuracy$	0.7008	0.3851	0.7063	0.6054	0.7197	**0.7243**	0.8733	0.8273	0.8562	0.8421	0.8694	**0.8762**
	R^2	0.8307	$*$	0.8370	0.6564	0.8266	**0.8422**	0.7121	$*$	0.6336	0.5583	0.6980	**0.7283**
	var	0.8307	0.0111	0.8379	0.6564	0.8267	**0.8423**	0.7121	$*$	0.5593	0.5593	0.6984	**0.7290**

(2) Autoregressive Integrated Moving Average model (ARIMA) [5], which fits a parametric model on the observed time series to predict future traffic data.

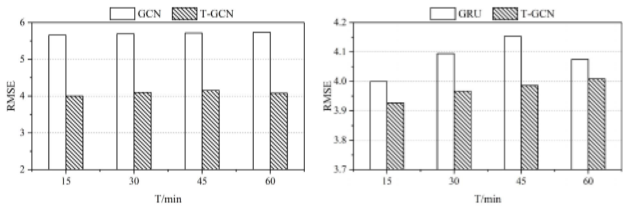
(3) Support Vector Regression model (SVR) [35], which uses historical data to train the model and obtains the relationship between the input and output, and then predicts the future traffic data by the trained model. The kernel function we use in this model is a linear kernel.

(4) Graph Convolutional Network model (GCN) [47]: see 3.2.1 for details.

(5) Gated Recurrent Unit model (GRU) [52]: see 3.2.2 for details.

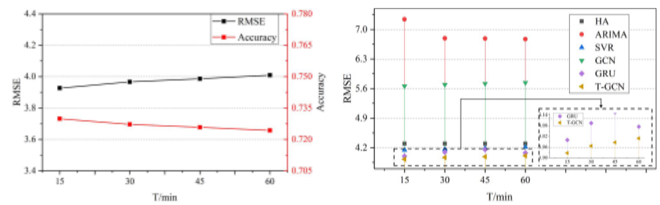
Table I shows the performance of T-GCN model and other baseline methods for 15 minutes, 30 minutes, 45 minutes and 60 minutes forecasting tasks on SZ-taxi and Los-loop datasets. $*$ means that the values are small enough to be negligible, indicating that the model's prediction effect is poor. It can be seen that the T-GCN model obtains the best prediction performance under almost all evaluation metrics for all prediction horizons, proving the effectiveness for spatio-temporal traffic forecasting tasks.

(1) High prediction precision. We can find that the neural network-based methods, including the T-GCN model, the GRU model, which emphasize the importance of modeling the temporal features, generally have better prediction precision than other baselines, such as the HA model, the ARIMA model and the SVR model. For example, for the 15-min traffic forecasting task, RMSE errors of the T-GCN model and the GRU model are reduced by approximately 8.58% and 6.88% compared with the HA model, and the accuracies are approximately 4.15% and 3.44% higher than that of HA. The RMSE errors of the T-GCN model and the GRU model are approximately 45.77% and 32.97% lower than that of the ARIMA model and the accuracies of these two models are improved by 63.54% and 62.42%. Compared with the SVR model, the RMSE errors of the T-GCN and the GRU models are reduced by 5.28% and 0.67%, and approximately 2.63% and 1.93% higher than that of the SVR model. This is mainly due to methods such as the HA, ARIMA, and SVR that find it difficult to handle complex, nonstationary time series data. The lower prediction effect of the GCN model is because the GCN considers the spatial features only and ignores that the traffic data is typical time series data. In addition, as a mature traffic forecasting method, the ARIMA's prediction precision is relatively lower than the HA, mainly because the ARIMA has difficulty dealing with long-term and nonstationary data



