Graph Attention Spatial-Temporal Network for Deep Learning Based Mobile Traffic Prediction

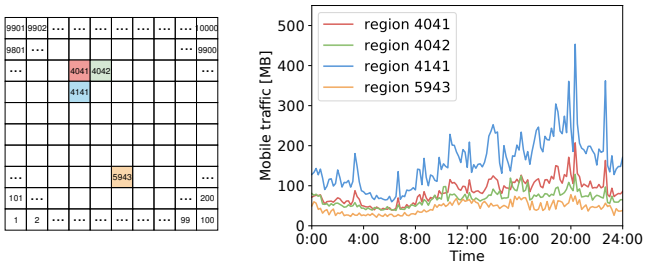
Kaiwen He, Yufen Huang, Xu Chen, Zhi Zhou, Shuai Yu
School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
{hekaiw, huangyf48}@mail2.sysu.edu.cn, {chenxu35, zhouzhi9, yushuai}@mail.sysu.edu.cn

***Abstract*—With the rapid development of mobile cellular technologies and the popularity of mobile devices, timely mobile traffic forecasting with high accuracy becomes more and more critical for proactive network service provisioning and efficient network resource allocation. Due to the complicated dynamic nature of mobile traffic demand, traditional time series methods cannot satisfy the requirements of prediction tasks well and often neglect the important spatial factors. In addition, while some recent approaches model mobile traffic prediction problem using temporal and spatial features, they only consider local geographical dependency and do not take influential distant regions into consideration. In this paper, we propose Graph Attention Spatial-Temporal Network (GASTN), a novel deep learning framework to tackle the mobile traffic forecasting problem. Specifically, GASTN considers spatial correlation through the geographical relation graph and utilizes structural recurrent neural networks to model the global near-far spatial relationships as well as capture the temporal dependencies between future demand for mobile traffic and historical traffic volume. Besides, two attention mechanisms are proposed to integrate different effects in a holistic way. Extensive experiments on a large-scale real-world mobile traffic dataset demonstrate that our model significantly outperforms the state-of-the-art methods.**

I. Introduction

In the past decade, mobile communication technology has developed unprecedentedly and the usage of mobile devices has been increasing rapidly, which result in a huge demand for mobile services. With the arrival of the fifth generation (5G) era, the demand is forecasted to bloom and increase by threefold over the next five years [1]. As a result, real-time accurate mobile traffic prediction is becoming more and more important for achieving smart network management and and providing high-quality proactive services for users. The more accurate traffic volume we predict, the better we can conduct network resource allocation in advance and finally enable intelligent mobile networking [2]–[4].

Even though network traffic shows strong dynamics over time, its periodic patterns make it possible to forecast mobile traffic using temporal features. Autoregressive integrated moving average (ARIMA) and recurrent neural network (RNN) are widely used for traffic prediction, which are time series forecasting models that focus on leveraging temporal factors [5]. Besides, there is a spatial correlation between the traffic of different regions. For example, areas with the same function have similar traffic patterns, e.g., residential areas would have low demand during the day and great demand at night while shopping districts have particularly high demand for mobile traffic on weekends compared to weekdays. In addition, suburbs generally present analogous demand patterns though they are far apart geographically. Actually, spatial dependency of mobile traffic is not only determined by real-world geographical proximity but also related to remote areas since network communication can overcome the distance factor. As an illustration, the real data shown in Fig. 1 indicates that the mobile traffic demand of region 4041 is similar to that of its adjacent region 4042 as well as the far region 5943 but it is quite different from that of its another adjacent region 4141. Therefore, it is critical to characterize spatial correlation based on such near-far effects for achieving accurate mobile traffic prediction.



(a) Geographical location (b) Traffic volume in one day

Fig. 1. The relationship between geographical location and traffic volume.

