2019 International Conference on Data Mining Workshops (ICDMW)

Traffic Flow Prediction Based on Deep Neural Networks

Yashan Wang
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
18125063@bjtu.edu.cn

Yuchun Guo
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
ychguo@bjtu.edu.cn

Zhongrui Wei
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
17125063@bjtu.edu.cn

Yin Huang
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
16211145@bjtu.edu.cn

Xiang Liu
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
16111017@bjtu.edu.cn

***Abstract*—Traffic forecasting plays an important role in intelligent transportation system. Accurate forecasting enables appropriate travel suggestions for commuters, and can further benefit transportation management. Recent research has led to significant progress in spatial-temporal feature extraction with the help of deep learning methods, such as graph convolutional neural network (GCN) and long short-term memory (LSTM) network. However, simply applying these methods to the specific situation of road network is not enough, hence in this paper, we propose an optimized spatial-temporal traffic flow prediction model. With line-graph transformation introduced into the construction of road traffic topology, the application of GCN has a more efficient and effective ability to learn spatial features in the road network compared to conventional convolutional neural network (CNN) operations; and the utilization of LSTM with weather and periodic features introduced as additional information makes the model more powerful in temporal feature extraction. Experimental results on real-world large-scale dataset verify the proposed models ability to capture spatial and temporal properties, showing the advantages of our model beyond several baseline models.**

***Index Terms*—spatial-temporal correlation, deep learning, traffic flow forecast**

I. Introduction

With the development of society, traffic problems such as traffic congestion have become increasingly prominent. As one of the effective means to solve traffic problems, intelligent transportation systems (ITS) have become more and more popular [1]–[3]. In this paper we study the problem of traffic flow forecasting, a key component of ITS.

Traffic prediction is challenging mainly due to that it needs to make good use of the temporal, spatial and other features in the road network, while most of the existing traffic flow prediction methods [4]–[6] are solely based on the time series characteristics, which will result in the poor accuracy of the prediction. Some other prediction methods [7], [8] comprehensively utilize the spatial-temporal features, nevertheless the spatial feature extraction method is not perfectly designed for traffic flow forecasting problem, and this insufficiency leads to the low prediction ability of the methods. Hence an effective approach used for traffic flow forecasting is in need, which not only captures the spatial-temporal features to make accurate predictions, but also contains methods that specifically fit for spatial information extractions under general road network structure.

Due to the powerful ability of deep learning in feature learning and representation, methods with deep learning models have been widely researched and applied in traffic area. Unlike problems where the coordinates of the underlying data representation has a grid structure [15], data used in traffic prediction are always presented in the form of non-Euclidean structure, and road network is more appropriate to be regarded as a graph which consists of nodes and edges.Through convolutional operations on graph, spatial features of the traffic flow can be better grasped [12], [14].

In this paper, we propose a traffic flow prediction model based on deep learning. We utilize the line-graph transformation [9] to construct road traffic topology, a combination of graph convolutional neural network (GCN) and Long Short-Term Memory (LSTM) to extract spatial-temporal features. Additionally, we adopted periodic feature and weather information to further improve the prediction accuracy. When evaluated on large-scale real-world traffic dataset, our approach obtained significantly better performance than baselines. In summary:

1) A spatial feature extraction method used for traffic flow forecasting is proposed and with the application of graph convolutional neural network, it has a better grasp of the road network characteristics;
2) Combining the spatial-temporal features extracted with deep learning methods and other related features, a novel traffic flow prediction model is proposed, and

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00040

compared with state-of-the-art baselines, the proposed model achieves higher predicting accuracy.

II. Related works

Most of the traditional traffic flow forecasting methods consider the traffic flow forecasting task a time series forecasting problem, and tend to reshape the traffic flow data into time series data. Base on that, methods such as Auto-Regressive Integrated Moving Average (ARIMA) model are proposed and remain popular. However, these data-driven methods are incapable to provide satisfying results unless with sufficient data, although models using machine learning methods such as HMM model [5] alleviate the problem to some extent, both simple time series models and machine learning methods strongly rely on the stationarity assumption, which is often violated by the traffic data [10]. Fortunately, deep learning models are proposed and have delivered new promise for traffic flow forecasting problem.

