Information Sciences 621 (2023) 580–595

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Dynamic traffic correlations based spatio-temporal graph convolutional network for urban traffic prediction

Yuanbo Xu, Xiao Cai, En Wang, Wenbin Liu, Yongjian Yang, Funing Yang *

Jilin University, Changchun, Jilin, China

A R T I C L E I N F O

Article history:
Received 2 December 2021
Received in revised form 15 November 2022
Accepted 18 November 2022
Available online 25 November 2022

Keywords:
Dynamic traffic correlations
Urban traffic prediction
Graph convolutional network
Long short term memory network
Attention networks

A B S T R A C T

Accurate urban traffic prediction is a critical issue in Intelligent Transportation Systems (ITS). It is challenging since urban traffic usually indicates high dynamic spatio-temporal correlations, leading to uncertainty and complexity of traffic status. Since the transportation network is a graph structure practically, existing works have applied Graph Convolutional Network (GCN) on urban traffic prediction with a pre-defined adjacency matrix based on node distance or connectivity. However, in many urban traffic scenarios, spatio-temporal dependencies among traffic data usually change over time, so using a fixed adjacency matrix cannot describe the dynamic dependencies. To track the dynamic spatio-temporal dependencies among traffic data, we propose a novel deep learning framework, Dynamic Traffic Correlation-based Spatio-Temporal Graph Convolutional network (DTC-STGCN), to forecast traffic flow and speed accurately. DTC-STGCN extracts a dynamic adjacency matrix from different traffic characters to describe dynamic spatio-temporal correlations. Moreover, an attention and dynamic adjacency matrix-based GCNs framework is proposed to capture urban traffic dynamic spatial features, while a long-short-term memory network (LSTM) is used to capture urban traffic temporal features, respectively. Finally, we feed the spatio-temporal features generated by GCN and LSTM, with real road segments into a hybrid graph convolution framework to simultaneously model the dynamic spatial and temporal dependencies for traffic predictions. The experiments on two real-world datasets demonstrate that the proposed DTC-STGCN model consistently outperforms the state-of-the-art traffic prediction baselines on MAE and RMSE over 10%, and achieve a stable performance for two specific tasks (long-term traffic prediction and peak time prediction). And ablation study validates the effectiveness of dynamic adjacency matrix, attention mechanism, respectively.

1. Introduction

Urban traffic prediction is one of the most challenging tasks in Intelligent Transportation Systems (ITS) [1]. Moreover, accurate and reliable traffic prediction has become a mission-critical work for developing a smart city, as it can provide insights for urban planning and dynamic traffic management, which will improve the efficiency of public transportation. Urban traffic prediction models are generally designed to accurately forecast future traffic states (the traffic speed and the traffic flow) of urban traffic networks by considering spatio-temporal correlations among sequential historical traffic data.

* Corresponding author.

https://doi.org/10.1016/j.ins.2022.11.086
0020-0255/

Recently, great efforts have been devoted to improving the urban traffic prediction accuracy [2]. Some studies apply traditional machine learning methods, such as autoregressive integrated moving average model (ARIMA) [3–5] and support vector regression (SVR) [6] and etc, to forecast future traffic states. As most of these methods are linear and unsuitable for handling complex spatio-temporal traffic data, the traffic forecasting accuracy is often low. In recent years, prediction methods based on deep learning technologies have received considerable attention. Some attempts have been made to apply deep Recurrent Neural Networks (RNN) [7] and Convolution Neural Network (CNN) [8–10] to predict traffic states. However, these methods are not suitable to apply to the data points with irregular graph relationship.

As the transportation network is a graph structure practically, Graph Convolutional Network (GCN) is an appealing choice [11]. Hence, many researchers use the method, which combined graph convolutional networks with other deep learning technologies, to deal with high-dimensional spatio-temporal urban traffic data [12–14]. These methods above that utilize GCN to handle signals effectively, which live on irregular or non-Euclidean domains, have outperformed the methods based on traditional deep learning technologies. Moreover, these GCN-based methods rely on the key assumption that the adjacency matrix is strictly unchanged (i,e. the input graph's adjacency matrix is constant). Nevertheless, due to the following dynamic spatio-temporal correlations, these existing frameworks still have some limitations, making them less efficient in urban traffic prediction:

1) Dynamic spatial correlations on traffic network. As shown in Fig. 1, the connected road segments affect each other at each timeslice, and the effect changes over time. For instance, the color of dotted line between C and D is different at t_1 and t_2, which indicates the spatial correlation strength between C and D at t_1 is stronger than that at t_2. The previous studies based on GCN use a fixed adjacency matrix to describe the spatial ontology structure of traffic networks, which cannot capture the dynamic spatial correlations.

2) Dynamic temporal correlations of urban traffic. As shown by the solid lines in the y-axis dimension in Fig. 1, the traffic states of road networks, such as traffic flow and traffic speed, are correlated with their previous states, and the correlations are dynamic. For instance, the correlation between the traffic states at t and t_1 has a greater strength than that between the traffic states at t_1 and t_2 on node B.

To capture the dynamic spatio-temporal correlations above, we have made some improvements in terms of data aspect and model aspect, respectively. From the data perspective, we calculate the dynamic adjacency matrix through a dynamic correlation matrix and feed it into the GCN framework to replace the original fixed adjacency matrix. From the model perspective, we design two attention mechanisms throughout the framework and a novel deep learning-based model structure that simultaneously takes the dynamic spatio-temporal features and road information to learn the dynamic spatio-temporal dependencies. Along with this line, we propose a novel spatio-temporal structure to forecast network-wide urban traffic

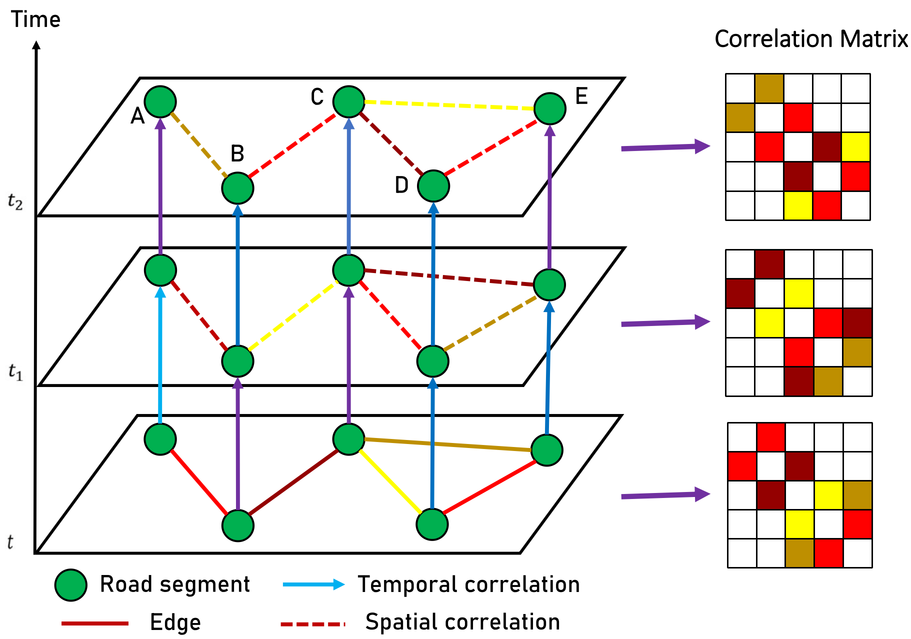

Fig. 1. This figure introduces an example of the transportation network at t, t1, and t2, and the dynamic spatio-temporal correlations of urban traffic. Each node denotes a road segment, and each edge represents the relationship between roads. The dotted lines indicate the spatial correlations, and the solid lines indicate the temporal correlations, respectively. Besides, the color of the edges indicates the strength of the spatio-temporal correlations. Moreover, the three matrices on the right of the figure represent the spatial correlation matrix of the urban traffic network at the three timestamps, respectively, and the color of each grid represents the size of the value, which is different apparently. The above figure shows that the strength of spatial–temporal correlations between nodes changes over time.

states more accurately. And we call it Dynamic-Traffic-Correlations based Spatio-Temporal Graph Convolutional Network (DTC-STGCN). Compared with existing GCN-based methods, our paper makes the following contributions:

- We use three different features to calculate the dynamic adjacency matrix correlated with the dynamic correlation matrix that from the real-world traffic data, which can adapt to the dynamic changes of the spatial relationship in urban traffic.
- We design a novel deep learning-based framework to learn dynamic spatio-temporal dependencies. A dynamic adjacency matrix and attention based GCN module is proposed to learn dynamic spatial features from dynamic graph-based transportation networks. Besides, an LSTM network is applied to capture the dynamic temporal dependencies. Finally, with a hybrid attention-based GCN framework, we can simultaneously model the dynamic spatio-temporal features and road information features.
- We conduct experiments on two real-world datasets in predicting urban traffic flow and traffic speed, respectively. The experimental results show that our model DTC-STGCN consistently outperforms the other state-of-the-art models on both prediction tasks.

