Research Article

Road Travel Time Prediction Based on Improved Graph Convolutional Network

Miao Xu and Hongfei Liu

College of Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Hongfei Liu; hongfeiliu@jlu.edu.cn

Received 31 August 2021; Revised 21 October 2021; Accepted 9 November 2021; Published 29 November 2021

Academic Editor: Nicola Biccocchi

Copyright © 2021 Miao Xu and Hongfei Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Travel time prediction is playing an increasingly important part in advanced traveler information system (ATIS), which is of great significance to alleviate urban traffic congestion. Although graph convolutional networks have been widely used in road network traffic prediction, spatiotemporal dynamic modeling of urban traffic is still an intractable task. In this study, we propose an improved graph convolutional network (IGC-Net) for travel time prediction. Specifically, we design a modified adjacency matrix by fusing distance and correlation matrix with original adjacency matrix to capture spatial dynamic feature. We then establish three components based on temporal property to capture recent, daily-periodic, and weekly periodic correlations. The comparison experiments with baseline models and variants on a real-world dataset in Beijing are conducted. The results show that the IGC-Net outperforms baseline models in different prediction horizons and has stronger robustness for dynamic traffic prediction.

1. Introduction

In recent years, the problem of urban traffic congestion has become more and more serious. The duration of commuting directly determines people's travel experience, which is not only an important index to measure the efficiency of urban operation but also an important factor affecting the life quality of residents. Road travel time can directly reflect the traffic congestion state of the road segment, which is an important basis for the development of intelligent transportation system (ITS) [1, 2]. Travel time prediction is of great significance and application value to both traffic users and traffic managers.

The study of traffic forecasting has been developing over the past few decades. Initially, the researchers used statistical methods including historical average (HA) and autoregressive integrated moving average (ARIMA) [3, 4] to predict temporal traffic parameters. However, simple statistical models are difficult to capture the nonlinear relationship of traffic data. Thanks to the progress of computer technology and the diversification of data acquisition system, neural networks modeling and multisource

information fusion are widely used in traffic prediction research, which greatly improves the accuracy of traffic prediction. Wu et al. [5] used the support vector regression model (SVR) for traffic prediction. Tian and Pan [6] utilized long short-term memory network (LSTM) to predict traffic flow, which demonstrated the effectiveness of recurrent neural network (RNN) to forecast time-series data. However, the traditional machine learning model still has deficiencies in capturing periodic features and selecting model parameters. Some researchers further proposed hybrid models to predict traffic parameters [7–9]. For example, Li et al. [10] established a model based on ensemble empirical mode decomposition and random vector functional link network for travel time prediction on highway network. They also proposed another model based on a deep belief network, which is optimized by the multiobjective particle swarm algorithm [11]. Moreover, traffic prediction research is also extended to the spatial scope. The researchers first analyzed the characteristics of spatiotemporal data. For instance, Zhang et al. [12] divided the city into a grid to explore the spatial distribution and correlation of the cellular traffic and analyzed temporal dynamics between different

cells using autocorrelation coefficient. Based on the analysis of temporal and spatial characteristics, the convolutional neural network (CNN) was widely applied [13, 14] to capture traffic features as images. However, due to the real traffic network is complex, the standard convolution with Euclidean grid is no longer suitable for general graphs.

In recent years, there are two ideas to explore how to generalize convolutional neural network into structured data forms. One is to expand the spatial definition of convolution network in the spatial domain [15], and the other is to process it by graph Fourier transform in spectral domain [16]. The former directly defines the convolution operation on the connection relationship of each node, which is more similar to the traditional convolution neural network. The latter is to realize convolution operation on the topological graph with the help of the spectral graph theory. The graph neural network (GNN) [17] has been widely used in traffic prediction. In addition, deep learning techniques can automatically extract features of multisource data [18] and model more complex spatial and temporal traffic patterns in various traffic scenarios. The sequence-to-sequence (Seq2Seq) model with encoder-decoder structure [19, 20] combined with graph convolutional network (GCN) which has been widely used to construct spatiotemporal prediction models. For instance, Yu et al. [21] adopted the method of spectral graph convolution combined with gated recurrent neural network to obtain spatiotemporal features. Guo et al. [22] proposed an attention-based spatial-temporal graph convolutional network (ASTGCN) model to realize the traffic flow forecasting. Nevertheless, due to the complexity and dynamics of the actual traffic network, the traditional adjacency matrix cannot effectively capture the time-varying spatial dynamic characteristics.