Specifically, Deep Neural Network (DNN) is widely used to complete forecasting tasks. Wu et al. [6] proposed DNN-BTF model utilizing daily and weekly features to improve the accuracy of traffic flow forecasting. Models based on Long Short-Term Memory (LSTM) have also been adopted to tackle time series predicting tasks. Yao et al. [11] designed a model which combined Convolutional Neural Network (CNN) with LSTM to extract spatial and temporal features respectively. However, noting that CNN fails to extract spatial features efficiently and effectively in road traffic flow forecasting circumstances, researchers constructed other deep learning models for optimization. A novel model ConvLSTM was proposed by Gang Y et al. [12], it recognized key segments of roads that have the strongest effects on local network, which could then be helpful in the traffic flow prediction of the whole network. In Lv Z et al. [7], authors reasonably utilized Recurrent Neural Network (RNN) and CNN to accomplish speed forecasting, in which a network-embedded convolution structure was proposed to capture topology aware features, with period and background information considered, the model achieved high predicting accuracy in speed forecasting area. Yu H et al. [8] integrated the advantage of deep CNN and LSTM, and proposed Spatial-temporal Recurrent Convolutional Networks (SRCN) which captured spatial dependencies and learned temporal dynamics of traffic flow to complete the forecasting task. Furthermore, considering the intrinsic character of road traffic, graph convolutional neural networks (GCN) have also been introduced into the process of traffic flow forecasting. Yu B et al. [13] constructed a new structure, Spatial-Temporal Graph Convolutional Network (STGCN), which contained GCN and Gated Convolutional Neural Network in a spatial-temporal convolutional module, to deal with the speed forecasting problem on road. And it outperformed existing methods in training speed, flexibility and scalability. Nevertheless, the aforementioned methods either concentrate on the problem of speed forecasting [7], [10], [14], or exhibit practicability deficiency in spatial feature extraction approach, and only few of them considers the other features such as those in period and weather field, besides spatial and temporal ones.

III. Preliminary

A. Formulation of Traffic Forecasting Problem

The goal of traffic forecasting is to predict the future vehicle flow given previously observed traffic flow from N correlated roads selected from the road network. The N correlated roads and the adjacency relationship among them constitute a graph, $G(V, \varepsilon)$, where V is a set of nodes, representing the selected roads, $|V| = N$; while ε is a set of edges, representing the convergence of the two roads. A simplified example is illustrated in Figure 1.

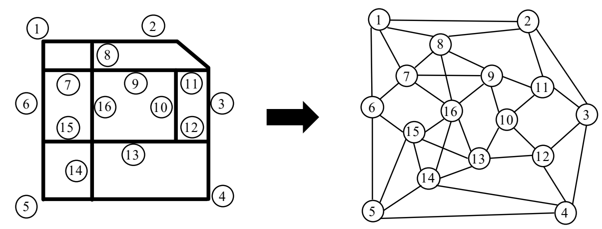


Fig. 1: Road network and its corresponding topology $G(V, \varepsilon)$

B. Deep Learning Models

Traffic flow forecasting is a classic problem in transportation and operational research, and several deep learning models are widely utilized to achieve state-of-the-art architectures.

1) Temporal Dynamics Modeling: Recurrent neural network (RNN) is a kind of model widely adopted to tackle time series problems. Basic RNNs are a network of neuron-like nodes organized into successive layers, each node in a given layer is connected with a directed connection to every other node in the next successive layer. Formula 1 gives the mathematical representation. Figure 2 gives the illustration of the structure of RNN.

$$h_t = tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{1}$$

where h_t represents the hidden state of RNN at time t, x_t denotes the input at time t, W_{xh} and W_{hh} are both weight matrix and b_h is the bias.

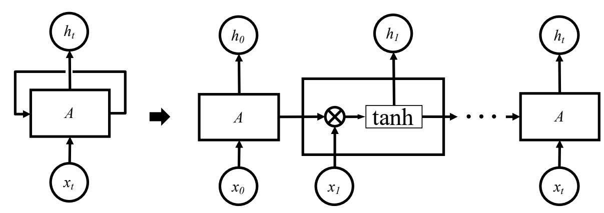


Fig. 2: An illustration of RNN structure, where A is a neuron in the network, X_i is the input to the hidden layer, and h_i is the hidden state of RNN. The right proportion of the picture can be regarded as the unfolded form of the structure on the left.

Theoretically, RNN is able to deal with time series with infinite length; however, it has the problem of gradient vanishing or exploding when faced with long series predicting tasks. As a solution to the problem, Long Short-Term Memory (LSTM) network is therefore proposed. It optimizes the internal structure of RNN neuron and achieve better performance. Formula 2 gives the mathematical representation of LSTM and Figure 3 gives the illustration of LSTM structure.