With the important insight above, in this paper, we propose a novel deep learning framework, Graph Attention Spatial-Temporal Network (GASTN) which integrates temporal and spatial features for traffic demand forecasting of the mobile network. For spatial domain, existing works tend to treat the traffic volume over a whole area as an image to capture the spatial dependency, which cannot get the best result because the grid structure fails to capture actual spatial relations about mobile traffic across all the regions and will introduce irrelevant regions as neighbors that hurts the performance. To tackle this issue, we first build a geographical relation graph by Dynamic Time Warping (DTW) algorithm to describe spatial relationship between regions instead of utilizing the given grid structure. Then, in order to integrate the graph structural information and fully utilize spatial information,

This work was supported in part by the National Science Foundation of China (No. U1711265); the Program for Guangdong Introducing Innovative and Enterpreneurial Teams (No.2017ZT07X355); and the Pearl River Talent Recruitment Program (No.2017GC010465).

Corresponding author: Xu Chen.

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

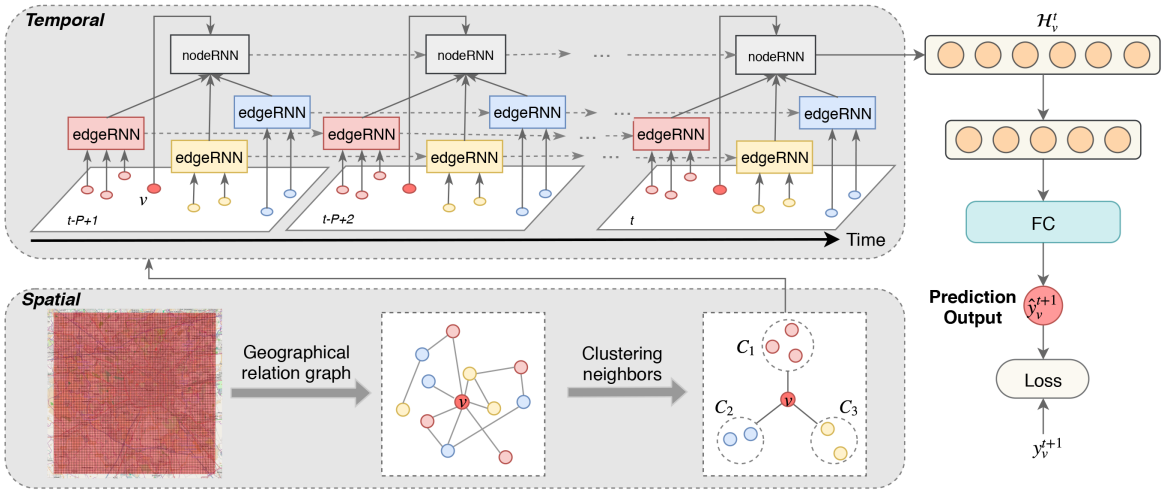


Fig. 2. The architecture of GASTN.

we construct a novel attention-based structural RNN which is able to extract spatial correlation while capturing temporal dependency.

In summary, the major contributions of this paper are as follows:

- We propose a novel deep learning model of Graph Attention Spatial-Temporal Network (GASTN) for precise mobile traffic prediction that jointly integrates temporal and spatial factors in a holistic manner.
- We build a geographical relation graph according to time series similarity of traffic demand across both near and far regions using the DTW algorithm. We then construct a novel attention-based structural RNN based on the graph to capture temporal dependency and spatial relationship simultaneously.
- Extensive experiments are conducted using a large-scale realistic dataset and the results reveal that our model significantly outperforms the state-of-the-art methods. For example, our proposed framework can achieve better performance than the most advanced STN scheme in literature and meanwhile possesses a lower running time.

The rest of this paper is organized as follows. Section II reviews the related work of mobile traffic prediction. Section III introduces the details of our proposed model. The experimental results are discussed in Section IV. The conclusion and future work are given in Section V.

II. Related Work

Numerous efforts have been made to study the dynamics of mobile networks. Mobile traffic prediction is a time series forecasting problem since there exists a strong dependency on time and lots of time series models have been applied to solve it. Autoregressive integrated moving average (ARIMA) [5], [6], and its variants [7] are widely used for network traffic forecasting which can explore the correlation between timestamps. However, ARIMA can't capture the rapid fluctuations of traffic volume since it merely depicts a linear relationship and makes predictions based on classical statistics. Mobile traffic has complicated temporal variation and nonlinear relationship so that ARIMA relatively lacks robustness.