To solve the above problems, an improved graph convolutional network is proposed in this paper to improve the accuracy of travel time prediction. The primary contributions of this paper are as follows:

- (i) We propose a modified adjacency matrix to better capture the spatial features by integrating dynamic weight information between road segments. The main idea is to construct distance weight matrix and correlation weight matrix, respectively, based on geographical location attributes and dynamic traffic information.
- (ii) According to the temporal property, we establish the recent, daily, and weekly component to model the temporal dependencies. Furthermore, we use the same improved graph convolutional network in three components to capture spatiotemporal characteristics.
- (iii) By using real-world datasets, we conduct baseline model comparison and ablation experiments to evaluate our model performance. The prediction results demonstrate the superior performance of our proposed model.

The remainder of this paper is organized as follows. In Section 2, we present the basic concepts and problem

formulation and describe the concrete modeling process. In Section 3, we introduce the experimental environment and setting. The experimental results are analyzed in detail in Section 4. Finally, Section 5 summarizes the study and looks forward to the prospects for the future.

2. Methodology

2.1. Preliminaries. In this section, we first describe the notations of variables and formalize the traffic prediction problem.

2.1.1. Traffic Network Topology. The road network is defined as a graph $G = (V, E, \mathbf{A})$, where V is the set of N vertices (i.e., road segments) $v_i \in V$ and E is edges between different vertices $(v_i, v_j) \in E$. The adjacency matrix $\mathbf{A}$ reflecting the connectivity between road segments can be indicated as $\mathbf{A} = (a_{ij})_{N \times N} \in \mathbb{R}^{N \times N}$. $a_{ij} = 1$ if node v_i and v_j are accessible; otherwise, $a_{ij} = 0$.

2.1.2. Travel Time. The travel time of each road segment is normalized to unit length time (s/m) considering the influence of unequal segment length. The travel time of the whole traffic network G at t th time slot is defined as $X_t = (x_t^1, x_t^2, \dots, x_t^N)^T \in \mathbb{R}^{N \times F}$, where F denotes the feature of vertices. Therefore, $\chi = (X_t, X_{t-1}, \dots, X_{t-p+1})^T \in \mathbb{R}^{N \times F \times p}$ denotes all node values over the first p time steps of input.

2.1.3. Meta Features. In this study, we aggregate traffic readings into 5 minutes' window. As shown in Figure 1, time-series traffic data have specific temporal characteristics, including recent, daily-periodic, and weekly periodic correlations. We set time labels' vector $l_1 = \{0, 1, \dots, 287\}$, $l_2 = \{0, 1, \dots, 6\}$, and $l_3 = \{0, 1\}$ to represent time-of-day, day-of-week, and weekday-or-weekend, respectively. For instance, 23:55 is labeled as $l_1 = 287(12 * 23 + 11)$, Friday as $l_2 = 4$, and weekdays as $l_3 = 0$. Hence, at each time slot t , the exogenous variable E_t , which is an additional inputs to the model, can be composed of the above time attribute characteristics.

2.1.4. Historical Average Travel Time. Historical traffic information can reflect the trend of daily traffic conditions. The historical average travel time at t th time slot of road segment i is denoted by $x_{t,h}^i$.