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ g_t &= tanh(W_g[h_{t-1}, x_t] + b_g) \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ h_t &= o_t \odot tanh(c_t) \end{aligned} \tag{2}$$

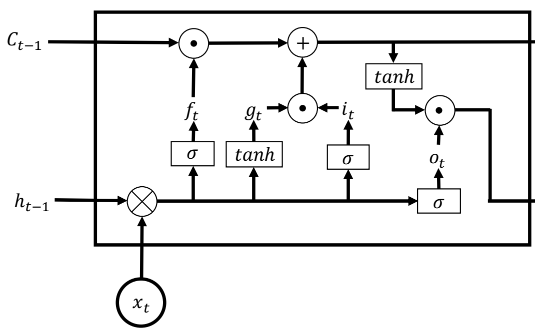


Fig. 3: The structure of memory cell in LSTM

2) Spatial Dependency Modeling: In order to make use of spatial information, Convolutional Neural Network (CNN) is frequently used in the field of computer vision, especially in image classification, image and video recognition, etc. A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. During the forward pass, each filter is convolved across the width and height of the input flow, computing the dot product between the entries of the filter and the output. Since that the road map at certain moment can also be regarded as a static image, CNN can be reasonably used to help solve traffic flow forecasting problem in extracting spatial information. Though CNN functions as a powerful tool in image information extraction, it still has some limitations, i.e., CNN is only suitable to tackle factors arrayed in Euclidean structure. Therefore, to deal with the non-Euclidean structures existing in real world, such as social network, Graph Convolutional Neural network (GCN) is proposed [15]. GCN bridges the spectral graph theory and deep neural networks. Assuming that the spectral convolution on a graph is defined as the multiplication of signal $x \in R^N$ and filter $g_\theta = diag(\theta)$, spectral graph convolution is formed as follows:

$$g_\theta \odot x = U g_\theta U^T x \tag{3}$$

where U represents the eigenvector matrix of $L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^T$, L is the Laplace operator of the graph, A is the adjacency matrix of the graph.

Due to the calculation needed in formula 8 is extremely large, Hammond [16] et al. proposed that an approximation of $g_\theta(\Lambda)$ could be reached through a K-step truncation of Chebyshev polynomial $T_k(x)$:

$$g_{\theta'}(\Lambda) \approx \sum_{k=0}^{K} \theta'_k T_k(\tilde{\Lambda}) \tag{4}$$

where $\tilde{\Lambda} = \frac{2}{\lambda_{max}} \Lambda - I_N$, and λ_{max} represents the largest eigenvalue of L, $\theta' \in R^K$ is the vector composed of Chebyshev coefficients. And Chebshev polynomial $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, given that $T_0(x) = 1, T_1(x) = x$.

Based on Formula 8 and 9, we can obtain the following results:

$$g'_\theta \odot x \approx \sum_{k=0}^{K} \theta'_k T_k(\tilde{L}) x \tag{5}$$

where $\tilde{L} = \frac{2}{\lambda_{max}} L - I_N$. Defferrard et al. [17] proposed ChebNet using the fast localized convolutions. This formula can be further simplified assuming $K = 1$ and $\lambda_{max} \approx 2$:

$$g'_\theta \odot x \approx {\theta_0}' x + \theta'_1 (L - I_N) x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \tag{6}$$

While in practice, considering the problem of overfitting, the convolution is transformed to:

$$g'_\theta \odot x \approx \theta (I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x \tag{7}$$

where $\theta = \theta'_0 = -\theta'_1$. To solve the problem of gradient exploding and gradient vanishing when training in neural networks, $I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ is changed to $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$, given that $\tilde{A} = A + I_N$, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. Consequently, the multilayer graph convolutional neural network finally adopted is:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l) \tag{8}$$

where $\tilde{A} = A + I_N$ is the adjacency matrix of undirected graph G with self-link, I_N is identity matrix, W^l is the trained weight matrix, and $H^l \in R^{N \times D}$ is the activation matrix of l-th layer, given that $H^0 = X$.

3) Dataset Description: The dataset used in our experiment is the GPS records of vehicles collected in urban area of Beijing, these records are uploaded through the control system placed in every vehicle, and every record contains multiple dimensions of information, such as record ID, time, longitude, latitude and so on, which is collected every 30 $\sim$ 60 seconds. After excluding invalid data, we finally get 2800000 pieces of record belonging to 33043 vehicles, which could be transformed to the traffic flow data after data processing.