In recent years, deep learning has made remarkable achievements in many prediction tasks [8], [9]. RNN is proposed to model sequence patterns, such as natural language processing [10] and speech recognition [11]. Long-Short Term Memory (LSTM) [12] is a variant of RNN which is able to prevent gradient vanishing and capture the long-term temporal dependency. It is utilized for mobile traffic forecasting and shows its superiority over traditional time series prediction models [13]. Nevertheless, all of the above approaches are limited to characterize spatial correlation of mobile traffic.

A line of studies applied convolutional structures to extract spatial correlation. Huang et al. [14] proposed an effective multitask learning (MTL) architecture with the combination of convolutional neural network (CNN) and RNN. Very recently, Zhang et al. [15] introduced a Spatio-Temporal neural Network (STN) architecture that simultaneously exploits spatial and temporal correlations of traffic patterns by ConvLSTM [16] and 3D-ConvNet [17]. These CNN-based methods map the spatial distribution to an image and only consider spatial dependency with adjacent areas. Thus, the impact of distant regions which have similar traffic demand patterns to the target area will be ignored. In fact, mobile traffic is not only related to local spatial correlation but also affected by global spatial dependency, so it's necessary to take remote regions into account.

Our proposed method considers both temporal dependency and spatial correlation of mobile traffic. Specifically, to capture near-far spatial correlation, we construct a geographical relation graph according to the similarity of network traffic patterns in different areas, regardless of the actual distance.

III. GASTN Framework for Mobile Traffic Prediction

In this section, we first formulate the problem of mobile traffic prediction and then introduce our proposed Graph

Attention Spatial-Temporal Network (GASTN) in detail. The whole architecture of GASTN is shown in Fig. 2.

A. Problem Formulation

In this study, The geographical area can be divided into a $n \times n$ grid map with totally M grids based on the longitude and latitude where a grid maps to a region. For each region v, it has a whole time series of mobile traffic $\{y_v^1, y_v^2, ..., y_v^T\}$ during a time period T and y_v^t represents the traffic volume in region v during the t^{th} time slot. Our objective is to forecast the demand for mobile traffic in region v at time interval $t+1$ given the past P observed values until time interval t. Since mobile traffic is affected by temporal and spatial factors collectively, the problem can be formulated as

$$\hat{y}_v^{t+1} = \mathcal{F}(\mathcal{Y}_v^{t-P+1:t}, \mathcal{X}_v^{t-P+1:t}), \tag{1}$$

for $v \in M$, where $\mathcal{Y}_v^{t-P+1:t} = \{y_v^{t-P+1}, ..., y_v^{t-1}, y_v^t\}$ is a set of the observations of mobile traffic demand in region v at the previous P time slots, and $\mathcal{X}_v^{t-P+1:t} = \{\mathcal{Y}_u^{t-P+1:t} | u \in NB(v)\}$ are the data of v's neighbors $NB(v)$ in the same time period.

B. Construction of Geographical Relation Graph

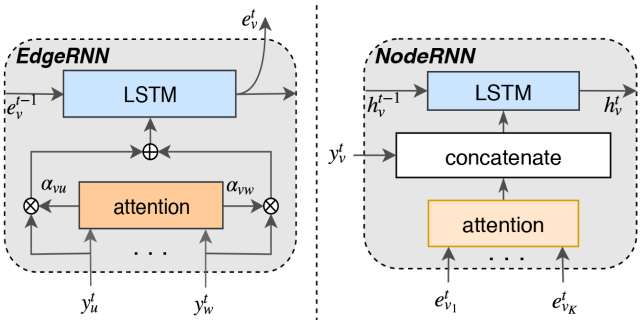
The geographical area is partitioned into a list of regular grids and each grid is regarded as a region. However, the grid structure is not suitable to reflect the actual relationship between regions about mobile traffic volume since it only considers the nearby regions as neighbors but ignores the distant ones.