2.1.5. Problem Definition. The task of travel time forecasting is to use the past traffic observations to predict the future value of each road segment in a certain period. Given traffic network graph G , the prediction problem is formulated as

$$Y_{t+1}, \dots, Y_{t+q} = \arg \max_{X_{t+1}, \dots, X_{t+q}} \Pr(X_{t+1}, \dots, X_{t+q} | X_t, X_{t-1}, \dots, X_{t-p+1}; G), \quad (1)$$

where Y_{t+q} denotes the predicted value of the q th time step, $\{X_t, X_{t-1}, \dots, X_{t-p+1}\}$ represents the historical values of the

FIGURE 1: Temporal correlations. Take two weeks of road travel time data as an example.

first p time steps, and $\Pr(\cdot|\cdot)$ is the conditional probability function.

2.2. Improved Graph Convolutional Network. In this study, the urban traffic network is regarded as a graph structure, and the features of each node are the signals on the graph [23]. In spectral convolution, the graph is represented by its corresponding Laplace matrix.

2.2.1. Spectral Graph Convolution. As it is difficult to express meaningful translation operator in the node domain [17], the spectral convolution on graphs is defined as the multiplication of a signal $X \in \mathbb{R}^N$ (a scalar of all nodes at a time slot) with a filter $g_\theta = \text{diag}(\theta)$ parameterized by $\theta \in \mathbb{R}^N$ in the Fourier domain [24]. According to the convolution theorem and Fourier transform of graphs, the spectral graph convolution is defined as

$$g_\theta *_G X = g_\theta(\mathbf{L})X = g_\theta(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)X = \mathbf{U}g_\theta(\mathbf{\Lambda})\mathbf{U}^T X, \quad (2)$$

where $*_G$ represents a graph convolution operation, $\mathbf{U} \in \mathbb{R}^{N \times N}$ is the Fourier basis composed of eigenvectors, and $\mathbf{\Lambda} \in \mathbb{R}^{N \times N}$ is a diagonal matrix of eigenvalues.

However, when the scale of the graph is large, it is computationally expensive to decompose the eigenvalue of the Laplace matrix in equation (2). Hammond et al. [25] used Chebyshev polynomials to circumvent this problem. Furthermore, a layerwise linear model can be built by stacking multiple localized graph convolution layers with 1-order approximation of graph Laplacian [26].

2.2.2. Modified Adjacency Matrix. The spatial dependence in the actual traffic network is complex, which is affected by the geospatial distance and dynamic traffic flow. We further construct distance matrix and correlation matrix to capture spatial attention.

Distance matrix $\mathbf{W}^d$: we first calculate the shortest distance d_{ij} between node v_i and v_j by Dijkstra algorithm and then use the reciprocal of the distance to represent the weight between two vertices:

$$w_{ij} = \begin{cases} 1, & i = j, \\ \frac{1}{d_{ij}}, & i \neq j \text{ and } d_{ij} \leq \delta, \\ 0, & i \neq j \text{ and } d_{ij} > \delta, \end{cases} \quad (3)$$

$$\mathbf{W}^d = \begin{bmatrix} 1 & w_{12} & \cdots & w_{1N} \\ w_{21} & 1 & & w_{2N} \\ \vdots & & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & 1 \end{bmatrix}, \quad (4)$$

where δ is set to 1000 to control the sparsity of the matrix.

Correlation matrix $\mathbf{W}^r$: due to the daily commuting rules, urban traffic has significant periodic characteristics. Therefore, based on the Pearson correlation coefficient method, we utilize historical average value $\bar{x}_{t,h}^i$ to obtain the dynamic correlation between road segments:

$$r_{t,ij} = \begin{cases} \frac{\sum_{\tau=t-p+1}^t (x_{\tau,h}^i - \bar{x}_{\tau,h}^i)(x_{\tau,h}^j - \bar{x}_{\tau,h}^j)}{\sqrt{\sum_{\tau=t-p+1}^t (x_{\tau,h}^i - \bar{x}_{\tau,h}^i)^2} \sqrt{\sum_{\tau=t-p+1}^t (x_{\tau,h}^j - \bar{x}_{\tau,h}^j)^2}}, & r_{ij} \geq 0.2, \\ 0, & r_{ij} < 0.2, \end{cases} \quad (5)$$

$$\mathbf{W}^r = \begin{bmatrix} 1 & r_{t,12} & \cdots & r_{t,1N} \\ r_{t,21} & 1 & & r_{t,2N} \\ \vdots & & \ddots & \vdots \\ r_{t,N1} & r_{t,N2} & \cdots & 1 \end{bmatrix}, \quad (6)$$

where the correlation coefficient $r_{ij} < 0.2$ means that the two vertices are almost uncorrelated.