IV. Model

A. Graph Construction

In this study we selected proportion of the urban area in Beijing for traffic flow forecasting, in which GPS records on 71 roads were contained. In order to obtain traffic flow on respective road in different time slices, data preprocessing was implemented.

Line-graph transformation. It can be anticipated that the preprocessed data of different time slices constructs a graph where roads in the network serve as the link, and the nodes in the obtained graph represent the intersections; Nevertheless traffic flow on different roads is expected to act as the information of nodes for the convenience of further analysis, this motivates us to perform a line-graph transformation for the purpose of the link-node switch of the graph.

According to graph theory, assuming that in an undirected graph G, $V(G)$ represents the set of nodes in G, and $E(G)$ represents the edges in G, then the corresponding line graph $L(G)$ of G can be:

$$\begin{aligned} &V(L(G)) = E(G) \\ &E(L(G)) = \{(e, e') | e \in G, e' \in G, e \bigcap e' = V \in G\} \end{aligned} \tag{9}$$

A brief example for Line-graph transformation is illustrated in Figure 4.

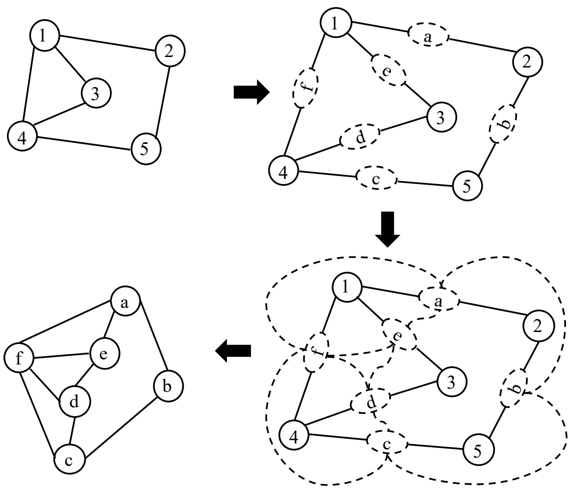


Fig. 4: Line-graph transformation

B. System Archietecture

Data processing and traffic flow forecasting constitute the main part of our traffic flow forecasting system.

For data processing module, raw GPS data is transformed to road traffic flow matrix, and road adjacency matrix is also obtained for further forecasting; and in traffic forecasting module, GCN and LSTM are utilized to extract spatial and temporal features, respectively. In addition, periodic and weather information are also included to achieve higher predicting accuracy. The architecture of the proposed system is displayed in Figure 5.

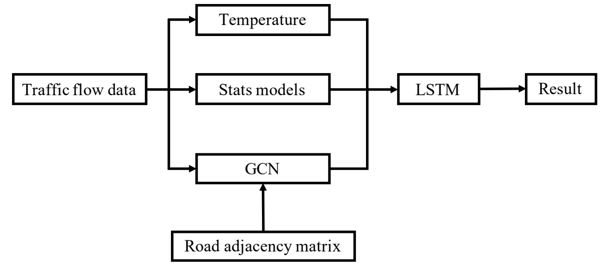


Fig. 5: system architecture

V. Experiment

A. Experiment settings

We conduct out experiment on real-world large-scale dataset, which contains the GPS records of $30,000 \sim 40,000$ vehicles in Beijing, and the size of everyday data generated is 4GB. We collect one month of data ranging from June 1st 2014 to June 30th 2014 for the experiment. We select 2/3 of the selected data for training and the remaining for testing. When training the neural networks, we chose Mean Square Error (MSE) as the loss function. In hidden layer the number of neurons is set to be 60. And the learning rate of the neural network is set to be 0.001, batch size is set to be 60, Adam optimizer is adopted as the optimizer. We compare the forecasting performance of our system with several baselines, including (1)HA: Historical Average, which models the traffic flow as a seasonal process, and uses weighted average of previous seasons as the prediction; (2) ARIMA: Auto-Regression Integrated Moving Average Model, which is frequently used in time series prediction; (3) HMM: Hidden Markov Model, a statistical Markov model with hidden states contained in Markov process, and it is usually used to deal with time series problem; (4) LSTM: Long Short-Term Memory, an optimized time series model derived from RNN; (5) CNN+LSTM: the two approaches are used for spatial and temporal dependency modeling, respectively, and it is an improvement compared with utilizing one deep learning model alone; (6) Multi-LSTM: an original approach in which traffic flow of the selected road network is obtained manually, and then the result serves as input of LSTM model. We select the traffic flow of the target roads and their adjacent roads to complete spatial dependency modeling. The utility evaluation adopted in our model is Root Mean Square Error (RMSE), its calculation is defined as below:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i}^{m}(y_i - \hat{y}_i)^2} \tag{10}$$

where m is the number of samples, y_i and $\hat{y}_i$ represent the measured data and forecasting result respectively.

