DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction

Jie Feng, Xinlei Chen, Rundong Gao, Ming Zeng, and Yong Li

Abstract

The past 10 years have witnessed the rapid growth of global mobile cellular traffic demands due to the popularity of mobile devices. While accurate traffic prediction becomes extremely important for stable and high-quality Internet service, the performance of existing methods is still poor due to three challenges: complicated temporal variations including burstiness and long periods, multi-variant impact factors such as the point of interest and day of the week, and potential spatial dependencies introduced by the movement of population. While existing traditional methods fail in characterizing these features, especially the latter two, deep learning models with powerful representation ability give us a chance to consider these from a new perspective. In this article, we propose Deep Traffic Predictor (*DeepTP*), a deep-learning-based end-to-end model, which forecasts traffic demands from spatial-dependent and long-period cellular traffic. DeepTP consists of two components: a general feature extractor for modeling spatial dependencies and encoding the external information, and a sequential module for modeling complicated temporal variations. In the general feature extractor, we introduce a correlation selection mechanism for a spatial modeling and embedding mechanism to encode external information. Moreover, we apply a seq2seq model with attention mechanism to build the sequential model. Extensive experiments based on large-scale mobile cellular traffic data demonstrate that our model outperforms the state-of-the-art traffic prediction models by more than 12.31 percent.

Introduction

Over the last 10 years, with the popularity of mobile devices and mobile applications, the demand of mobile traffic has grown rapidly around the world. According to the technical report from Cisco [1], the global cellular network traffic from mobile devices is expected to surpass 48.3 EB (1018 B) per month in 2021. To provide stable Internet service with guaranteed quality of service (QoS) [2], predicting mobile traffic demands accurately becomes extremely important for both service and infrastructure providers. For example, with the help of accurate prediction of traffic demands, they can realize a timely traffic schedule to avoid traffic jams by pulling partial traffic demands away from busy cell towers to free ones.

By modeling it as a general time series forecasting problem, some research efforts [3–7] have been made to improve the performance of traffic demand forecasting. Seasonal AutoRegression Integrated Moving Average (SARIMA) and support vector regression (SVR) are the most widely used methods in these works. However, SARIMA fails to capture the rapid traffic variation since it relies on the mean value of historical data. In addition, SARIMA cannot model the nonlinear relationship in a real system. Although SVR can model the nonlinear relationship, it requires well-tuned key parameters to achieve accurate prediction results. Nevertheless, the major drawback of these methods is that they totally ignore potential correlations between traffic series, like spatial dependencies, which are very common and important in a mobile network. Due to the popularity of mobile networks, more and more people use their mobile devices to access cellular networks while moving. Thus, the traffic demand movement caused by human movement leads to remarkable spatial dependencies between the traffic of different base stations. Besides, the basic traffic demands in a specific area can be affected by the environment around it, like the point of interest (PoI) distribution and weekly effect, which are also beyond the ability of existing traditional methods. Even in capturing temporal characteristics, existing traditional methods are limited to only keeping short memory because of the limited parameters and computing efficiency. "Infinite" memory is impossible for them to model, while such long-term memory can be useful in practice. These observations encourage us to explore more powerful modeling tools to capture and characterize them.

Recently, as a kind of specially designed neural network, the recurrent neural network (RNN) has been widely used to model complicated nonlinear sequence patterns, which achieves promising results in many fields, such as natural language processing, speech recognition, and video processing. Because of the gradient vanishing and the exploding risk of the general RNN, long short-term memory (LSTM) [8] is used in these tasks. Theoretically, the RNN is designed to be capable of capturing infinite temporal relations. Meanwhile, a neural network model is good at importing discrete factors into the model by directly encoding them as vectors. Hence, the RNN is regarded as a promising tool to model complicated traffic time series. Recent works [3, 5] utilize LSTM to model the traffic

Digital Object Identifier:
10.1109/MNET.2018.1800127

Jie Feng, Rundong Gao, Ming Zeng, and Yong Li are with Beijing National Research Center for Information Science and Technology, Dept. of Electronic Engineering, Tsinghua University.; Xinlei Chen is with Carnegie Mellon University.