Based on the above, the modified adjacency matrix is defined as

$$A' = A \odot W^d \odot W^r, \quad (7)$$

where $\odot$ is the Hadamard product operator.

2.2.3. Improved Graph Convolution (IGC). With the modified adjacency matrix A' , the calculation formula between multiple graph convolutional layers is as follows:

$$H^{(l+1)} = f\left(\tilde{D}^{l-1/2} \tilde{A} \tilde{D}^{l-1/2} H^{(l)} \Theta^{(l)}\right). \quad (8)$$

To capture a larger range of spatial correlations, we leverage the residual learning to our model, which has been proved to achieve better results in deep network training [21]. The residual units are denoted as

$$H^{(l+1)} = H^{(l)} + f\left(\tilde{D}^{l-1/2} \tilde{A} \tilde{D}^{l-1/2} H^{(l)} \Theta^{(l)}\right), \quad (9)$$

where $\tilde{A} = A' + I_N$ is the modified adjacency matrix added self-connections and $\tilde{D}_{ii}' = \sum_j \tilde{A}_{ij}'$. $H^{(l)}$, $H^{(l+1)}$, and $\Theta^{(l)}$ are the input, output, and a trainable matrix of the l th layer, respectively. $f(\cdot)$ denotes a nonlinear activation function, e.g., ReLU.

2.3. Framework of the Prediction Model. Figure 2 presents the overview of our proposed prediction model, which is mainly composed of four components to model temporal dependencies and spatial correlations. IGC_1, IGC_2, and IGC_3 have the same network structure, which can capture recent, daily-periodic, and weekly periodic correlations, respectively. According to the definitions in Section 2.1, the sampling frequency of historical observations is 288 times per day, the current time slot is t , and the prediction horizon is q in this study. Therefore, the input of the above three components can be defined as follows, where l_r , l_d , and l_w indicate the lengths of three parts of data.

- (i) Recent: $\chi_r = (X_{t-l_r+1}, X_{t-l_r+2}, \dots, X_t)^T$
- (ii) Daily: $\chi_d = (X_{t-288(l_d/q)+1}, \dots, X_{t-288(l_d/q)+q}, X_{t-288(l_d/q-1)+1}, \dots, X_{t-288(l_d/q-1)+q}, X_{t-288+1}, \dots, X_{t-288+q})^T$
- (iii) Weekly: $\chi_w = (X_{t-7 \times 288(l_w/q)+1}, \dots, X_{t-7 \times 288(l_w/q)+q}, X_{t-7 \times 288(l_w/q-1)+1}, \dots, X_{t-7 \times 288(l_w/q-1)+q}, X_{t-7 \times 288+1}, \dots, X_{t-7 \times 288+q})^T$

Then, the result Y_{IGC} is obtained from the output of each component by means of parametric-matrix-based fusion:

$$Y_{IGC} = W_1 \odot Y_{IGC-1} + W_2 \odot Y_{IGC-2} + W_3 \odot Y_{IGC-3}, \quad (10)$$

where W_1 , W_2 , and W_3 denote trainable matrices.

In the rightmost component of Figure 2, the metadata are transformed into the binary vector by one-hot encoding and the fully connected neural network is utilized to process binary feature. Finally, by integrating Y_{IGC} with the output Y_{meta} , we yield the predicted value by means of Tanh function:

$$Y_{t+q} = \text{Tanh}(Y_{IGC} + Y_{meta}). \quad (11)$$

We leverage minimizing mean squared error to calculate the loss between predicted value and historical observation. The loss function is expressed as

$$L(\delta) = \|X_{t+q} - Y_{t+q}\|_2^2, \quad (12)$$

where δ represents all learnable parameters in the model.