To depict the neighborhood relation between regions regardless of the real geographical location, we put forward to construct a weighted graph $G = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, which is able to represent the degree of relevance between every two regions. In general, regions are closely related in terms of mobile traffic when they have similar traffic patterns. Therefore, we adopt Dynamic Time Warping (DTW) algorithm to measure the traffic sequences similarity between two regions [18]. Firstly, we treat each grid as a node and the weight of an edge is calculated based on DTW. The weight w_{uv} of the edge connecting node u and node v is estimated as

$$w_{uv} = \exp(-\text{DTW}(ts_u, ts_v)), \tag{2}$$

where ts_u and ts_v represent the time series of traffic in training data for region u and region v respectively, and $\text{DTW}(\cdot)$ is the normalized dynamic time warping distance between the two sequences. Then we get a dense weight graph where every two nodes are connected by an edge.

Since a target region has spatial dependency mainly on a few regions, it's essential to find out the closely related neighbors for each region. In order to focus on the strong correlation as well as reduce computation, we propose to build a sparse graph based on the dense one. We select top-N nearest neighboring nodes for each node according to the weights of its edges and we consider the relationship between nodes is symmetric that node u would also be chosen as a neighbor of node v if v is u's neighbor even though u is not in the top-N list of v. Then we eliminate the edges between each node and its unselected neighbors and get a sparse weighted graph. This geographical relation graph can reflect the spatial relationship among regions without regard for the distance factor.



(a) edgeRNN (b) nodeRNN

Fig. 3. The architecture of edgeRNN and nodeRNN. (a) The inputs $\{y_u^t, ..., y_w^t\}$ of edgeRNN are the traffic volumes of neighbors of target node v at time interval t. (b) The output of edgeRNN i for target node v, denoted as $y_{v_i}^{'t}$, acts as an input of nodeRNN.

C. Attention-based Structural RNN for Traffic Forecasting

In order to model our problem on the geographical relation graph defined above, we propose a graph-based method, attention-based structural RNN (S-RNN) which is inspired by [19], and the major difference is that two attention mechanisms are applied to S-RNN for capturing the diverse influence of neighbors and distinct importance of different kinds of edges. Traditionally, S-RNN is proposed to cast a spatio-temporal graph (st-graph) as a rich RNN mixture and is exploited mainly for computer vision, such as modeling human motion and detecting object interactions [19], and to the best of our knowledge, we are the first to leverage S-RNN for mobile traffic prediction. For the S-RNN part that we develop for mobile traffic prediction, nodes in the graph are first classified into K classes $\{C_1, ..., C_K\}$ according to the overall mobile traffic demand and thus there are $\frac{K(K+1)}{2}$ kinds of edges $\{E_{ij} | i, j = 1, ..., K\}$, where E_{ij} is a set of the edges connecting a node $u \in C_i$ and a node $v \in C_j$. For each class C_i of nodes, there is a corresponding nodeRNN R_{C_i} to handle node information. Likewise, for each kind E_{ij} of edges, an edgeRNN $R_{E_{ij}}$ is used for extracting edge information so as to capture spatial dependency. The architecture of edgeRNN and nodeRNN is shown in Fig. 3, which will be described in detail next.

For the approach we proposed, in the forward pass for a target node $v \in C_i$, the temporal sequences of its neighbors of class C_j, denoted as $\{\mathcal{Y}_u^{t-P+1:t} | u \in (NB(v) \cap C_j)\}$, are input to edgeRNN $R_{E_{ij}}$ and integrated via the following attention mechanism. Since distinct neighbors have different influence on the target node, we first propose a soft attention module to explore different degrees of importance of neighbors. Specifically, for node v in the geographical relation graph, the attention coefficient α_{vu}^t of its neighbor u is calculated as

follows:

$$e_{vu}^{t} = \sigma(\boldsymbol{W_a}[\mathcal{Y}_v^{t-P+1:t}, \mathcal{Y}_u^{t-P+1:t}] + \boldsymbol{b_a}), \tag{3}$$