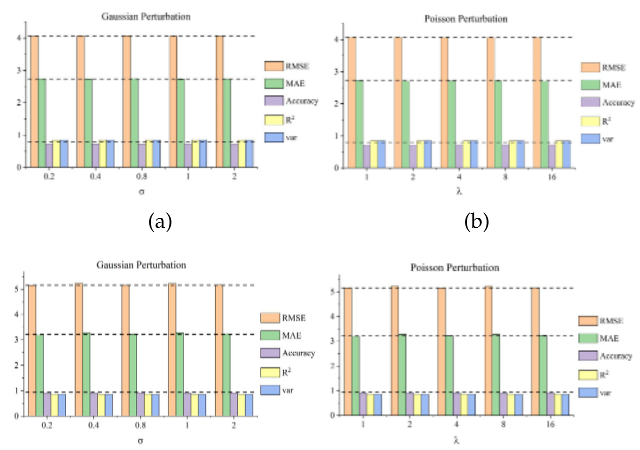
(a) (b)

Fig. 9. Spatio-temporal prediction capability. (a) The RMSE of the T-GCN model lower than the GCN model, which considers spatial features only, indicating the effectiveness of the TGCN to capture spatial features. (b) The RMSE of the T-GCN model lower than the GRU model, which considers temporal features only, indicating the effectiveness of the T-GCN to capture temporal features.



(a) (b)

Fig. 10. Long-term prediction ability. (a) Under different prediction horizons, the change in RMSE and Accuracy of the T-GCN model. (b) Under different prediction horizons, the RMSE errors of the T-GCN model and all baseline methods.



(a) (b)

(c) (d)

Fig. 11. Perturbation analysis. (a) The results of adding Gaussian perturbation on SZ-taxi dataset. (b) The results of adding Poisson perturbation on SZ-taxi dataset. (c) The results of adding Gaussian perturbation on Los-loop dataset. (d) The results of adding Poisson perturbation on Los-loop dataset.

and the ARIMA is calculated by calculating the error of each node and averaging; if there are fluctuations in some data, it will also increase the final total error.

(2) Spatio-temporal prediction capability. To verify whether the T-GCN model has the ability to portray spatial and temporal features from traffic data, we compare the T-GCN model with the GCN model and the GRU model. As shown in Figure 9, we can clearly see that method based on the spatio-temporal features (T-GCN) has better prediction precision than those based on single factor (GCN, GRU), indicating that the T-GCN model can capture spatial and temporal features from traffic data. For example, for the 15-min traffic forecasting, the RMSE is reduced by approximately 30.62% compared with the GCN model, which considers only spatial features and for 30-min traffic forecasting, the RMSE of the T-GCN model is reduced by 30.32%, indicating that the T-GCN model can capture spatial dependence. Compared with the GRU model, which considers only temporal features, for 15-min and 30-min traffic forecasting, the RMSE errors of the T-GCN model are decreased by approximately 1.82% and 3.12%, indicating that the T-GCN model can capture temporal dependence well.

(3) Long-term prediction ability. No matter how the horizon changes, the T-GCN model can obtain the best prediction performance through training and the prediction results have less tendency to change, indicating that our approach is insensitive to prediction horizons. Thus, we know that the T-GCN model can be used not only for short-term prediction but also for long-term prediction. Figure 10(a) shows the change of RMSE and Accuracy at different prediction horizons, which represent the prediction error and precision of the T-GCN model, respectively. It can be seen that the trends of error increase and precision decrease are small, with a certain degree of stability. Figure 10(b) shows the comparison of RMSE for baselines at different horizons. We observe that the T-GCN model can achieve the best results regardless of the prediction horizon.

E. Perturbation Analysis and Robustness

There is inevitably noise during the data collection process in the real world. To test the noise immunity of the T-GCN model, we test the robustness of the model through perturbation analysis experiments.

We add two types of common random noise to the data during the experiment. The random noise obeys the Gaussian distribution $N \in (0, \sigma^2)$, where $\sigma \in (0.2, 0.4, 0.8, 1, 2)$ and the Poisson distribution $P(\lambda)$, where $\lambda \in (1, 2, 4, 8, 16)$. Then, we normalize the values of the noise matrices turn to be between 0 and 1. Using different evaluation metrics, the results are shown as following. Figure 11(a) shows the results of adding Gaussian noise on SZ-taxi dataset, where the horizontal axis represents σ, the vertical axis represents the change of each evaluation metrics, and different colors indicate different metrics. Similarly, Figure 11(b) shows the results of adding Poisson noise on SZ-taxi dataset. Figure 11(c) and 11(d) are the results of adding Gaussian noise and Poisson noise on Los-loop dataset. It can be seen that the change in metrics is small whatever the noise distribution is. Thus, the T-GCN model is robust and is able to handle high noise issues.