3. Experiments

3.1. Data. The data used in our experiments includes road network geographic information (as shown in Figure 3) and road segments travel time data from October 1 to December 31, 2019, in Beijing, for a total of 13 weeks. In the datasets' division, we divide the data of the first ten weeks into nonoverlapping training set and verification set samples on the timeline and take the data of the last two weeks as the test set.

3.2. Experimental Settings. The modeling process is carried out on the Anaconda3 using Python. The experiments are performed on a server with an Intel Core i9-9900 KF 3.60 GHz CPU and 32-GB RAM. Furthermore, a GPU with 8G memory is used to accelerate the model learning process.

In the experiments, the model is trained by the Adam optimizer [27], where the learning rate is set to 0.001 and the batch size is 64. Taking the mean square error (MSE) as the objective function, the early stop technique is applied on the verification set to avoid overfitting. For our proposed model, we use the data of previous 60 minutes to predict single-step (5 min) and multistep (15 min and 30 min) traffic in the future. Take 30 min as an example, the prediction horizon $q = 6$, and the lengths of three parts are set as $l_r = 12$, $l_d = 18$, and $l_w = 12$.

3.3. Comparison Algorithms and Metrics. In order to verify the performance of the proposed model, we construct five baseline models for comparison, including the statistical model (HA and ARIMA), the traditional neural network model (LSTM), and the graph convolutional network-based deep learning model (DCRNN, STGCN, and graph WaveNet). Also, IGC-Net_E is utilized as variant of IGC-Net to compare the influence of exogenous variables, respectively.

- (i) HA: we predict the future travel time by the average value of travel time in the corresponding time interval.
- (ii) ARIMA: the autoregressive integrated moving average method has been widely used in time-series forecasting.
- (iii) LSTM: it is a variant model of RNN [28]. Here, the model has two hidden layers with 64 units.
- (iv) DCRNN: we built the diffusion convolutional recurrent neural network according to the open source code provide by [29].
- (v) STGCN: it stands for spatial-temporal graph convolutional network. For the sake of fairness, the

FIGURE 2: Architecture of IGC-Net.

hyperparameters are set according to the original model [21].

- (vi) Graph WaveNet [30]: graph WaveNet captures spatial-temporal correlations by combining graph convolution with dilated casual convolution.
- (vii) IGC-Net_T: it is a variant of our proposed model without the classification modeling of temporal property. That is, the input data are not distinguished by recent, daily-periodic, and weekly periodic correlations.
- (viii) IGC-Net_E: It is a variant of our proposed model without metadata input.

Furthermore, the prediction results of models are evaluated by MAE, MAPE, and RMSE. The calculation formula is as follows:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|, \quad (13)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \frac{|X_i - Y_i|}{|Y_i|}, \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}, \quad (15)$$

where m is the number of test samples and X_i and Y_i are the i th real and prediction value, respectively.

4. Results and Discussion

4.1. Model Comparisons. In this section, we use the data of previous 60 minutes to predict the travel time of the next 15 minutes and 30 minutes, i.e., the input step is set to 12, and the output step is set to 3 and 6, respectively. Table 1 shows the prediction performance of different algorithms. It can be seen that our proposed model outperforms baseline models in different prediction horizons. The prediction accuracy of the statistical model tends to be worse, which is caused by the model's inability to model nonlinear and complex traffic. Besides, LSTM only models the time dependence without considering the spatial correlation, and the prediction performance is also greatly reduced.

FIGURE 3: Map of experimental road network.

Compare our IGC-Net with the graph convolutional network-based model, all of which model the spatial and temporal characteristics. However, our model performs the best. The reasons are as follows. On the one hand, we consider the daily-periodic and weekly periodic correlations of time-series data; on the other hand, we modify the adjacency matrix of graph convolution.