The rest of this paper is organized as follows: We first review the relevant works about urban traffic prediction in Section 2. Then, Section 3 introduces the background knowledge and relevant definitions in our research. We present the technical details of our proposed DTC-STGCN model in Section 4. After that, we evaluate the performance of our proposed model DTC-STGCN through experiments on two real-world datasets to predict urban traffic flow and traffic speed respectively in Section 5, and conclude our work in Section 6.

2. Related work

Urban traffic forecasting accurately has a long history already. Most of the early works for urban traffic forecasting were based on statistics and time series model such as history average (HA), vectors auto regression (VAR), auto regressive integrated moving average (ARIMA) [3] and ARIMA based variants model [4,5]. These models require the data to satisfy some patterns and only consider urban traffic dependencies in the temporal dimension. However, urban traffic data generally takes on intricate spatial and temporal patterns, so they usually perform poorly in practice.

Machine learning models such as K-nearest neighbors (KNN) [15], support vectors machine (SVM) [16], support vectors regression (SVR) [17], and Bayesian model [18] are of alleviating above challenges and modeling more complex data by feature engineering to extract multi-dimensional features to adapt to specific application problems better. However, they need to pay much attention to feature engineering. With the rapid development of deep learning technology in recent years, many researchers are committed to applying deep learning technology to predict urban traffic conditions owing to automatic features extracting and excellent performance on big and complex data [19,20]. Considering the need for historical data in urban traffic prediction, some researchers make use of recurrent neural network (RNN), LSTM, and gated recurrent unit (GRU) to capture the temporal correlations in traffic data [21],22–24. Nevertheless, it is not enough to only focus on temporal correlations because urban traffic also has complex spatial patterns. Hence, the combination of convolutional neural network (CNN), residual neural network and deep learning models appear above in many researches for simultaneously capturing spatio-temporal patterns of urban traffic [25–28]. Although traditional deep learning models can effectively extract the spatio-temporal features in traffic data, they can only be applied to the standard grid data.

The transportation network is a graph structure practically so that it is well-suited to the graph neural network (GNN), which is specifically designed for the graph data structure. Researchers are shifting to GCN-based models to develop more general and widely-used traffic forecasting methods in recent years. STGCN [11] introduced the complete convolutional structures to mine the spatial–temporal patterns of urban traffic. MRes-RGNN [12] first proposed to adopt residual neural network in graph neural network to make it more sensitive to sudden changes in urban traffic. MVGCN [29] builds a multi-view graph convolutional network to capture the multiple temporal correlations among different time intervals. DCRNN [30] re-formulates the spatial dependency of traffic as a diffusion process and extends the previous GCN to a directed graph. Following DCRNN, Graph Wavenet [31] combines GCN with dilated causal convolution networks for saving computation cost in handling long sequence and propose a self-adaptive adaptive adjacency matrix as a complement for the predefined adjacent matrix to capture spatial correlations. Considering the high dynamics in urban traffic, some researchers developed the dynamic GCN structures for urban traffic predictions [32,33]. DGCNN [32] incorporated the tensor operation into the neural network to estimate the dynamic Laplacian matrices for achieving more accurate predictions. AGCRN [34] can capture fine-grained spatial and temporal correlations in traffic series automatically without pre-defined graphs. Moreover, More recent works such as ASTGCN [13], GMAN [14], ST-MetaNet+ [35], ST-GDN [36], GAT [37] and STSGCN [38] further add more complicated spatial and temporal attention mechanisms with GCN to capture the dynamic spatial and temporal correlations. These GNN-based models have effectively learned the spatial dependencies on the graph structure, especially those combined with the attention mechanism, which has shown excellent performance. However, these models utilize the fixed adjacency matrix that describes the geographic connectivity or distance or the adaptive adjacency matrix without practical significance.

3. Preliminaries

3.1. Urban traffic network

In our work, we define the traffic network as a dynamic graph $G_t = (V, E, A_t)$as shown in Fig. 2. G_t denotes the dynamic traffic graphs at t timestamps. V is a node set, corresponding to the observations of N road segments in traffic networks. E is a set of edges, indicating the connectivity between the nodes. A_t is a set of adjacency matrix at t timestamps, corresponding to the fixed connectivity of traffic graph and the dynamic correlations between nodes in graph G simultaneously.

3.2. Dynamic adjacency matrix

The correlations between road segments in traffic graphs change over time. Hence, this paper utilizes one dynamic adjacency matrix A_t in the GCN framework, corresponding to the initial dynamic spatial correlations information between nodes. A novel calculated adjacency matrix is used for graph convolutional networks at each timeslice. Furthermore, we propose three methods for calculating the dynamic adjacency matrix correlated with the spatial correlation matrix. These three calculation methods represent the number of vehicles that transfer between roads, feature ratio, and feature influence, respectively.

3.3. Traffic prediction on dynamic graphs

Given a set of dynamic traffic graph G_n and a set of traffic observations O_n: traffic speed and traffic flow, in last n timestamps, the traffic prediction can be formulated as follows,

$$O_k = f_o(G_n, O_n), \tag{1}$$

where O_k is the traffic observations in the future k timestamps, the n denotes the observed urban traffic data horizon, the k denotes the urban traffic prediction horizon and the f_o denotes the model on traffic prediction task proposed by us.

4. Dynamic traffic correlations based spatio-temporal graph convolutional networks

This section elaborates on the proposed architecture of dynamic traffic correlations based on spatio-temporal graph convolutional network (DTC-STGCN). As shown in Fig. 3, DTC-STGCN mainly consists of four components: dynamic adjacency matrix that describes dynamic spatial correlations of traffic network, attention-based graph convolutional networks with a dynamic adjacency matrix, which capture the dynamic spatial correlations, LSTM network for capturing the dynamic temporal correlations and attention-based GCN framework to learn the dynamic spatio-temporal dependencies and road information features simultaneously. The details of each module are described as follows.

4.1. Dynamic adjacency matrix

GCN heavily depends on the adjacency matrix, which is defined as the spatial correlations between nodes. The spatial correlations between road segments change over time in urban traffic, but the adjacency matrix is stable in previous GCN-based methods, which cannot capture the change. Therefore, we calculate the dynamic adjacency matrix correlated with the dynamic spatial correlation matrix to replace the original fixed adjacency matrix and feed it into the GCN framework in our work.

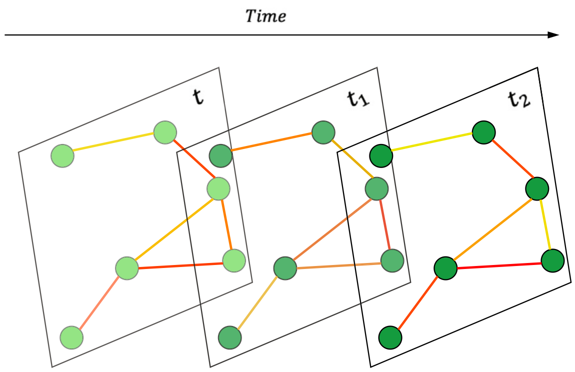

Fig. 2. The spatial–temporal structure of traffic data, where the data at each time slice t form a graph.