B. Traffic Flow Forecasting Performance Comparison

Table I and Figure 6 show the comparison of different approaches for 10 minutes ahead forecasting on the given dataset. After analyzing the results, we observe the following phenomenon: (1) The proposed system achieves the best performance compared with all of the baselines, which suggests the effectiveness of the combination of GCN, LSTM, periodic feature extraction and weather feature extraction; (2) The integration of GCN and LSTM outperforms other spatial-temporal feature extraction model. Reasons can be explained from two perspectives: on the one hand, the utilization of GCN fits the structure of the real road network better than CNN, which accounts for the corresponding better spatial dependency modeling performance; on the other, the comprehensive consideration of spatial-temporal factors outperforms the design which only emphasizes the importance of dependency in one dimension; (3) For temporal dynamics modeling, neural networks tend to have better performance than other models, this is mainly due to that models like HA, HMM and ARIMA lack the ability to tackle complex non-stable time series data.

TABLE I: Performance comparison of different approaches for traffic flow forecasting

Model	RMSE
HA	18.4332
ARIMA	9.5155
HMM	27.0129
LSTM	6.7031
CNN+LSTM	4.3512
Multi-LSTM	2.3669
GCN+LSTM+P+W	**1.6203**

C. Effects of Spatial Dependency Modeling with GCN

To further validate the effects of GCN in spatial dependency modeling, we design a variant of the proposed approach, which only uses LSTM to complete forecasting with periodic and weather information contained. The performance comparison is displayed as below in Table II. We can observe that the forecasting accuracy of the proposed method is 75% higher than that of the variant model, which presents the necessity of spatial dependency modeling, especially with GCN.

TABLE II: Performance Analysis of Spatial Feature Extraction Module

Model	RMSE
LSTM+P+W	6.5362
GCN+LSTM+P+W	**1.6203**

D. Effects of Periodic and Weather Information

To further validate the effects of adding periodic and weather information, we design another three variants of the proposed system: (1) GCN+LSTM, this method does not include the periodic and weather information; (2) GCN+LSTM+P, this method incorporates the periodic information based on (1); (3) GCN+LSTM+W, this contains weather information based on (1). Table III exhibits the experimental results. From the results comparison we can observe that: (1) With the participation of periodic and weather information, the prediction accuracy increases about 13.7%; (2) The periodic feature of traffic flow plays a more important role in accuracy improvement, compared with weather information. This is because that the weather does not have an obvious fluctuation during the selected time in the experiment, therefore the weather information can not reveal its influence completely; In addition, periodic feature of traffic flow is observed to be a strong feature for forecasting, this explains the apparent accuracy improvement brought by the participation of periodic feature extraction.

TABLE III: Effects of periodic and weather information

Model	RMSE
GCN+LSTM	1.8774
GCN+LSTM+P	1.6376
GCN+LSTM+W	1.8594
GCN+LSTM+P+W	**1.6203**

VI. Conclusion

In this paper, we formulated the traffic flow forecasting on road network as a spatial-temporal forecasting problem, and proposed a traffic flow forecasting system based on deep learning methods. Specifically, we utilized GCN and line-graph transformation to complete spatial dependency modeling; additionally, besides the spatial-temporal feature extraction, we adopted periodic feature and weather information to complete the forecasting task, which functioned as significant factors for prediction accuracy improvement. When evaluated on large-scale real-world traffic dataset, our approach obtained significantly better prediction than baselines.

For further work, we are going to evaluate the proposed model on other real-world large-scale datasets, in order to realize traffic prediction in a wider spatial and temporal range,and conduct experiments with more state-of-the-art models involved for comparison. In additon, optimizion which investigates on the following aspects is also anticipated: (1) adding more abundant spatial and temporal features to achieve more accurate forecasting, such as the consideration of POI and weekday-weekend information; (2) applying machine learning methods in the weight distribution of different forecasting modules to achieve optimized prediction results.