As can be seen from Figures 4 and 5, the deep learning model has better prediction performance compared with the traditional neural network model, but consumes longer calculation time. DCRNN has the lowest running efficiency because of the time-consuming sequence calculation in recurrent neural network. STGCN has

relatively poor prediction performance, despite the short running time. On the basis of integrating spatial dynamic features, the running time of IGC-Net is only slightly higher than those of Graph WaveNet. As the prediction horizon increases, our model becomes more advantageous.

In addition, IGC-Net_E removes exogenous variables, which reduces prediction performance by 1.13%. The prediction performance of IGC-Net_T decreases more significantly due to not considering the temporal dependencies. However, IGC-Net_T outperforms LSTM, which proves the importance of considering spatial dynamic correlation in traffic prediction.

TABLE 1: Performance comparison of prediction models.

Model	15 min			30 min		
	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE
HA	0.0401	25.37	0.0689	0.0431	28.48	0.0703
ARIMA	0.0394	23.96	0.0663	0.0419	25.06	0.0691
LSTM	0.0365	20.04	0.0624	0.0400	22.97	0.0652
IGC-Net_T	0.0326	18.57	0.0643	0.0387	21.18	0.0640
STGCN	0.0327	18.41	0.0638	0.0375	20.03	0.0653
DCRNN	0.0301	16.78	0.0611	0.0336	18.37	0.0641
Graph WaveNet	0.0293	16.21	0.0602	0.0317	17.11	0.0634
IGC-Net_E	0.0287	14.83	0.0596	0.0298	15.97	0.0610
IGC-Net	0.0269	13.70	0.0527	0.0292	15.49	0.0601

FIGURE 4: Comparison of MAE.

FIGURE 5: Comparison of running time.

4.2. Effect of Input Sequence Length and Prediction Horizon. For a time-series model, data acquisition and prediction requirements directly affect the final prediction performance. Therefore, we conduct sensitivity analysis on the input sequence length and prediction horizon of the proposed model.

As shown in Figure 6, for each input sequence length, the prediction performance decreases as the prediction horizon increases. Intuitively, increasing the prediction horizon will make the model require more input sequence to capture temporal correlation information. However, the prediction results become complex when we increase the input sequence length in two datasets. We find that it cannot meet the requirement of long-term prediction when the input is 3 steps, especially the prediction horizon of 30 min; however, when the input is increased to 12 steps, the prediction result also becomes worst. The reason is that too much historical information causes data redundancy, which will weaken the temporal correlation and reduce prediction accuracy.

Therefore, to learn more relevant historical information for accurate prediction, it is beneficial to appropriately increase the input sequence length. However, too much or too little input will reduce model performance.

4.3. Effect of Spatial Dynamic Modeling. To verify the performance of the improved graph convolutional network, we use a variant GC-Net, which only contains the adjacency matrix for spatial modeling without considering the spatial dynamic correlation, to carry out comparative experiments.

Table 2 shows the comparison results of two models in 5 minutes, 15 minutes, and 30 minutes ahead forecasting. Furthermore, we can more intuitively analyze the comparison results of model performance from Figure 7. IGC-Net achieves better performance in different prediction horizons, which demonstrates the effectiveness of the modified adjacency matrix in spatial dynamic modeling.

Further, to analyze the prediction performance of different road segments more intuitively, the results on December 20-21 are taken as an example. As shown in Figure 8, the GC-Net has a poorer ability to capture the dynamic change of travel time, especially in peak periods.

FIGURE 6: Comparison of different input sequence lengths and prediction horizons.

TABLE 2: Performance comparison of GC-Net and IGC-Net in different prediction horizons.

Prediction horizon	Metrics	GC-Net	IGC-Net
5 min	MAPE	12.97%	10.08%
	RMSE	0.0523	0.0455
15 min	MAPE	15.02%	12.31%
	RMSE	0.0549	0.0511
30 min	MAPE	22.15%	18.77%
	RMSE	0.0674	0.0616

FIGURE 7: Performance comparison of GC-Net and IGC-Net.