$$\alpha_{vu}^{t} = \frac{\exp(e_{vu}^{t})}{\sum_{k\in(NB(v)\cap C_j)} \exp(e_{vk}^{t})}, \tag{4}$$

where $\boldsymbol{W_a}$ and $\boldsymbol{b_a}$ are trainable parameters and $\sigma(\cdot)$ is an activation function. It represents the importance of neighboring node u in class C_j for target node v. Therefore, each neighboring node u is assigned an attention weight α_{vu}^{t} and then the spatial dependency of v on neighboring nodes of class C_j is represented as

$$Y_{v_j}^{t-P+1:t} = \sum_{u\in(NB(v)\cap C_j)} \alpha_{vu}^{t} \cdot \mathcal{Y}_u^{t-P+1:t}, \tag{5}$$

which is fed to the time series module in $R_{E_{ij}}$. We utilize LSTM as the time series module to capture the useful information of neighbors and get the hidden representation as $Y_{v_j}^{'t-P+1:t} \in \mathbb{R}^{P\times H}$, where H is the dimension of hidden representation of LSTM. Then we get outputs from edgeRNNs $\{R_{E_{ik}}|k = 1, ..., K\}$ respectively, which are associated with the target node $v \in C_i$.

Besides, the second attention module, which is a component of nodeRNN, is suggested to solve the problem that different kinds of edges exert different effects for the target node. We derive the attention weight $\beta_{v_j}^{t}$ of edgeRNN $R_{E_{ij}}$ for target node v using following equations:

$$s_{v_j}^{t} = \sigma(\boldsymbol{W_b}[\mathcal{Y}_v^{t-P+1:t}, Y_{v_j}^{'t-P+1:t}] + \boldsymbol{b_b}), \tag{6}$$

$$\beta_{v_j}^{t} = \frac{\exp(s_{v_j}^{t})}{\sum_{k=1}^{K} \exp(s_{v_k}^{t})}, \tag{7}$$

where $\boldsymbol{W_b}$ and $\boldsymbol{b_b}$ are parameters to be learned. The coefficient $\beta_{v_j}^{t}$ indicates the significance of edgeRNN E_{ij} for v, and the output of attention is computed by weighted summation

$$X_v^{t-P+1:t} = \sum_{j=1}^{K} \beta_{v_j}^{t} \cdot Y_{v_j}^{'t-P+1:t}. \tag{8}$$

Finally, the output $X_v^{t-P+1:t} \in \mathbb{R}^{P\times H}$ and the target node v's historical time series $\mathcal{Y}_v^{t-P+1:t} \in \mathbb{R}^{P\times 1}$ are concatenated and fed to the time series module in R_{C_i}, and then we get an output of vector representation $\mathcal{H}_v^{t} \in \mathbb{R}^{H}$ for node v at time interval t. It is formulated as:

$$\mathcal{H}_v^{t} = \text{LSTM}([\mathcal{Y}_v^{t-P+1:t}, X_v^{t-P+1:t}]). \tag{9}$$

The attention mechanisms determine the importance of different areas and the combination of nodeRNN and edgeRNN considers both self temporal dependency and complicated spatial impact simultaneously which enables a precise mobile traffic prediction. In the end, $\mathcal{H}_v^{t}$ is fed to the fully connected layer to get the final prediction value $\hat{y}_v^{t+1}$.

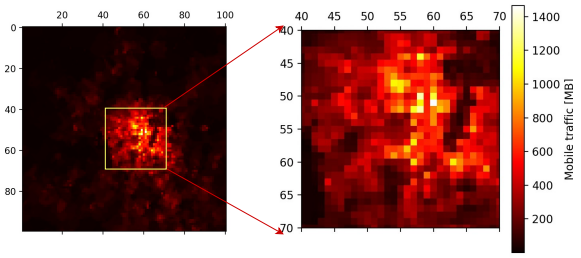


Fig. 4. Milan mobile traffic heat map and the coverage area with 30×30 grids is selected to be studied.