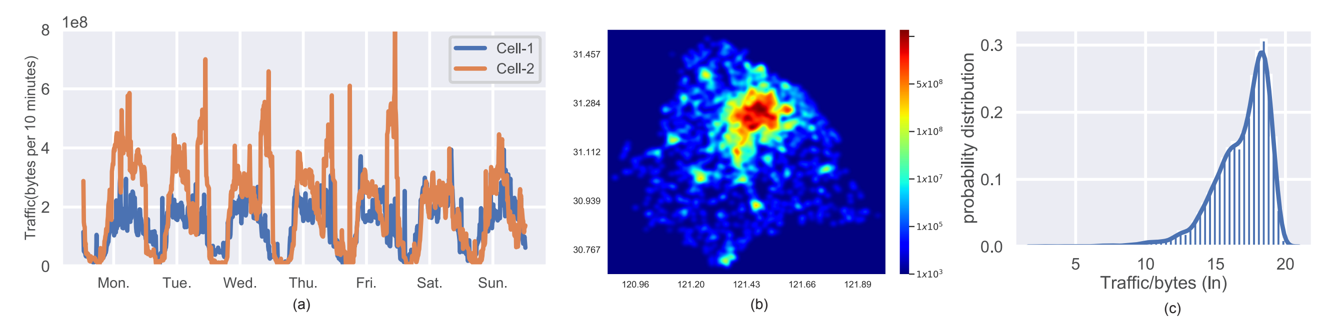

FIGURE 1. Multi-view distribution of mobile traffic in Shanghai: a) temporal distribution of the mobile traffic of two typical cells; b) spatial distribution of the mobile traffic at 17:00; c) probability distribution of the mobile traffic.

flow for wireless networks. Furthermore, Wang *et al*. [5] applied an auto-encoder [8] to capture both local and global spatial dependencies of traffic between adjacent cell towers. However, their approach needs to preprocess the data for the neural network by projecting the traffic of each cell tower into a square, which limits its application scenarios and causes unnecessary errors. Besides, their auto-encoder can only model the adjacent spatial dependencies by predesigned adjacent areas, which are strict in a real system. Beyond the spatial-temporal relationships, many other discrete factors also influence the trend and volume of mobile traffic, like the location of the cell tower and the PoI distribution of adjacent areas, which suggests the functions. Because of the heterogeneity and sparsity, no existing methods are designed for directly processing the external information. Therefore, it is challenging to take these discrete factors into account in the forecasting model.

In this article, we propose *DeepTP*, a deep-learning-based end-to-end framework for traffic demands forecasting from heterogeneous and periodic traffic data of mobile networks. In DeepTP, we first design a general feature extractor to model the spatial relationship and external information from a new perspective. In this feature extractor, a discrete embedding module is designed to encode external information, such as the PoI category and the day of the week, into a unified dense vector. Meanwhile, we build an attention-based module, which selects traffic features from other cell towers by the traffic of a certain cell to represent their influence. The auto-selection is based on the "correlation" of these traffic histories, which is not only suitable for modeling the influence from adjacent cell towers but also capable of capturing distant spatial relations. Particularly, our module mines two kinds of "correlations" from the traffic features: the "positive correlation," which represents the influence of similar-traffic-pattern cell towers, and the "negative correlation," which represents the influence of opposite-traffic-pattern cell towers. Finally, with the features from the general feature extractor, we utilize an LSTM-based RNN as the basic sequential model to capture the temporal information. Specifically, this sequential module is implemented by a seq2seq model with the attention mechanism. While the seq2seq model enables the multi-step prediction, the attention mechanism makes the sequential module observe more deeply and work robustly.

The contributions of this article can be summarized as:

- We propose DeepTP, a deep-learning-based end-to-end framework to predict the traffic demands of data from Shanghai, which includes about 10,000 base stations. The prediction results demonstrate that DeepTP outperforms state-of-the-art traffic prediction models by more than 12.31 percent. DeepTP captures the spatial-temporal features of the mobile cellular traffic and models the influence of external information for traffic prediction.

We structure the article as follows. We first introduce the challenges and overview of our investigated problem, followed by an overall visualization of the temporal features of mobile big data. Then we introduce our end-to-end neural network framework for mobile cellular traffic prediction. We then introduce the evaluation environment with the dataset and also compare the performance of our system with the state-of-the-art solutions. Finally, we summarize our study and discuss future work.

Challenges and Overview

For mobile traffic, as a kind of spatial-temporal series, its forecasting has to face similar key challenges of modeling sequential information as the traditional time series forecasting problem. Furthermore, accurate forecasting requires addressing the challenges of modeling the complex and underlying spatial correlation and influence from external information.

Challenge 1: Complicated Temporal Variation: As the typical mobile traffic series shown in Fig. 1a, it is not only multi-level periodical but also highly bursty. Mobile traffic demands show a similar temporal trend as that of human daily life. Meanwhile, influenced by the natural burstiness of the Internet, mobile traffic demands also vary rapidly. Both of them make the mobile traffic series difficult to forecast. Existing methods like ARIMA and SVR fail to handle the burstiness. Recently, as a powerful sequential modelling tool, the RNN has applied for prediction in many sequence modeling tasks, including traffic forecasting, and has achieved promising results. Similarly, we choose the RNN as the basic component of our model to handle the complicated sequential information. Furthermore, we apply two advanced structures, the seq2seq module and attention mechanism, to strengthen the model.