(a)

(b)

(c)

FIGURE 8: Comparison of prediction results between IGC-Net and GC-Net in different road segments. (a) N.3th Ring Rd Middle. (b) Anding Rd. (c) G6 Fwy.

5. Conclusions

In this study, we propose an improved graph convolutional network (called IGC-Net) for travel time prediction. We construct distance weight matrix and correlation weight matrix, respectively, to modify the adjacency matrix of traditional GCN. Furthermore, we establish recent, daily, and weekly component to model the temporal dependencies according to the temporal property. Our proposed model can not only capture the static spatiotemporal characteristics but also realize the modeling of spatial dynamic correlation. The comparison experiments are carried out using Beijing road network traffic data. The results prove that our proposed model baseline models and the modified adjacency matrix can significantly improve the model's prediction accuracy.

For future work, we will investigate the applicability of our model to other urban traffic forecasting tasks and further explore the method of dynamic spatial modeling in graph convolutional network.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by grants from National Key Research and Development Program of China (no. 2018YFB1601600).

References

- [1] M. Abdollahi, T. Khaleghi, and K. Yang, "An integrated feature learning approach using deep learning for travel time prediction," *Expert Systems with Applications*, vol. 139, 2019.
- [2] D. Wang, J. Zhang, W. Cao, J. Li, and Y. Zheng, "When will you arrive? Estimating travel time based on deep neural networks," in *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.
- [3] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [4] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time series data by using box-jenkins techniques," *Transportation Research Record*, vol. 773, pp. 1–9, 1979.
- [5] C.-H. Wu, C. C. Wei, D. C. Su, M. H. Chang, and J. M. Ho, "Travel time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.
- [6] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proceedings of the IEEE International Conference Smart City/SocialCom/SustainCom (SmartCity)*, pp. 153–158, Chengdu, China, December 2015.
- [7] L. Li, J. Zhu, H. Zhang, H. Tan, B. Du, and B. Ran, "Coupled application of generative adversarial networks and conventional neural networks for travel mode detection using gps data," *Transportation Research Part A: Policy and Practice*, vol. 136, pp. 282–292, 2020.
- [8] Y. Lin, L. Li, H. Jing, B. Ran, and D. Sun, "Automated traffic incident detection with a smaller dataset based on generative adversarial networks," *Accident Analysis & Prevention*, vol. 144, Article ID 105628, 2020.
- [9] C. Zhang, S. Dang, B. Shihada, and M.-S. Alouini, "Dual attention-based federated learning for wireless traffic prediction," in *Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pp. 1–10, May 2021.
- [10] L. Li, X. Qu, J. Zhang, H. Li, and B. Ran, "Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network," *Applied Soft Computing*, vol. 73, pp. 921–932, 2018.
- [11] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowledge-Based Systems*, vol. 172, pp. 1–14, 2019.
- [12] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1656–1659, 2018.
- [13] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, 2019.
- [14] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [15] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proceedings of the ICML*, pp. 2014–2023, New York, NY, USA, June 2016.
- [16] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, pp. 14–16, Banff, AB, Canada, April 2014.
- [17] K. Guo, Y. Hu, Z. Qian, Y. Sun, J. Gao, and B. Yin, "Optimized graph convolution recurrent neural network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 1–10, 2020.
- [18] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: an overview," *Information Fusion*, vol. 53, pp. 123–133, 2020.
- [19] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3104–3112, Montreal Canada, December 2014.
- [20] L. N. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, "An effective spatial-temporal attention based neural network for traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 12–28, 2019.
- [21] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3634–3640, Stockholm, Sweden, July 2018.

- [22] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 922–929, 2019.
- [23] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
- [24] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3844–3852, Barcelona, Spain, September 2016.
- [25] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the ICLR*, Toulon, France, April 2017.
- [27] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Banff, AB, Canada, April 2014.
- [28] H. Sepp and S. Jürgen, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [29] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," 2017, <https://arxiv.org/abs/1707.01926>.
- [30] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 1907–1913, Macao, China, August 2019.