F. Model Interpretation

To better understand the T-GCN model, we select one road on SZ-taxi dataset and visualize prediction results of the test set. Figure 12, Figure 13, Figure 14, and Figure 15 show the visualization results for prediction horizons of 15 minutes, 30 minutes, 45 minutes, and 60 minutes, respectively. These results show:

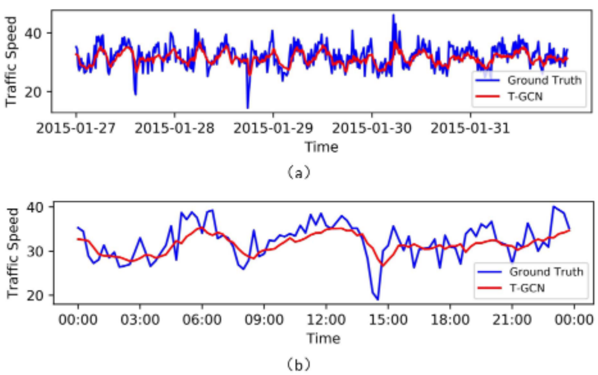


Fig. 12. The visualization results for prediction horizon of 15 minutes.

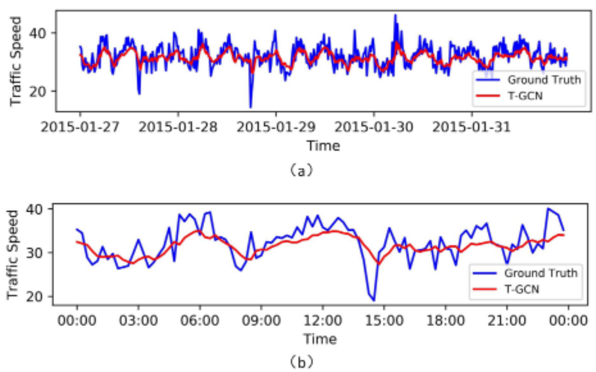


Fig. 13. The visualization results for prediction horizon of 30 minutes.

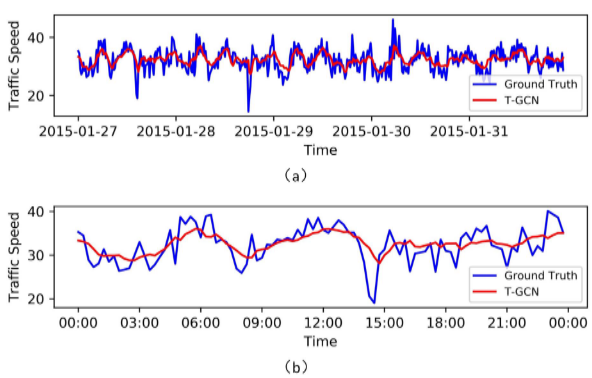


Fig. 14. The visualization results for prediction horizon of 45 minutes.

(1) The T-GCN model predicts poorly at local minima/ maxima. We speculate that the main cause is that the GCN model defines a smooth filter in the Fourier domain and captures spatial feature by constantly moving the filter. This process leads to a small change in the overall prediction results, which makes peaks smoother.

(2) There is a certain error between the real traffic information and the prediction results caused by "zero taxi value". Zero taxi value is the phenomenon which traffic feature matrix, whose true value is not zero, will be set to zero because of no taxis on the road.