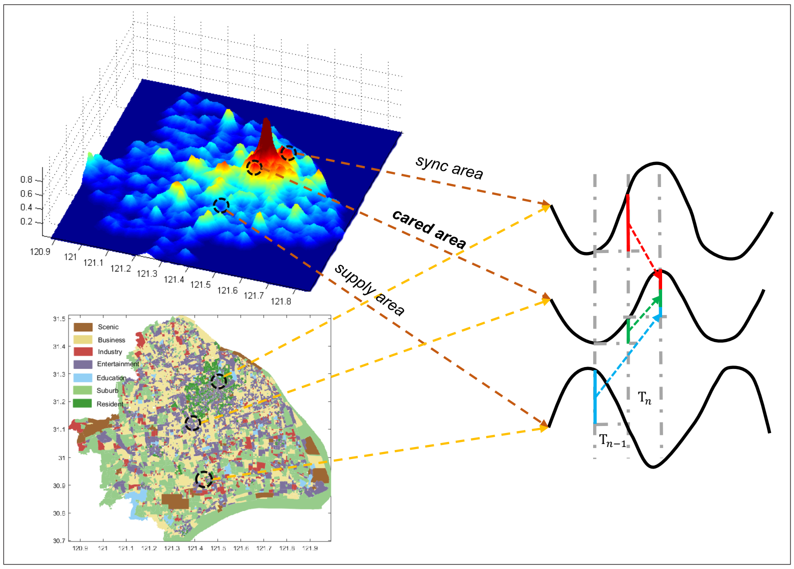

FIGURE 2. The intuition of our solution: considering the spatial dependencies and external information into modeling and forecasting.

Challenge 2: Multi-Variant Influence Factors: As Fig. 1c shows, the characteristics of traffic like volume among different base stations are totally different. According to Wang *et al.* [10, 11], a lot of factors influence the traffic pattern of large-scale cell towers around a city. For example, the traffic of a cell tower near transportation facilities can be highly bursty during rush hour, while those located in entertainment areas may grow rapidly in the evening. Different locations mean different service groups and city functions, which lead to totally different traffic demand patterns. To solve this problem, traditional methods usually first cluster the traffic and train different independent predictors for different traffic patterns. While working to some extent, the division solution causes the error spreading among two steps, and the clustering results in high interference with the prediction accuracy. Therefore, we propose to directly encode these influence factors into the model by embedding. In this way, we can combine these two steps and optimize them together. Furthermore, with the help of embedding, we are able to take various discrete factors into consideration such as the day of the week and road density.

Challenge 3: Spatial Correlations among Base Stations: Different from the general time series forecasting problem, where different time series vary independently, the traffic variations of different base stations in a city are correlated (i.e., spatial correlations). As Fig. 1b shows, adjacent base stations share similar traffic patterns, while some faraway base stations share reversed traffic patterns. According to Xu *et al*. [12], the movement of users who require and consume mobile traffic contributes to spatial correlations. In addition, cells located adjacently share similar infrastructure and serve similar user groups with the same habits. Beyond the simple close spatial locations, cells that are distant can also be contacted because of their similar city function or the connection of advanced transportation systems like subways and buses. On one hand, these various spatial correlations make it difficult to predict traffic accurately. On the other hand, they also give us great chances to improve the performance of prediction by forecasting them together. However, limited by the capability of modeling tools, previous methods like ARIMA and SVR fail to handle the spatial correlations. Some researchers applied denoising auto-encoders [9] to model the spatial correlation and achieved interesting results. However, they can only capture the adjacent spatial correlation, and both of them are difficult to interpret for the forecasting results, which limits their application scenarios. Furthermore, their solution requires pre-defining the related cell towers by hand, which may introduce errors and cause prediction performance deterioration. In this article, we design a spatial correlation extractor by applying a "temporal attention" mechanism. With this auto-extractor, the spatial correlation beyond the constraint of adjacent locations is modeled, which makes the prediction results easy to interpret.