Acknowledgment

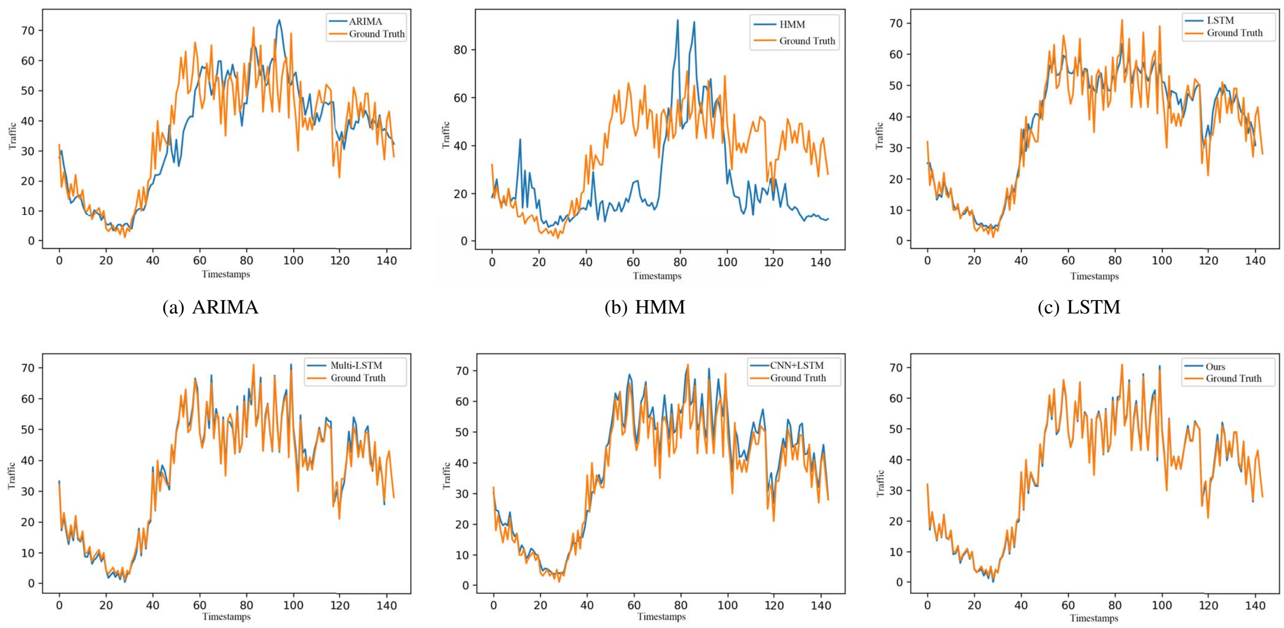
This work was supported in part by the National Science Foundation of China under Grant No. 61572071 and No.61872031 and the Fundamental Research Funds for the Central Universities under Grant No. W19JB500180.

References

[1] I. Kalamaras, A. Zamichos, A. Salamanis, A. Drosou, D. D. Kehagias, G. Margaritis, S. Papadopoulos, and D. Tzovaras, "An interactive visual analytics platform for smart intelligent transportation systems management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 487–496, 2017.

[2] H. Ge, S. Dai, L. Dong, and Y. Xue, "Stabilization effect of traffic flow in an extended car-following model based on an intelligent transportation system application," *Physical Review E*, vol. 70, no. 6, p. 066134, 2004.

[3] W.-H. Lee, S.-S. Tseng, and W.-Y. Shieh, "Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system," *Information Sciences*, vol. 180, no. 1, pp. 62–70, 2010.



(a) ARIMA (b) HMM (c) LSTM

(d) Multi-LSTM (e) CNN+LSTM (f) GCN+LSTM+P+W

Fig. 6: Forecasting results with different methods

[4] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal arima model with limited input data," *European Transport Research Review*, vol. 7, no. 3, p. 21, 2015.

[5] Y. Qi and S. Ishak, "A hidden markov model for short term prediction of traffic conditions on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95–111, 2014.

[6] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, 2018.

[7] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou, "Lc-rnn: A deep learning model for traffic speed prediction." in *IJCAI*, 2018, pp. 3470–3476.

[8] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[9] J. Moon, "On the line-graph of the complete bigraph," *The Annals of Mathematical Statistics*, pp. 664–667, 1963.

[10] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.

[11] H. Yao, X. Tang, H. Wei, G. Zheng, Y. Yu, and Z. Li, "Modeling spatial-temporal dynamics for traffic prediction," 03 2018.

[12] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, no. 7, p. 2287, 2018.

[13] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[14] Z. Cui, R. Ke, and Y. Wang, "Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction," *arXiv preprint arXiv:1801.02143*, 2018.

[15] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.

[16] 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.

[17] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in neural information processing systems*, 2016, pp. 3844–3852.