IV. Experiments

A. Data Description

In this paper, we evaluate our proposed method on a large-scale real-world telecommunications dataset, which is provided by Telecom Italia and publicly available [20]. The dataset contains records of mobile traffic volume observed over 10-minute intervals for the city of Milan where is partitioned into 100×100 regions and the size of each region is $235m \times 235m$. The data we use is from 1st Nov. 2013 to 30th Nov. 2013 and each record in the dataset includes region ID, the beginning of the time interval and the mobile traffic volume during the time interval. In the experiment, we choose data from 01/11/2013 to 20/11/2013 (20 days) as training data and the rest 10 days as testing data.

B. Experimental Settings

Preprocessing. Fig. 4 shows a heat map of Milan's total cellular traffic volume during November 2013. It indicates that the central part of Milan generally has a great demand for mobile traffic while little traffic is required in other regions. Since people don't care about the low-demand situation in practical applications, we focus on the regions (30×30 grids) with great demand and evaluate our proposed method on the corresponding data. For the traffic data, we scale them to [0,1] by Min-Max Normalization before prediction and anti-normalize the predicted values for evaluation after prediction. Samples for training and testing are generated by a sliding window on the data.

Hyperparameter Settings. We set the length of input sequence $P = 6$ (i.e., previous 1 hour) and apply our model to predict the traffic volume in the next time step. The threshold of the size of each neighborhood is set as $N = 15$ and the number of classes for nodes is set to $K = 3$. In our model, we utilize LSTMs as the recurrent neural networks in edgeRNN and nodeRNN, and the number of hidden units of each LSTM is 64.

Baselines. We compare our GASTN with the following widely used methods:

- **HA**: Historical average (HA) predicts the value using the average of previous mobile traffic demands of the given region in the same relative time interval (i.e., the same time of the day) [21].
- **ARIMA**: ARIMA is a well-known time series model and widely used for mobile traffic prediction [7].

TABLE I
COMPARISON WITH DIFFERENT BASELINES

Method	NRMSE	MAE
HA	0.4338	69.9062
ARIMA	0.1817	32.6042
MLP	0.2089	37.9893
LSTM	0.1771	32.6483
CNN-LSTM	0.2089	37.7972
STN	0.1750	33.0419
GASTN	**0.1705**	**30.9307**

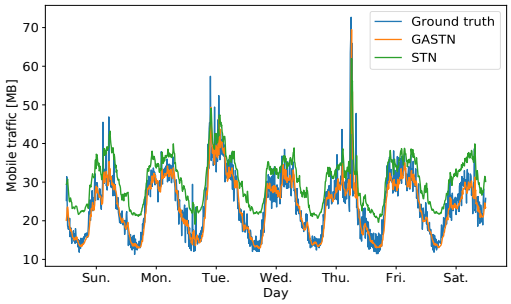


Fig. 5. The ground truth and predicted values of mobile traffic volume in the last week of November for region 6643.

- **MLP**: Multiple layer perceptron (MLP) is a neural network with three fully connected layers that the number of hidden units are 64, 128, 64 respectively.
- **LSTM**: LSTM is able to capture short-term and long-term temporal dependency and it has been widely used in time series forecasting problems including mobile traffic prediction [13], [22].
- **CNN-LSTM**: CNN and LSTM are integrated into a model to capture spatial dependency and temporal relationship respectively [14].
- **STN**: STN is a deep spatio-temporal network that combines ConvLSTM and 3D-ConvNet to capture temporal and spatial relationships simultaneously. It obtains the state-of-the-art result in mobile traffic forecasting [15].

Evaluation Metric. We evaluate our model by two commonly used metrics: Normalized Root Mean Square Error (NRMSE) [15] and Mean Absolute Error(MAE) [14], which are defined as follows:

$$NRMSE = \frac{1}{\bar{y}}\sqrt{\frac{1}{z}\sum_{i=1}^{z}(\hat{y}_i - y_i)^2}, \tag{10}$$

$$MAE = \frac{1}{z}\sum_{i=1}^{z}|\hat{y}_i - y_i|, \tag{11}$$

where $\hat{y}_i$ and y_i are the predicted value and ground truth, z is the total number of predicted values, and $\bar{y}$ is the mean of all ground truth values.