Figure 2 illustrates the basic intuition of our work. The heatmap in the upper left shows the spatial distribution of traffic at 5 p.m. in Shanghai, while the image in the lower left presents the function distribution of Shanghai. The right part shows three traffic series curves of three selected base stations from 10,000 candidates. As the middle curve is the traffic we currently care about, traditional methods predict its future value only based on its own history without any spatial dependencies and external information. Besides, existing methods [6] take the spatial dependencies into consideration to predict the future value by directly encoding the traffic of a fixed number adjacent cell towers. However, not all the adjacent cell towers influence the traffic trend of the current cell tower. Meanwhile, some distant base stations may also influence its traffic flow with the help of an advanced transportation system. To overcome the drawbacks of existing methods and model the mentioned potential spatial influence, we propose to design a specific module to not only model the adjacent spatial influence, defined as sync influence, but also model the distant spatial influence, defined as supply influence. The details of the model are discussed later.

Model and Solution

Figure 3 shows the framework of our DeepTP, which is a deep-learning-based end-to-end framework for mobile cellular traffic prediction. It consists of two components: the *features extractor*, which generates temporal features and spatial features of the data traffic series, and the *sequential module*, which models the sequential relationship among the traffic features to derive the prediction result.

The input of our model consists of three parts: one primary traffic series, some auxiliary traffic series (selected from all the candidate traffic series based on distance), and discrete factors of primary traffic series like PoI distribution. At first, the traffic series are fed into a spatial correlation selector to generate one primary traffic feature series. Meanwhile, the discrete factors are fed into the discrete embedding module to generate external vectors. Further, external vectors are copied and concatenated with the primary traffic feature series. Then the sequential model takes this

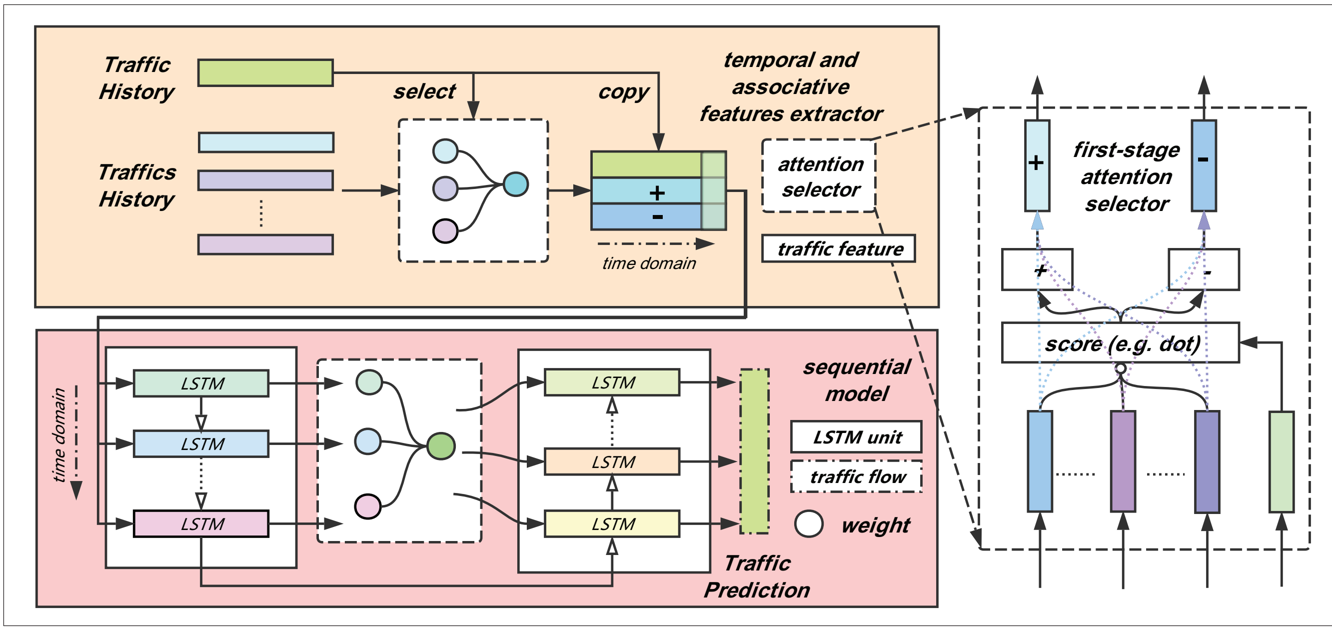

FIGURE 3. Main architecture of DeepTP including two components: features extractor and sequential model.

new future series as input to generate the future value step by step. The details of our models are as follows.