(3) Regardless of the prediction horizons, the T-GCN model always achieves better results. The T-GCN model can capture

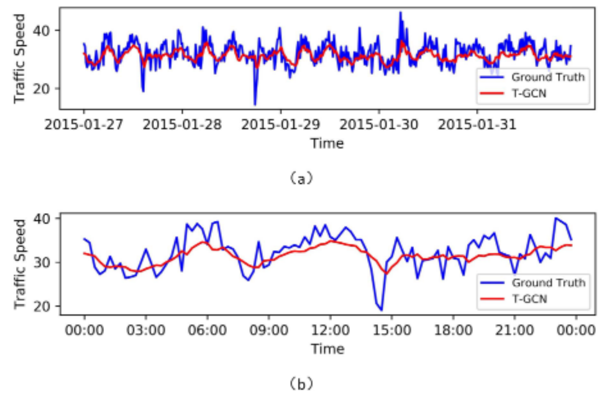


Fig. 15. The visualization results for prediction horizon of 60 minutes.

the spatio-temporal features and obtain the variation trend of traffic information on the road. Moreover, the T-GCN model detects the start and end of the rush hour and makes prediction results with a similar pattern with the real traffic speed. Those properties are helpful for predicting traffic congestion and other traffic phenomena.

V. Conclusion

This research develops a novel neural network based approach for traffic forecasting called T-GCN, which combines the GCN and the GRU. We use a graph network to model the urban road network in which the nodes on the graph represent roads, the edges represent the connection relationships between roads, and the traffic information on the roads is described as the attribute of the nodes on the graph. On one hand, the GCN is used to capture the spatial topological structure of the graph for obtaining the spatial dependence; on the other hand, the GRU model is introduced to capture the dynamic change of node attribute for obtaining the temporal dependence. Eventually the T-GCN model is used to tackle spatio-temporal traffic forecasting tasks. When evaluated on two real-world traffic datasets and compared with the HA model, the ARIMA model, the SVR model, the GCN model, and the GRU model, the T-GCN model achieves a better performance under different prediction horizons. In addition, the perturbation analysis illustrates the robustness of our approach. In summary, the T-GCN model successfully captures the spatial and temporal features from traffic data so that can be applied to other spatio-temporal tasks.

References

[1] H. Huang, "Dynamic modeling of urban transportation networks and analysis of its travel behaviors," *Chin. J. Manage.*, vol. 2, no. 1, pp. 18–22, Jan. 2005.

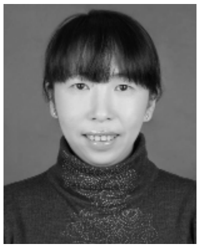
[2] J. Liu and W. Guan, "A summary of traffic flow forecasting methods," *J. Highway Transp. Res. Develop.*, vol. 21, no. 3, pp. 82–85, Mar. 2004.

[3] J. Yuan and B. Fan, "Synthesis of short–term traffic flow forecasting research progress," *Urban Transp. China*, vol. 10, no. 6, pp. 73–79, Jun. 2012.

[4] C. J. Dong, C. F. Shao, Z. Cheng-Xiang, and M. Meng, "Spatial and temporal characteristics for congested traffic on urban expressway," *J. Beijing Univ. Technol.*, vol. 38, no. 8, pp. 128–132, 2012.

[5] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, no. 722, pp. 1–9, 1979.