C. Experimental Results

Performance Comparison. Table I shows the experimental results of our proposed method and other baselines on Milan dataset. It can be seen that GASTN generally outperforms all baselines, which achieves the lowest NRMSE (0.1705) and the lowest MAE (30.9307) in the experiment. Specifically, the traditional time series models (HA and ARIMA) don't perform well. For deep learning methods, because of the ability of LSTM to capture temporal dependency, LSTM performs better than MLP. CNN-LSTM and STN consider the relations of both space and time but their performance is worse than GASTN since these methods capture geographical relation based on the grid structure. They regard all nearby areas of a region as its neighbors and predict the traffic volume of the target region according to its historical information as well as the neighbors' that will introduce noise provided by some actually irrelevant neighbors. Therefore, it further illustrates the essentiality of selecting neighbors by constructing the geographical relation graph. GASTN utilizes temporal dependency and spatial correlation for modeling in a holistic manner that achieves the best performance among all state-of-the-art baselines. The experimental results demonstrates the effectiveness of our model.

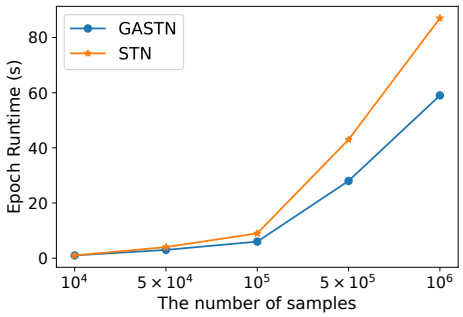
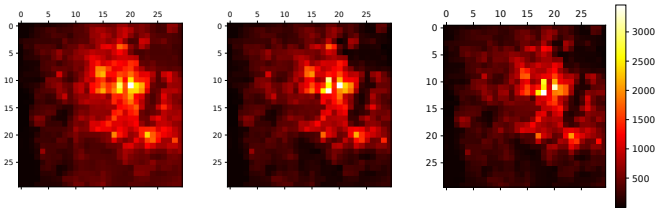


Fig. 6. The training time of GASTN and STN for each epoch under different sample sizes.

In Fig. 5, we plot the ground truth and predicted values for mobile traffic in the last week of November for a given region and the predicted values are generated by our model GASTN and the up-to-date method STN respectively. It illustrates that the curve generated by our model fits the real traffic curve better while STN is unable to grasp the dynamics well especially when traffic demand declines. Compared to the training time of GASTN and STN for each epoch under the same experimental equipment in Fig. 6, it shows that the training time of GASTN is always less than STN's and as the number of samples increases, the gap between them becomes more and more obvious. It implies that the complexity of GASTN is lower and further indicates that the effectiveness of GASTN depends on its reasonable design rather than model complexity. In addition, we give an example of the comparison of the predicted traffic distribution and the real distribution in Fig. 7, which indicates that GASTN can predict mobile traffic distribution in a whole city more precisely compared with STN.

Effect of attention mechanisms. In this section, we analyze the effectiveness of the attention mechanisms of GASTN and several variants are listed as follows:



(a) STN (b) GASTN (c) Ground truth

Fig. 7. An example of the predicted mobile traffic distributions using STN and GASTN respectively and the corresponding ground truth.

TABLE II
COMPARISON WITH VARIANTS OF GASTN

Method	NRMSE	MAE
GASTN-AT1	0.1729	32.2622
GASTN-AT2	0.1730	32.5218
GASTN-ATs	0.1734	32.5469
GASTN	**0.1705**	**30.9307**

- **GASTN-AT1**: GASTN-AT1 removes the first attention mechanism in GASTN that just averages the temporal sequences of neighbors in edgeRNN.
- **GASTN-AT2**: we simply concatenate the outputs of edgeRNNs instead of using the second attention mechanism and then feed it into nodeRNN.
- **GASTN-ATs**: GASTN-ATs is applied without the two attention mechanisms mentioned above.