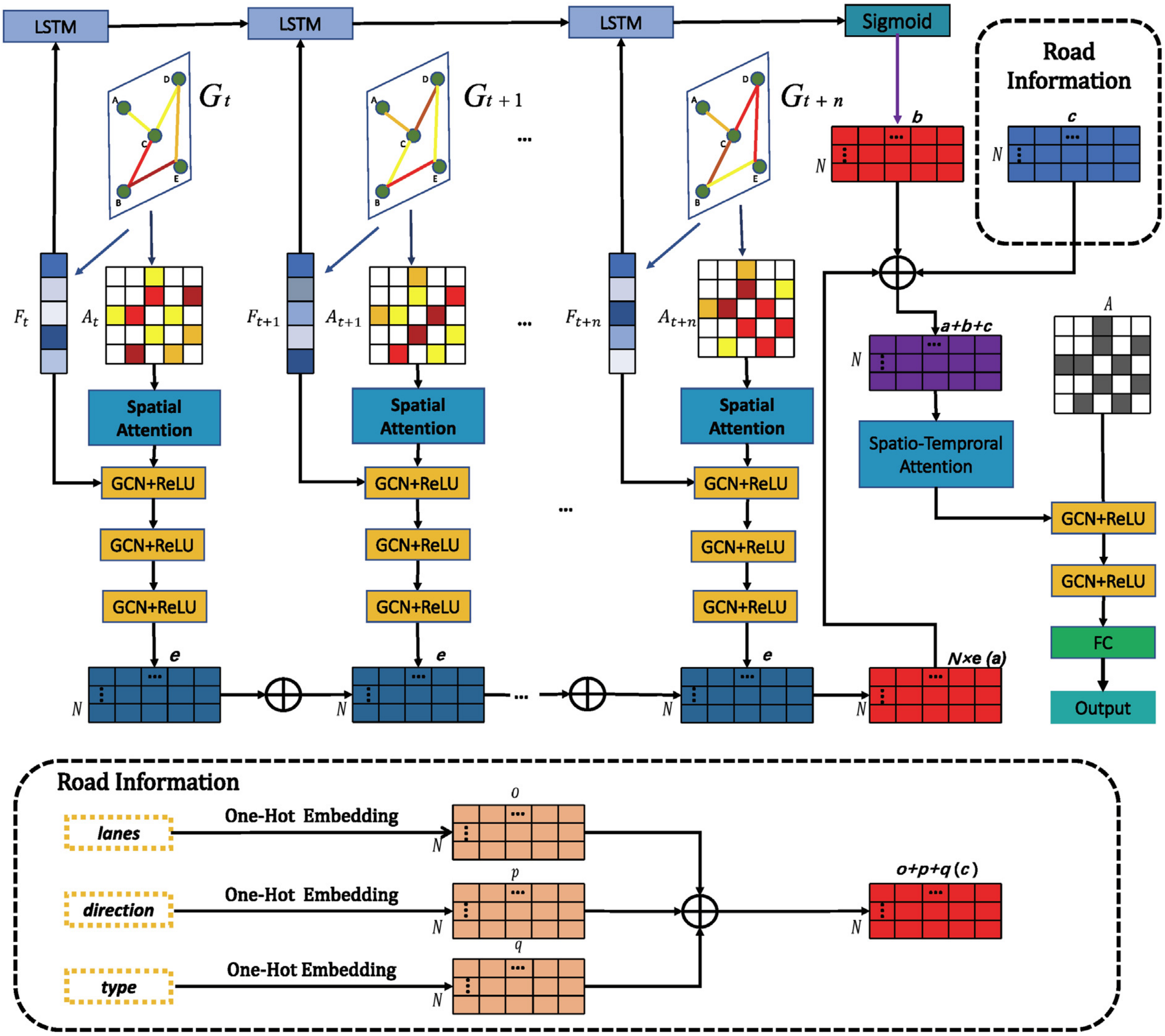

Fig. 3. The architecture of DTC-STGCN for urban traffic forecasting.

This paper proposed three different dynamic adjacency matrices, which denote the dynamic spatial correlations between all road segments in the urban traffic network. The three dynamic adjacency matrices describe the number of taxis that transfer between roads, feature difference, and feature ratio, respectively. The first dynamic adjacency matrix will be utilized when we have the detailed trajectories of taxis. The other two methods will be used when we only have traffic status data.

The first type of spatial correlation matrix S^t describes the number of vehicles that transfer between road segments at each timestamp t. This type of adjacency matrix is only applied on the CC-taxi dataset, where each value S^t_{ij} is calculated as follow,

$$S^t_{ij} = N^t_{ij}, \tag{2}$$

where N^t_{ij} denotes the number of vehicles which transfer from the road segment i to the road segment j at timestamp t.

The second type of spatial correlation matrix S^t describes the ratio of feature values observed between all nodes in the urban traffic network at each timestamp t. Each value S^t_{ij} in spatial correlation matrix is calculated as,

$$S^t_{ij} = \frac{F^t_i}{F^t_i + F^t_j}, \tag{3}$$

where F^t_i denotes the feature value observed in the road segment i at t, while F^t_j denotes the feature value observed in the road segment j at t.

The third type of spatial correlation matrix S^t describes the absolute influence values of the feature value observed between all nodes in the graph at each timestamp, where each value S^t_{ij} is calculated as,

$$S^t_{ij} = |F^t_i - F^t_j|, \tag{4}$$

where F^t_i denotes the observed feature value in the road segment i at t, and F^t_j denotes the feature value observed in the road segment j at t.

In (3) and (4), the traffic feature F observed denotes traffic flow on the CC-Taxi dataset, and traffic speed on the SZ-Taxi dataset, respectively.

Correspondingly, we obtain three types of adjacency matrix A_t from the above three types of spatial correlation matrix S^t, which is calculated as,

$$A_t = S^t \odot A, \tag{5}$$

where $\odot$ denotes the element-wise Hadamard product. The A denotes the fixed adjacency matrix as follows,

$$A_{ij} = \begin{cases} 1, & if\, v_i and\, v_j areconnected, \\ \\ 0, & \text{ifnot.} \end{cases} \tag{6}$$

We can infer that the adjacency matrix A_t we generated is dynamic, describing traffic networks' fixed connectivity and representing the dynamic spatial correlation of road segments.

4.2. Attention based graph convolutional networks

The urban traffic network is a graph structure, but the urban traffic forecasting methods, based on traditional deep learning approaches, apply to grid-based data only. While the graph neural network is designed for graph-based data, so it is well suited for traffic research. In this paper, to make full use of the urban traffic network's dynamic spatial properties, we conduct attention-based graph convolutional operations on a dynamic adjacency matrix at each timestamp.

GCN is an efficient variant of convolutional neural networks that can operate directly on graphs. Formally, assuming that there are N nodes with M-dimensional features (or attributes) in a graph, the topological structure and node attributes can be represented by an adjacency matrix $A \in \mathbb{R}^{N\times N}$ and a feature matrix $F \in \mathbb{R}^{N\times M}$ (in which the *i-th* row of F corresponds to the feature vector of node i), respectively. Moreover, in graph analysis, a graph can be represented by its corresponding Laplacian matrix. Moreover, the properties of the graph structure can be obtained by analyzing the Laplacian matrix. Laplacian matrix of a graph is defined as $L = D - A$, and its normalized form is $L = I_N - D^{-1/2}AD^{-1/2} \in \mathbb{R}^{N\times N}$, where A is the adjacency matrix, I_N is a unit matrix, and the degree matrix $D \in \mathbb{R}^{N\times N}$ is a diagonal matrix, consisting of node degrees, $D_{ii} = \sum_j A_{ij}$.

In this work, we utilize the GCN to capture the spatial dependencies of each traffic snapshot firstly. Given a feature matrix $F_t \in \mathbb{R}^{N\times M}$ where N denotes the number of nodes, and M denotes the dimension of the feature of nodes, and an adjacency matrix A_t at t, the graph convolutional operation is an iterative process, which can be briefly defined as follows:

$$H^i_t = f\left(H^{i-1}_t, A_t\right), \tag{7}$$

where the i denotes the *i-th* graph convolution layer, the f is the spreading function that aggregates the feature information of neighbor nodes, and $H^i_t \in \mathbb{R}^{N\times F^i}$ denotes the feature vector representation matrix of all nodes in the *i-th* graph convolution layer at t. $H^0_t = F$ in each layer is a feature matrix at t, where each row denotes the feature representation of a node and F^i denotes the number of hidden units in the *i-th* layer.