[6] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in Urban arterials," *J. Transp. Eng.*, vol. 121, no. 3, pp. 249–254, 1995.
[7] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, 1984.
[8] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.
[9] Z. S. Yao, C. F. Shao, and Y. L. Gao, "Research on methods of short-term traffic forecasting based on support vector regression," *J. Beijing Jiaotong Univ.*, vol. 30, no. 3, pp. 19–22, 2006.
[10] X. L. Zhang, H. E. Guo-Guang, and L. U. Hua-Pu, "Short-term traffic flow forecasting based on K-nearest neighbors non-parametric regression," *J. Syst. Eng.*, vol. 24, no. 2, pp. 178–183, Feb. 2009.
[11] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006.
[12] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.
[13] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Automat. (YAC)*, Wuhan, China, Nov. 2016, pp. 324–328.
[14] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 1655–1661.
[15] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," Dec. 2016, *arXiv:1612.01022*. [Online]. Available: https://arxiv.org/abs/1612.01022
[16] X. Cao, Y. Zhong, Y. Zhou, J. Wang, C. Zhu, and W. Zhang, "Interactive temporal recurrent convolution network for traffic prediction in data centers," *IEEE Access*, vol. 6, pp. 5276–5289, 2018.
[17] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2016, pp. 3844–3852.
[18] X.-Y. Xu, J. Liu, H.-Y. Li, and J.-Q. Hu, "Analysis of subway station capacity with the use of queueing theory," *Transp. Res. C, Emerg. Technol.*, vol. 38, no. 1, pp. 28–43, Jan. 2014.
[19] P. Wei, Y. Cao, and D. Sun, "Total unimodularity and decomposition method for large-scale air traffic cell transmission model," *Transp. Res. B, Methodol.*, vol. 53, pp. 1–16, Jul. 2013.
[20] W. Qi, L. I. Li, H. U. Jianming, and B. Zou, "Traffic velocity distributions for different spacings," *J. Tsinghua Univ. Sci. Technol.*, vol. 51, no. 3, pp. 309–312, Mar. 2011.
[21] F. F. Xu, Z. C. He, and Z. R. Sha, "Impacts of traffic management measures on urban network microscopic fundamental diagram," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 13, no. 2, pp. 185–190, Apr. 2013.
[22] E. I. Vlahogianni, "Computational intelligence and optimization for transportation big data: Challenges and opportunities," in *Engineering and Applied Sciences Optimization*, vol. 38, N. Lagaros and M. Papadrakakis, Eds. Cham, Switzerland: Springer, 2015.
[23] Z. Shan, D. Zhao, and Y. Xia, "Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst.*, The Hague, Netherlands, Oct. 2013, pp. 118–123.
[24] S. G. X. Xiangjie, "Short-term traffic volume intelligent hybrid forecasting model and its application," *Syst. Eng.—Theory Pract.*, vol. 31, no. 3, pp. 562–568, 2011.
[25] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, Sep. 2004.
[26] H. van Lint and C. van Hinsbergen, "Short-term traffic and travel time prediction models," *Artif. Intell. Appl. Critical Transp.*, vol. 22, pp. 22–41, Nov. 2012.
[27] H. Sun, C. Zhang, and B. Ran, "Interval prediction for traffic time series using local linear predictor," in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, Washington, DC, USA, Nov. 2004, pp. 410–415.
[28] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 130, pp. 139–147, Jan. 2016.
[29] M. van der Voort, M. Dougherty, and S. Watson, "Combining kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.
[30] S. Lee and D. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1678, no. 1, pp. 179–188, 1999.
[31] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.
[32] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.
[33] C. P. I. J. van Hinsbergen, T. Schreiter, F. S. Zuurbier, J. W. C. van Lint, and H. J. van Zuylen, "Localized extended Kalman filter for scalable real-time traffic state estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 385–394, Mar. 2012.
[34] L. L. Ojeda, A. Y. Kibangou, and C. C. de Wit, "Adaptive Kalman filtering for multi-step ahead traffic flow prediction," in *Proc. Amer. Control Conf.*, Washington, DC, USA, Jun. 2013, pp. 4724–4729.
[35] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[36] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.
[37] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
[38] D. Silver *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.
[39] M. Moravčik, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker," *Science*, vol. 356, no. 6337, pp. 508–513, May 2017.
[40] D. Park and L. R. Rilett, "Forecasting freeway link travel times with a multilayer feedforward neural network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 14, no. 5, pp. 357–367, 1999.
[41] Q. Wang, J. Wan, and Y. Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognit.*, vol. 75, pp. 272–281, Mar. 2018.
[42] W. Qi, W. Jia, and L. Xuelong, "Robust hierarchical deep learning for vehicular management," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4148–4156, May 2019.
[43] J. W. C. Van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Freeway travel time prediction with state-space neural networks: Modeling state-space dynamics with recurrent neural networks," *Transp. Res. Rec.*, vol. 1811, no. 1, pp. 30–39, Jan. 2002.
[44] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[45] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *J. Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.
[46] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, Jun. 2017.
[47] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," Sep. 2016, *arXiv:1609.02907*. [Online]. Available: https://arxiv.org/abs/1609.02907
[48] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," Jul. 2017, *arXiv:1707.01926*. [Online]. Available: https://arxiv.org/abs/1707.01926
[49] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun, "Spectral networks and locally connected networks on graphs," Dec. 2013, *arXiv:1312.6203*. [Online]. Available: https://arxiv.org/abs/1312.6203
[50] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[52] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," Sep. 2014, *arXiv:1409.1259*. [Online]. Available: https://arxiv.org/abs/1409.1259
[53] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," Dec. 2014, *arXiv:1412.3555*. [Online]. Available: https://arxiv.org/abs/1412.3555