The results of these variants are shown in Table II. It reveals that GASTN outperforms its variants because GASTN-AT1 overlooks the different effects caused by different neighbors of a target node and GASTN-AT2 ignores the discrepancy of importance of distinct edgeRNNs. GASTN-ATs performs worst among these variants since it pays no attention to the differences in neighbors or edge types. The results manifest the effectiveness of the attention mechanisms in GASTN to collectively capture the impact of spatial dependency on mobile traffic prediction.

V. CONCLUSION

In this paper, we propose a novel Graph Attention Spatial-Temporal Network (GASTN) for mobile traffic forecasting. Our approach integrates spatial and temporal views together to extract important information for prediction. The near-far spatial correlation is captured by the geographical relation graph and the spatial-temporal factors are modeled by attention-based structural RNN. We evaluate our model on a large-scale mobile traffic dataset and the experimental results demonstrate our proposed method can outperform the state-of-the-art method with faster running time.

In the future, we plan to consider other types of data (e.g., social data, SMS and call activity data) and find out the relationship between mobile traffic and these volumes, then use the multi-source dataset for a more accurate prediction.

REFERENCES

[1] Cisco, *Cisco visual networking index: forecast and trends, 2017-2022*, 2018.

[2] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE transactions on services computing*, vol. 9, no. 5, pp. 796–805, 2016.

[3] A. Furno, M. Fiore, and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[4] R. Li, Z. Zhao, Z. Xuan, G. Ding, C. Yan, Z. Wang, and H. Zhang, "Intelligent 5g: When cellular networks meet artificial intelligence," *IEEE Wireless Communications*, vol. PP, no. 99, pp. 2–10, 2017.

[5] H. W. Kim, J. H. Lee, Y. H. Choi, Y. U. Chung, and H. Lee, "Dynamic bandwidth provisioning using arima-based traffic forecasting for mobile wimax," *Computer Communications*, vol. 34, no. 1, pp. 99–106, 2011.

[6] A. Adas, "Traffic models in broadband networks," *Comm. Mag.*, vol. 35, no. 7, pp. 82–89, Jul. 1997.

[7] Y. Shu, M. Yu, J. Liu, and O. W. W. Yang, "Wireless traffic modeling and prediction using seasonal arima models," *Ieice Transactions on Communications*, vol. E88B, no. 10, pp. 1675 – 1679, 2003.

[8] J. Liu and Y. L. Huang, "Nonlinear network traffic prediction based on bp neural network," *Journal of Computer Applications*, vol. 27, no. 7, pp. 1770–1772, 2007.

[9] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *AAAI Conference on Artificial Intelligence*, 2017.

[10] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[11] S. Kombrink, T. Mikolov, M. Karafiát, and L. Burget, "Recurrent neural network based language modeling in meeting recognition," in *Twelfth annual conference of the international speech communication association*, 2011.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile traffic prediction from raw data using LSTM networks," in *29th IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2018, pp. 1827–1832.

[14] C. W. Huang, C. T. Chiang, and Q. Li, "A study of deep learning networks on mobile traffic forecasting," in *IEEE International Symposium on Personal*, 2018.

[15] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proceedings of the Nineteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc*, 2018, pp. 231–240.

[16] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, 2015, pp. 802–810.

[17] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.

[18] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proceedings of the First SIAM International Conference on Data Mining, SDM 2001, Chicago, IL, USA, April 5-7*, 2001, pp. 1–11.

[19] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, June 2016, pp. 5308–5317.

[20] G. Barlacchi, M. D. Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, "A multi-source dataset of urban life in the city of milan and the province of trentino," *Scientific Data*, vol. 2, p. 150055, 2015.

[21] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *AAAI Conference on Artificial Intelligence*, 2018.

[22] R. Vinayakumar, K. Soman, and P. Poornachandran, "Applying deep learning approaches for network traffic prediction," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 2353–2358.