In our work, the spreading rule we adopted is as follow,

$$f\left(H^i_t, A_t\right) = GCN\left(A_t, H^{i-1}_t\right) = ReLU\left(A_t H^{i-1}_t W^i_t\right), \tag{8}$$

where the $W^i_t \in \mathbb{R}^{F^i\times F^{i+1}}$ denotes the learnable weight matrix in the *i-th* convolution layer, and *ReLU* denotes activation function.

Moreover, we calculate the Laplacian matrix of each traffic snapshot as,

$$L_t = D^{-1/2}\left(D - \hat{A}_t\right)D^{-1/2} = \widetilde{D}^{-1/2}\tilde{A}_t\widetilde{D}^{-1/2}, \tag{9}$$

where $\tilde{A}_t = \hat{A}_t + I_N, \widetilde{A} = \widehat{A} + I_N$ and $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$.

Finally, the graph convolutional operation in our work is set up as,

$$f\left(H^i_t, A_t\right) = ReLU\left(\widetilde{D}^{-1/2}\tilde{A}_t\widetilde{D}^{-1/2}H^{i-1}_t W^i_t\right). \tag{10}$$

As shown in Fig. 3, we utilize three graph convolution layers on each traffic graph at each timestamp for capturing the dynamic spatial correlation more completely. In order to adaptively capture the most significant spatial correlations between nodes for higher accuracy in future urban traffic forecasting, we adopt an attention algorithm P^t for each historical timestamp in the first GCN layer. Considering the strong temporal correlation of traffic condition and the dynamic spatial pattern in urban traffic data, the current traffic state is closely related to the road features and spatial correlations between roads in the last timestamp. Therefore, we set the attention P_t with the traffic feature F_{t-1} and adjacency matrix $\widehat{A}_{t-1}$ as input,

$$P^t = V_t^s \sigma\left(\left(F_{t-1} W_t^s\right) W_t^p \left(W_t^a \widehat{A}_{t-1}\right) + b_t^s\right), \tag{11}$$

$$P_{i,j}^t = \frac{\exp\left(P_{i,j}^t\right)}{\sum_{j=1}^{N} \exp\left(P_{i,j}^t\right)}, \tag{12}$$

where $P^t \in \mathbb{R}^{N \times N}$ denotes the spatial attention matrix at t, F_{t-1} is the feature matrix at the last timestamp, $\widehat{A}_{t-1}$ denotes the adjacency matrix at the last timestamp, $V_t^s, W_t^p, W_t^a, b_t^s \in \mathbb{R}^{N \times N}$ and $W_t^s \in \mathbb{R}^{M \times N}$ are learnable parameters and the sigmoid $\sigma(\cdot)$ is used as the activation function. Then a softmax function is used to ensure the attention weights of a node sum to 1. The value of each element $P_{i,j}^t$ in P^t semantically represents the correlation strength between node i and node j. Note that there are n different spatial attention matrix P^t in model from the Fig. 4. In our work, when performing the first graph convolution, we will accompany the adjacency matrix A_t with the spatial attention matrix P^t to dynamically adjust the impacting weights between nodes.

Eventually, we obtain the spatial feature vector representation of traffic snapshots at the past n timestamps, $H_1, H_2, \cdots, H_n \in \mathbb{R}^{N \times e}$, where e denotes the number of hidden units in the last graph convolution layer. we concate all vector representations as the dynamic spatial features of all nodes $H_s \in \mathbb{R}^{N \times (e \times n)}$,

$$H_s = H_1 \oplus H_2 \oplus \cdots \oplus H_n. \tag{13}$$

4.3. Long short term memory network

Attention-based graph convolutional network has learned the dynamic spatial correlations of traffic data. However, the urban traffic is a typical time-series study, so we feed the historical traffic observation values O_n into the LSTM network to capture the dynamic temporal dependencies between different timestamps.

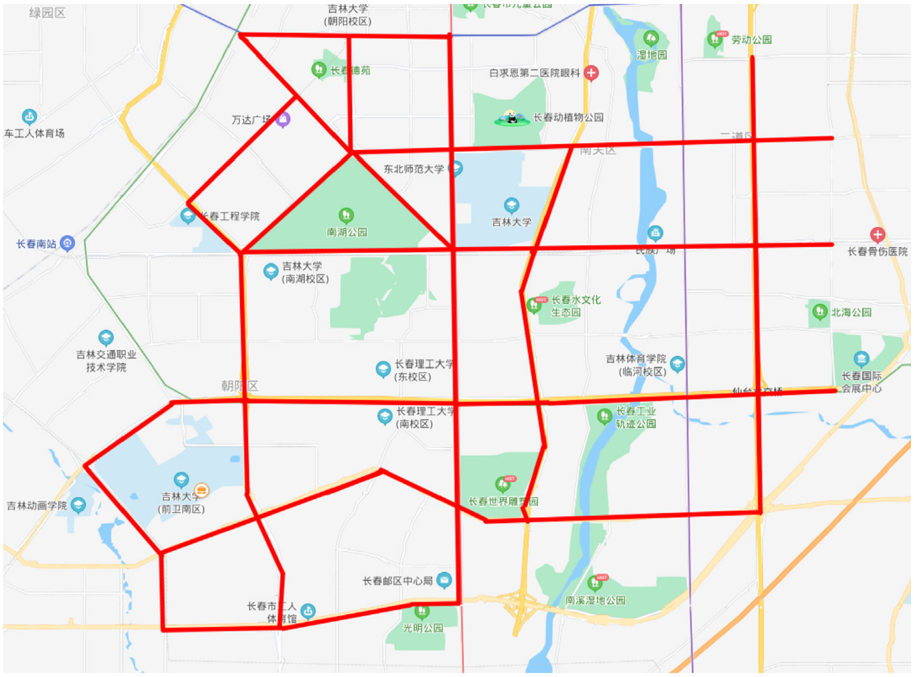

Fig. 4. The roads we selected in the central district of ChangChun.

The LSTM network has a powerful capacity to learn the long-term dependencies of sequential data, to capture the evolving patterns of the weighted dynamic traffic networks. The standard LSTM architecture can be described as an encapsulated cell with several multiplicative gate units. For a certain timestamp t, the LSTM cell takes current input vector x_t as well as the state vector of last timestamp h_{t-1} as the input, and then output the state vector in the current time step h_t:

$$i_t = \sigma\left(W_x^i x_t + W_h^i h_{t-1} + b^i\right), \tag{14}$$

$$f_t = \sigma\left(W_x^f x_t + W_h^f h_{t-1} + b^f\right), \tag{15}$$

$$o_t = \sigma\left(W_x^o x_t + W_h^o h_{t-1} + b^o\right), \tag{16}$$

$$s_t = f_t \odot s_{t-1} + i_t \odot .\widetilde{s_t}, \tag{17}$$

$$\widetilde{s_t} = \sigma\left(W_x^s x_t + W_h^s h_{t-1} + b^s\right), \tag{18}$$

$$h_t = o_t \odot \tanh(s_t), \tag{19}$$

where i_t, f_t, o_t and s_t represent the input gate, forget gate, output gate and memory cell, respectively, W_x, W_h, b are the parameters of the corresponding unit, $\sigma(\cdot)$ is the sigmoid activation function, and the $\odot$ denotes the element-wise multiplication.

Eventually, we treat the last hidden state h_{t-1}as the distributed temporal correlation representation of the historical traffic snapshots.

4.4. Attention based spatio-temporal graph convolutional network

We have obtained the dynamic spatial and temporal dependencies from GCN and LSTM framework, respectively. Then, we consider utilizing a unified graph convolutional network to learn the dynamic spatio-temporal dependencies and road information features simultaneously.