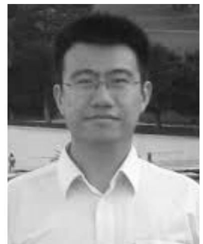


Ling Zhao received the Ph.D. degree from the School of Geosciences and Info-Physics, Central South University, Changsha, China. She is currently an Associate Professor with the School of Geosciences and Info-Physics, Central South University. She has published more than 10 peer-reviewed journal articles. Her research interests mainly concentrate in humanities and social science based on big geodata and artificial intelligence.

Yujiao Song received the B.S. degree in 2016. She is currently pursuing the master's degree with the School of Geosciences and Info-Physics, Central South University. Her research interests include humanities and social science based on big geodata and artificial intelligence.

Chao Zhang received the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign. He is currently an Assistant Professor with the School of Computational Science and Engineering, Georgia Institute of Technology. He has published more than 50 papers in top-tier conferences and journals. His work has been honored by multiple awards, including the ACM SIGKDD Dissertation Runner-Up Award and the ECML/PKDD Best Student Paper Runner-Up Award. His research interests include data mining and machine learning. He is particularly interested in developing label-efficient and robust learning techniques, with applications in text mining and spatiotemporal data mining.

Yu Liu received the B.S., M.S., and Ph.D. degrees from Peking University in 1994, 1997, and 2003, respectively. He is currently a Professor with the Institute of Remote Sensing and Geographic Information System, Peking University. His research interest mainly concentrates in humanities and social science based on big geo-data.



Pu Wang received the B.S. degree in physics from the University of Science and Technology of China, Hefei, China, in 2005, and the Ph.D. degree in physics from the University of Notre Dame, Notre Dame, IN, USA, in 2010. From May 2010 to December 2011, he was a Post-Doctoral Researcher with the Department of Civil and Environmental Engineering, MIT. He is currently a Full Professor with the School of Traffic and Transportation Engineering, Central South University, Changsha, China. He has authored or coauthored several papers in international leading journals, such as *Science*, *Nature Physics*, and *Nature Communications*. His research interests include complex networks, traffic analysis, human dynamics, and data mining. He is a Guest Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Tao Lin received the M.S. and Ph.D. degrees in agricultural and biological engineering from the University of Illinois at Urbana–Champaign. He joined the Faculty of the Biosystems Engineering Department, Zhejiang University, by 100 Talents Program of Zhejiang University and serviced as a Full Research Professor. He has published more than ten peer-reviewed journal articles. His research focuses on agricultural big data and AI systems informatics and analytics, ranging from spatio-temporal analysis, GIS, optimization modeling analysis, to high performance cyberinfrastructure enabled decision support systems.

Min Deng is currently a Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. His research interests mainly concentrate in humanities and social science based on big geodata and artificial intelligence.

Haifeng Li (M'15) received the master's degree in transportation engineering from the South China University of Technology, Guangzhou, China, in 2005, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009. He was a Research Associate with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, in 2011, and a Visiting Scholar with the University of Illinois at Urbana–Champaign, Urbana, IL, USA, from 2013 to 2014. He is currently a Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. He has authored more than 30 journal papers. His current research interests include geo/remote sensing big data, machine/deep learning, and artificial/brain-inspired intelligence. He is a reviewer for many journals.