In our work, the road information features consist of the number of lanes, the direction, and types of road. We utilize one-hot embedding methods to encode the three types of road information features respectively as,

$$F_l = \text{One-Hot Embedding } (\textit{lanes}), \tag{20}$$

$$F_d = \text{One-Hot Embedding } (\textit{directions}), \tag{21}$$

$$F_y = \text{One-Hot Embedding } (\textit{types}), \tag{22}$$

Eventually, we concat F_l, F_d, F_y as the unified road information feature representations as,

$$F_r = F_l \oplus F_d \oplus F_y. \tag{23}$$

Supposing that the spatial feature generated by attention based GCN framework is a-dimension $\left(F_s \in \mathbb{R}^{N\times a}\right)$, the temporal dependence generated by LSTM network is b-dimension $\left(F_t \in \mathbb{R}^{N\times b}\right)$ and the road information feature is c-dimension $\left(F_r \in \mathbb{R}^{N\times c}\right)$, we concat the three vector representations as the node feature representation as,

$$F_o = F_s \oplus F_t \oplus F_r. \tag{24}$$

In order to adjust the contribution of each feature for urban traffic prediction adaptively, we set a soft attention mechanism R_o implemented by a full connected layer as,

$$R_o = softmax(tanh(w_o F_o + b_o)), \tag{25}$$

where $w_o, b_o$are learnable parameters.

The spatio-temporal attention R_o will be accompanied by feature matrix F_o to be fed into GCN. Furthermore, we use two graph convolutional layers on complete nodes features F_o and a fixed adjacency matrix with taking spatio-temporal dependencies and road information into consideration simultaneously as follows,

$$H_{o1} = ReLU\left(D^{-1/2} A D^{-1/2} (F_o \odot R_o) W_{o1}\right), \tag{26}$$

$$H_o = ReLU\left(D^{-1/2} A D^{-1/2} H_{o1} W_o\right), \tag{27}$$

where the H_{o1} is the output of the first graph convolutional layer, the H_o is the output of second graph convolutional layer, the $\odot$ is element-wise Hardmard product, the $W_{o1} \in \mathbb{R}^{(a+b+c)\times m}, W_o \in \mathbb{R}^{m\times d}$ are learnable parameters and $ReLU$ is activation function.

Finally, we use a fully connected layer to adjust the dimension of output H_o.

$$Output = FC(H_o). \tag{28}$$

4.5. Optimization

The training process aims to minimize the error between the real traffic feature observations on all roads in the traffic graph and the predicted values. We use Y_t and $\hat{Y}_t$ to denote the real traffic feature observations and the predicted traffic feature values. The loss function of the DTC-TGCN model is shown as follows,

$$loss = ||Y_t - \hat{Y}_t|| + \lambda||L_{reg}||_{l2}, \tag{29}$$

where the first term is used to minimize the error between the real traffic feature observations and the predicted traffic feature values, the second term is an $L2$ regularization term, which helps avoid an overfitting problem, the L_{reg} denotes the set of all parameter matrix, and the λ is a hyperparameter.

5. Experiment and evaluations

This section evaluates the prediction performance of our proposed model DTC-STGCN compared with other baselines on two real-world datasets: the CC-taxi dataset and the SZ-taxi dataset. Since the two datasets are related to traffic flow and traffic speed, the experiments setting on the two datasets are slightly different.

5.1. Dataset descriptions and preprocessing

CC-taxi. This dataset consists of traffic trajectories collected from 2000 taxies in Changchun city, Jilin Province, China, from June. 1 to July. 10, 2017. We select 258 road segments in the central district of Changchun for experiments as Fig. 4. Some road segments in the dataset are short, and the observation values are always 0, we aggregate the 258 roads to 47 road segments. Hence, the input experimental dataset contains three parts: a $47 * 47$ adjacency matrix, which denotes the fixed connectivity between nodes at each timestamp, a feature matrix describing all roads' traffic flow time series, and a set of road information matrix of selected roads selected. Finally, we aggregate the traffic flow on each road every 15 min and divide the dataset: 30 days as the training set, 10 days as the test set. We predict traffic flow on this dataset.

SZ-taxi [39].This dataset was the taxi trajectories of Shenzhen from Jan. 1 to Jan. 30, 2015. We select 156 major roads of Luohu District as the study area. The experimental input data mainly includes two parts: a $156 * 156$ adjacency matrix, which describes the fixed connectivity of transportation network where each row represents one road and the values in the matrix represent the connectivity between the roads, and a feature matrix denoting the traffic speed time series of selected sections where rows are indexed by road sections and columns are indexed by the timestamps. Finally, we aggregate the traffic speed on each road every 15 min and divide the datasets: 20 days as the training set, 10 days as the test set. We predict traffic speed on this dataset.

5.2. Experiment setting

We implement the DTC-STGCN model based on the TensorFlow framework. We utilize the historical two hours urban traffic observations for predicting traffic features in the next, second, third, fourth timestamps, i.e., 15 min, 30 min, 45 min, 60 min. In other words, the k is set to 8, and the n is set as 1, 2, 3, 4. The input dimension of the feature value in each node is 1.

The road information contains lanes, directions, and roads, where the lanes are 1 to 6, the direction contains one-way and two-way roads, and the road types consist of the trunk, major roads, side roads, and highways. Hence, the dimension of the road information features representation c = 12 after one-hot embedding.

The optimization method utilized in our work is Adam optimizer. The hyperparameters of the DTC-STGCN model mainly include: learning rate, training epoch, drop rating, and the number of hidden layers. In the experiment, we manually adjust and set the learning rate to 0.001, the training epoch to 100, the input drop learning of LSTM is 0.25, and the output drop rating of LSTM is 0.3. The number of hidden units is an essential parameter of the DTC-STGCN model, as different hidden units may significantly affect the prediction precision. We conduct experiments with different hidden units and select the optimal value by comparing the predictions to choose the best value. In our work, the best hidden units of all neural network layers are chosen by using the Grid Search method. The range of hidden units is within [8,16,32] in GCN layers, and the range of hidden unit of LSTM is within [2,4,8,16,32]. Especially, the range of spatial features dimension a at each historical timestamp and the temporal feature dimension b is within [1,2,4,6,8]. Finally, the best hidden units of the first two graph convolution layers in Attention-based Graph Convolutional Networks are 32 and 64, the best hidden units in Attention-based

Spatio-Temporal Graph Convolutional Networks module are 32 and 16. The a and b are set to 6 and 4 respectively on CC-Taxi while 6 and 6 on SZ-Taxi as Fig. 5 and Fig. 6.

5.3. Baselines

We compare our models with the following several baselines:

- FC-LSTM [40]: Recurrent Neural Network with fully connected LSTM hidden units.
- GCN [41]: Graph Convolutional Network, which is a deep learning framework for capturing local spatial topology features of graphs.
- GAT [37]: Graph Attention Network, which assigns different weights to different neighbor nodes in convolution.
- STGCN [11]: Spatial–Temporal Graph Convolution Network, which captures spatial and temporal dependencies with complete convolutional structures for traffic forecasting.
- DCRNN [30]: Diffusion convolution recurrent neural network, which combines graph convolution networks with recurrent neural networks in an encoder-decoder manner.
- Graph Wavenet [31]: A model combines dilated casual convolution and graph convolutional network which utilizes adaptive adjacency matrix to mine implicit graph structure.
- DTC-STGCN-FR: Our proposed DTC-STGCN model where the adjacency matrix describes the feature value ratio between different roads at each timestamp.
- DTC-STGCN-FD: Our proposed DTC-STGCN model where the adjacency matrix denotes the absolute difference in feature observations of different roads at each timestmap.
- DTC-STGCN-TN: Our proposed DTC-STGCN model where the adjacency matrix describes the number of taxis transformed between road segments at each timestamp.

5.4. Metrics

To evaluate the performance of our proposed models, We introduce two commonly used performance metrics in this paper.

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{30}$$

- Rooted Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{31}$$

where n is the number of samples, y_i is the ground truth, and $\hat{y}_i$ is the prediction result. Specifically, RMSE and MAE are used to measure the prediction error: the smaller the value, the better the prediction effect.

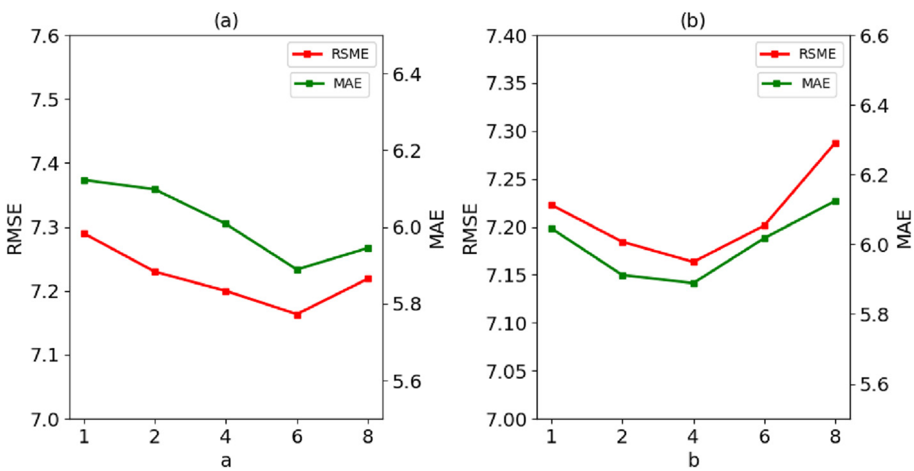

Fig. 5. Comparison of predicted performance of a and b under different hidden units on dataset CC-Taxi.

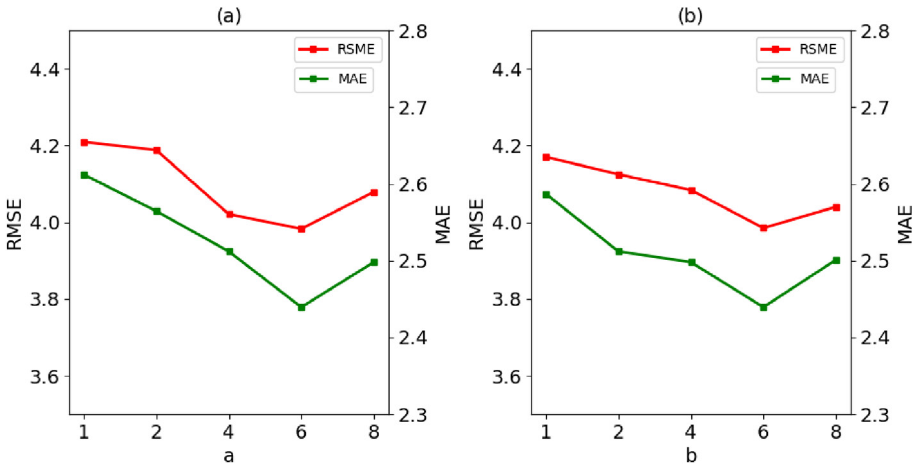

Fig. 6. Comparison of predicted performance of *a* and *b* under different hidden units on dataset SZ-Taxi.

5.5. Performance comparison and result analysis

In this section, we compare DTC-STGCN models' performance with other baselines on datasets CC-Taxi and SZ-Taxi.

5.5.1. Overall performance comparison

The Table 1 shows the average performance of DTC-STGCN models and other baseline methods for urban traffic forecasting on CC-Taxi and SZ-Taxi datasets over the next one hour.

It can be seen from Table 1 and Table 2 that our DTC-STGCN models achieve the best performance on both two datasets in terms of all evaluation metrics. We can observe that traditional deep learning methods (FC-LSTM and GCN) are usually not ideal because they only pay attention to temporal features or spatial patterns. Significantly, the GCN and GAT model performs worse, demonstrating the importance of temporal patterns in urban traffic. Among them, the models which simultaneously take both the temporal and spatial correlations into account, including STGCN, DCRNN, Graph Wavenet, and our models, are superior to the other models, which illustrates that it is necessary to capture the spatio-temporal features in urban traffic prediction domain. Our DTC-STGCN models achieve better performance than the previous state-of-the-art models, proving the advantages of our model in describing spatial–temporal correlations of urban traffic data.

5.5.2. Effect of dynamic adjacency matrix

To verify the models' advantage with the dynamic adjacency matrix, we design a baseline where the adjacency matrix is fixed with static values: 0 and 1, which only denotes the fixed connectivity of the urban traffic road network. The experimental settings are the same. The Table 3 and the Table 4 show the DTC-STGCN models, which are fed with dynamic adjacency matrix have better performance than it with fixed adjacency matrix. It indicates that the input of the dynamic adjacency matrix can precisely describe the dynamic spatio-temporal correlation among urban traffic data and is beneficial for forecasting urban traffic accurately.

Table 1
The average performance comparison of different approaches on dataset CC-Taxi under the RMSE, MAE and MAPE metrics for urban traffic forecasting.

Model	RMSE	MAE	MAPE
FC-LSTM	9.0559	7.8471	20.25%
GCN	9.7112	8.6945	22.46%
GAT	9.2960	8.4004	20.45%
STGCN	8.7349	7.5510	19.57%
DCRNN	8.1254*	6.8102*	19.21%*
Graph Wavenet	8.7349	7.5510	19.57%
DTC-STGCN-FR	7.2463	6.1073	17.54%
DTC-STGCN-FD	7.2917	6.0174	17.33%
DTC-STGCN-TN	**7.1635**+	**5.8891**+	**17.23%**+
Best + VS Baseline *	↑**11.8%**	↑13.5%	↑11.9%

Table 2
The average performance comparison of different approaches on dataset SZ-Taxi under the RMSE, MAE and MAPE metrics for urban traffic forecasting.

Model	RMSE	MAE	MAPE
FC-LSTM	5.0579	2.926	14.66%
GCN	5.4675	4.1811	15.75%
GAT	5.3486	3.7562	15.19%
STGCN	4.1711*	2.8079*	12.92%*
DCRNN	4.5343	3.3211	13.87%
Graph Wavenet	4.6942	3.4772	14.28%
DTC-STGCN-FR	3.9884	2.5533	12.35%
DTC-STGCN-FD	**3.9851**$^{+}$	**2.4393**$^{+}$	**12.15%**$^{+}$
DTC-STGCN-TN	-	-	-
Best $^{+}$ VS Baseline *	↑4.45%	↑13.1%	↑5.96%

Table 3
Average performance comparison of DTC-STGCN and its variant on CC-Taxi dataset to verify the effectiveness of the dynamic adjacency matrix module.

Model	RMSE	MAE	MAPE
DTC-STGCN-0/1	7.4645	6.125	18.15%
DTC-STGCN-FR	7.2463^{-}	6.1073	17.54%
DTC-STGCN-FD	7.2917	6.0174^{-}	17.33%$^{-}$
DTC-STGCN-TN	**7.1635**$^{+}$	**5.8891**$^{+}$	**17.23%**$^{+}$
Best $^{+}$ VS Variants $^{-}$	↑1.14%	↑2.13%	↑0.58%

Table 4
Average performance comparison of DTC-STGCN and its variant on SZ-Taxi dataset to verify the effectiveness of the dynamic adjacency matrix module.

Model	RMSE	MAE	MAPE
DTC-STGCN-0/1	4.0645	2.785	12.75%
DTC-STGCN-FR	3.9884^{-}	2.5533^{-}	12.35%$^{-}$
DTC-STGCN-FD	**3.9851**$^{+}$	**2.4393**$^{+}$	**12.15%**$^{+}$
Best $^{+}$ VS Variants $^{-}$	↑0.0017%	↑4.46%	↑1.61%

5.5.3. Effect of spatio-temporal attention mechanism

Besides, to verify the impact of the spatio-temporal attention mechanism proposed in this paper, we design a degraded version of DTC-STGCN-FD model that eliminates the spatio-temporal attention. The Table 5 and the Table 6 show our DTC-STGCN combined with the spatio-temporal attention mechanisms achieves better prediction results, proving the advantages of our spatio-temporal mechanism in capturing dynamic spatio-temporal changes of the urban traffic data.

5.5.4. Performance of long-term urban traffic forecasting

The Fig. 7 and the Fig. 8 show the changes of prediction performance of various methods as the prediction interval increases on dataset CC-Taxi and SZ-Taxi, respectively. Overall, as the prediction interval adds, the corresponding difficulty of prediction increases. Hence, the prediction errors also increase.

As can be seen from the Fig. 7 and the Fig. 8, the methods that only take the spatial correlations into account can achieve good results in the short-term prediction, such as GCN and GAT. However, with the increase of the prediction interval, their

Table 5
Average performance comparison of DTC-STGCN and its variant on CC-Taxi dataset to verify the effectiveness of the attention mechanism.

Model	RMSE	MAE	MAPE
DTC-STGCN(w/o att)	7.4652	6.2204	17.91%
DTC-STGCN-FD	7.2917	6.0174	17.33%
Performance Gain	↑2.32%	↑3.26%	↑3.18%

Table 6
Average performance comparison of DTC-STGCN and its variant on SZ-Taxi dataset to verify the effectiveness of the attention mechanism.

Model	RMSE	MAE	MAPE
DTC-STGCN(w/o att)	4.0312	2.6204	12.67%
DTC-STGCN-FD	3.9851	2.4393	12.15%
Performance Gain	↑1.14%	↑7.42%	↑4.10%

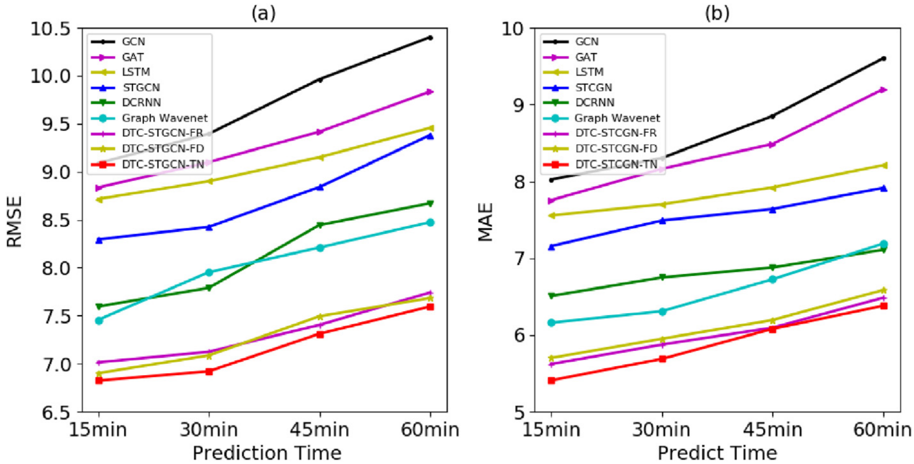

Fig. 7. The prediction results of different methods on CC-Taxi dataset as the prediction interval increases.

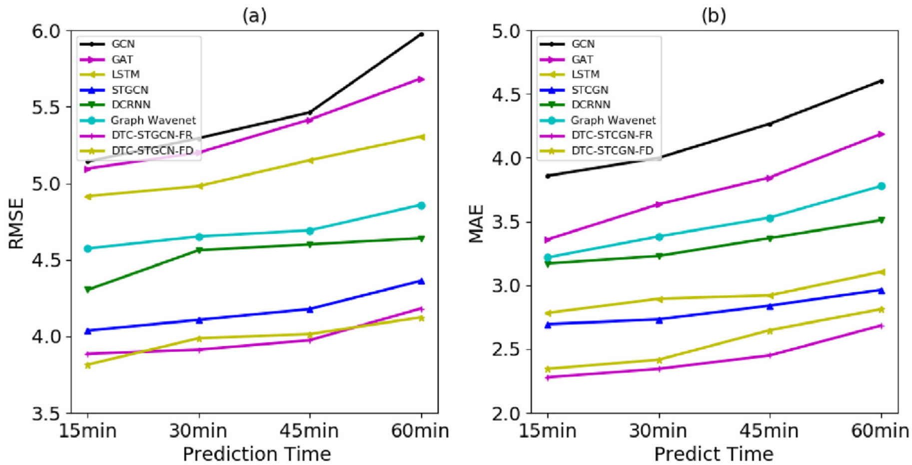

Fig. 8. The prediction results of different methods on SZ-Taxi dataset as the prediction interval increases.

prediction accuracy drops dramatically because the urban traffic prediction is a typical time-series study. By comparison, the performance of STGCN, DCRNN, Graph Wavenet, and our models drops slower than those methods. This is mainly because these models can simultaneously consider the spatial–temporal correlations, which are more critical in long-term prediction. Our DTC-STGCN models achieve the best prediction performance almost all the time. Especially in the long-term prediction, the performance differences between our models and other baselines are more significant, showing that our models can better learn the dynamic spatial–temporal patterns of urban traffic data.

5.5.5. Performance of urban traffic prediction in peak hours

Urban traffic prediction in the morning peak and evening rush hours is the most complicated task, owning to the intricate spatial and temporal patterns. The Fig. 9 and the Fig. 10 show the prediction result of different methods in the morning peak

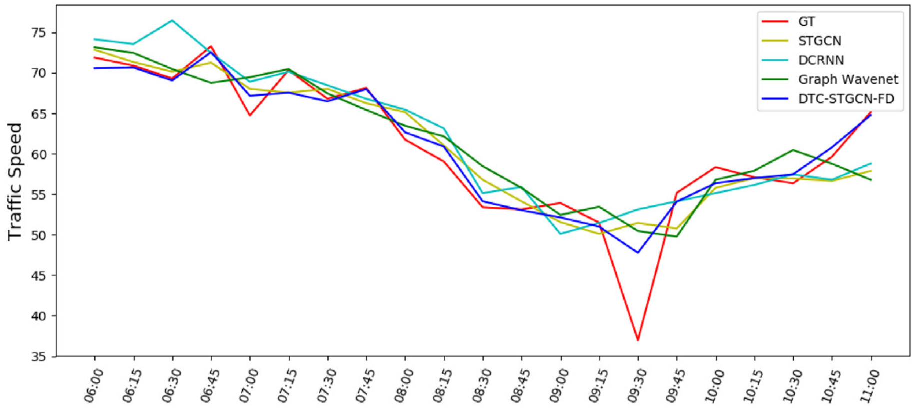

Fig. 9. Speed prediction in the morning peak hours on the dataset SZ-Taxi.

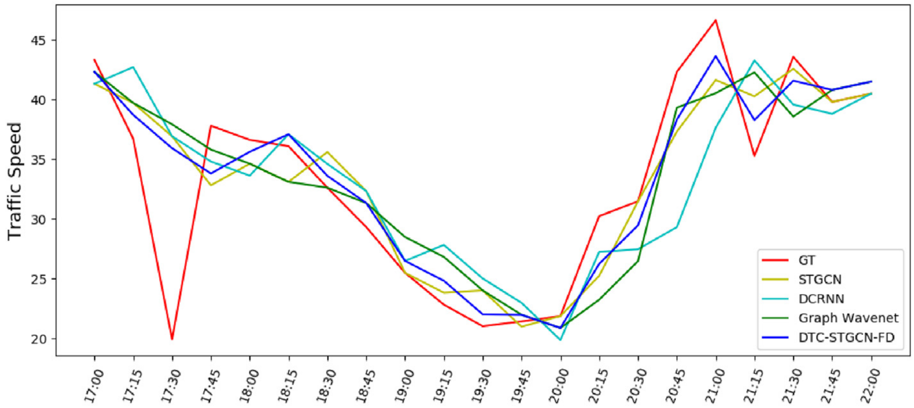

Fig. 10. Speed prediction in the evening rush hours on the dataset SZ-Taxi.

and evening rush hours on the dataset SZ-Taxi, respectively. It is easy to observe that our proposed DTC-STGCN-FD model is more sensitive to sudden changes in traffic conditions, which can capture the trend of rush hours more accurately than other methods. In addition, whether in the morning peak hours or evening rush hours, the prediction result of our model is the closest to the ground truth.

6. Conclusion remarks

Urban traffic prediction is a significant part of urban traffic research, which can serve many traffic applications. However, because of the dynamic, complex, nonlinear, and spatio-temporal patterns in urban traffic, there are many challenges in urban traffic forecasting. In this paper, we design a novel deep learning model for urban traffic forecasting and successfully apply it to urban traffic forecasting, which mainly consists of four components: Dynamic adjacency matrix for describing the dynamic spatio-temporal correlation, Attention and dynamic adjacency matrix-based graph convolutional networks module for capturing the dynamic spatial patterns, LSTM for learning the dynamic temporal dependencies and Attention-based graph convolutional network for learning dynamic spatio-temporal dependencies and road information features simultaneously. The experimental results demonstrate that our model outperforms other state-of-the-art methods on two real-world datasets, indicating that DTC-STGCN has advantages in capturing dynamic spatio-temporal patterns in predicting urban traffic.

In the future, we consider accompanying external features with the DTC-STGCN model to explore the impact of external features, like weather, point of interest, and social events, and find a hybrid method for calculating the adjacency matrix to predict urban traffic more accurately. Meanwhile, we will apply our model DTC-STGCN to deal with the time series problems in other domains.

CRediT authorship contribution statement

Yuanbo Xu: Conceptualization, Writing – original draft, Writing – review & editing, Data curation, Methodology. **Xiao Cai:** Methodology, Writing – original draft. **En Wang:** Validation. **Wenbin Liu:** Validation. **Yongjian Yang:** Supervision, Funding acquisition. **Funing Yang:** Data curation, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Natural Science Foundations of China under Grant No. 61772230, No.61976102, No.U19A2065, and No. 61972450, Natural Science Foundation of China for Young Scholars No. 61702215 and No. 62002132, China Postdoctoral Science Foundation No. 2020M681040 and Changchun Science, and Technology Development Project No.18DY005, and National Defense Science and Technology Key Laboratory Fund Project No. 61421010418 and Science Foundation of Jilin Province No. 20190201022JC and China National Postdoctoral Program for Innovative Talents No. BX20180140.

References

[1] M.R. Jabbarpour, H. Zarrabi, R.H. Khokhar, S. Shamshirband, K.R. Choo, Applications of computational intelligence in vehicle traffic congestion problem: a survey, Soft Comput. 22 (2018) 2299–2320.
[2] Z. Diao, D. Zhang, X. Wang, K. Xie, S. He, X. Lu, Y. Li, A hybrid model for short-term traffic volume prediction in massive transportation systems, IEEE Trans. Intell. Transp. Syst. 20 (2019) 935–946.
[3] B.M. Williams, L.A. Hoel, Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results, J. Transp. Eng. 129 (2003) 664–672.
[4] B.M. Williams, P.K. Durvasula, D.E. Brown, Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models, Transp. Res. Rec. 1644 (1998) 132–141.
[5] C. Chen, J. Hu, Q. Meng, Y. Zhang, Short-time traffic flow prediction with ARIMA-GARCH model, in: IEEE Intelligent Vehicles Symposium (IV), 2011, Baden-Baden, Germany, June 5–9, 2011, IEEE, 2011, pp. 607–612.
[6] R. Chen, C. Liang, W. Hong, D. Gu, Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, Appl. Soft Comput. 26 (2015) 435–443.
[7] Y. Wu, H. Tan, Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework, CoRR abs/1612.01022 (2016).
[8] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, Dnn-based prediction model for spatio-temporal data, in: S. Ravada, M.E. Ali, S.D. Newsam, M. Renz, G. Trajcevski (Eds.), Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2016, Burlingame, California, USA, October 31 - November 3, 2016, ACM, 2016, pp. 92:1–92:4.
[9] J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in: S.P. Singh, S. Markovitch (Eds.), Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA, AAAI Press, 2017, pp. 1655–1661.
[10] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, Y. Wang, Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction, Sensors 17 (2017) 818.
[11] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, in: J. Lang (Ed.), Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13–19, 2018, Stockholm, Sweden, ijcai.org, 2018, pp. 3634–3640.
[12] C. Chen, K. Li, S.G. Teo, X. Zou, K. Wang, J. Wang, Z. Zeng, Gated residual recurrent graph neural networks for traffic prediction, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019, AAAI Press, 2019, pp. 485–492.
[13] S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019, AAAI Press, 2019, pp. 922–929.
[14] C. Zheng, X. Fan, C. Wang, J. Qi, GMAN: A graph multi-attention network for traffic prediction, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020, AAAI Press, 2020, pp. 1234–1241.
[15] S. Cheng, F. Lu, Short-term traffic forecasting: A dynamic ST-KNN model considering spatial heterogeneity and temporal non-stationarity, in: N. Augsten (Ed.), Proceedings of the Workshops of the EDBT/ICDT 2018 Joint Conference (EDBT/ICDT 2018), Vienna, Austria, March 26, 2018, volume 2083 of CEUR Workshop Proceedings, CEUR-WS.org, 2018, pp. 133–140.
[16] X. Feng, X. Ling, H. Zheng, Z. Chen, Y. Xu, Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction, IEEE Trans. Intell. Transp. Syst. 20 (2019) 2001–2013.
[17] C. Wu, J. Ho, D. Lee, Travel-time prediction with support vector regression, IEEE Trans. Intell. Transp. Syst. 5 (2004) 276–281.
[18] S. Sun, C. Zhang, G. Yu, A bayesian network approach to traffic flow forecasting, IEEE Trans. Intell. Transp. Syst. 7 (2006) 124–132.
[19] T. Pamula, Impact of data loss for prediction of traffic flow on an urban road using neural networks, IEEE Trans. Intell. Transp. Syst. 20 (2019) 1000–1009.
[20] Y. Lv, Y. Duan, W. Kang, Z. Li, F. Wang, Traffic flow prediction with big data: A deep learning approach, IEEE Trans. Intell. Transp. Syst. 16 (2015) 865–873.
[21] B. Yang, S. Sun, J. Li, X. Lin, Y. Tian, Traffic flow prediction using LSTM with feature enhancement, Neurocomputing 332 (2019) 320–327.
[22] D. Kang, Y. Lv, Y. Chen, Short-term traffic flow prediction with LSTM recurrent neural network, in: 20th IEEE International Conference on Intelligent Transportation Systems, ITSC 2017, Yokohama, Japan, October 16–19, 2017, IEEE, 2017, pp. 1–6.
[23] J. Zhu, C. Huang, M. Yang, G.P.C. Fung, Context-based prediction for road traffic state using trajectory pattern mining and recurrent convolutional neural networks, Inf. Sci. 473 (2019) 190–201.

[24] R. Fu, Z. Zhang, L. Li, Using lstm and gru neural network methods for traffic flow prediction, in: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 324–328.
[25] B. Du, H. Peng, S. Wang, M.Z.A. Bhuiyan, L. Wang, Q. Gong, L. Liu, J. Li, Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction, IEEE Trans. Intell. Transp. Syst. 21 (2020) 972–985.
[26] J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in: S.P. Singh, S. Markovitch (Eds.), Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA, AAAI Press, 2017, pp. 1655–1661.
[27] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, Z. Li, Deep multi-view spatial-temporal network for taxi demand prediction, in: S.A. McIlraith, K.Q. Weinberger (Eds.), Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018, AAAI Press, 2018, pp. 2588–2595.
[28] H. Yao, X. Tang, H. Wei, G. Zheng, Z. Li, Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press, 2019, pp. 5668–5675.
[29] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Zheng, Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks, CoRR abs/1903.07789 (2019).
[30] Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, OpenReview.net, 2018.
[31] Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph wavenet for deep spatial-temporal graph modeling, in: S. Kraus (Ed.), Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019, ijcai.org, 2019, pp. 1907–1913.
[32] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, S. He, Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019, AAAI Press, 2019, pp. 890–897.
[33] F. Manessi, A. Rozza, M. Manzo, Dynamic graph convolutional networks, Pattern Recogn. 97 (2020).
[34] L. Bai, L. Yao, C. Li, X. Wang, C. Wang, Adaptive graph convolutional recurrent network for traffic forecasting, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6–12, 2020, virtual.
[35] Z. Pan, W. Zhang, Y. Liang, W. Zhang, Y. Yu, J. Zhang, Y. Zheng, Spatio-temporal meta learning for urban traffic prediction, IEEE Trans. Knowl. Data Eng. (2020).
[36] X. Zhang, C. Huang, Y. Xu, L. Xia, P. Dai, L. Bo, J. Zhang, Y. Zheng, Traffic flow forecasting with spatial-temporal graph diffusion network, in: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2–9, 2021, AAAI Press, 2021, pp. 15008–15015.
[37] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, CoRR abs/1710.10903 (2017).
[38] C. Song, Y. Lin, S. Guo, H. Wan, Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020, AAAI Press, 2020, pp. 914–921.
[39] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, H. Li, T-GCN: A temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. 21 (2020) 3848–3858.
[40] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pp. 3104–3112.
[41] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, CoRR abs/1609.02907 (2016).