Features Extractor

As discussed above, the mobile cellular traffic is difficult to predict because of the complicated influential factors and comprehensive spatial-temporal correlations. To handle these various features, we design a features extractor module that employs the embedding and attention mechanism. As Fig. 3a shows, the features extractor consists of two modules:

- *Spatial correlation selector*, which models the influence from the adjacent and related areas of the cared area
- *Discrete embedding module*, which models the individual characters of the considered area itself

Spatial Correlation Selector: As presented in Fig. 3b, the spatial correlation selector adopts the traffic flow of the considered area (green box) and the traffic flow from other areas (blue, pink, and purple boxes) as the input and outputs of two kinds of spatial traffic features:

- *Positive (+ box) spatial traffic feature*, which models the influence from the areas that share similar patterns with the considered area
- *Negative (– box)* spatial traffic feature, which models the influence from the areas sharing reversed patterns compared to the considered data

To get these features, the spatial selector first calculates "correlation" between the considered area's traffic and the traffic from other areas, which is implemented as a feed-forward neural network. Then the traffic features with positive correlations are summed up as the positive spatial traffic feature, while the traffic features with negative correlations are summed up to obtain the negative spatial traffic feature. In this way, with the help of the spatial correlation selector, the single-dimensional feature of the considered area's traffic at every time step is extended into a three-dimensional feature, which embeds the spatial influence from other areas.

Discrete Embedding Module: The discrete embedding module is designed to model various discrete factors such as the PoI, the hour of day, and the day of the week, which play an important role in describing the unique property of an area with its traffic usage patterns. For convenience of computing, all these discrete factors are encoded in terms of one-hot vector. Especially, we use the frequency distribution of PoI categories to understand the function of the area, which is an important indicator of the potential trend and volume of traffic. We design a sparse linear embedding layer for every factor and concatenate the outputs into one assembled vector. For the capability of modeling and limited range, we add *tanh* function as the final nonlinear activation layer. With various discrete factors of the considered area as input, the embedding module outputs a fixed-length vector as its semantic representation.

Sequential Module

Based on the sequential features obtained from the former extractor module, we design a powerful sequential module to model the important temporal trend and the period of mobile traffic. Our model is an RNN-based model and employs three important properties in the applications of an RNN in sequential modeling. The architecture of our sequential module is presented in Fig. 3c. The first core component is the *LSTM unit*. Further, we utilize two chaining LSTMs to implement the *seq2seq module*. Finally, we apply the *attention mechanism* to strengthen its capability to capture the long and complicated dependencies.

LSTM: The RNN is a widely used sequential modeling tool in many scenarios like neural language processing and time series forecasting. Different from the general neural network, the RNN has a cycle connection and a memory cell to calculate and store the past state, which is expected to record the knowledge of the sequence history.

Dataset	Location	Time	Group	Items
Mobile cellular traffic	Shanghai	Aug. 2014	9600 BSs	1,000,000+ users
Points of interest (PoIs)			21 casses	839,128 records
Function	**Utilized POI**			
Residence	Residence, life services			
Entertainment	Food, hotel, gym, shopping, leisure.			
Business	Finance, office building, company, trading area.			
Industry	Factory, industrial estate, economic development zone.			
Education	School, campus.			
Scenery spot	Scenery spot.			
Suburb	Villages, towns			

TABLE 1. Basic statistics of utilized datasets.

While simple, the original RNN is not suitable for our bursty and long-period mobile cellular traffic forecasting because of its risk of gradient exploding and vanishing problem for modeling long time series. Thus, we choose LSTM with delicate control gates as the basic unit of our sequential model. In general, LSTM consists of a memory cell and three information control gates. It first generates the "forget gate" and "input gate" from the combination of input and hidden state. With the forget gate telling the memory cell which to forget and the input gate telling the memory cell which to remember, the memory cell is updated. Similarly, LSTM generates the "output gate" to cooperate with the new memory cell to generate the new hidden state, which is also the output of it. With the help of "information gates," LSTM is capable of stably capturing the long-range dependency of real mobile cellular traffic.

Seq2seq: While widely used in time series forecasting, one LSTM is not enough for mobile cellular traffic forecasting because of the burstiness and complexity. Besides, one LSTM is not able to predict multiple steps in advance. Thus, following the popular model in language modeling, we further utilize seq2seq structure to enable the multi-step prediction and enhance the capability of handling complex bursty cellular traffic. Our seq2seq model consists of two submodules: encoder and decoder. Each submodule is an independent LSTM or other kind of recurrent unit like a general resource unit (GRU), a popular variation of LSTM. During training, the input cellular traffic sequence is fed into the encoder, and the final hidden state is regarded as the representation of the traffic sequence. Then this final hidden state is fed into the decoder as the initial seed to generate variable length prediction of cellular traffic. Compared to only one LSTM, which simply considers the final hidden state for the next hop traffic prediction, the seq2seq model not only forecasts multiple steps, but also enables longer and deeper relation extraction from bursty and long-period cellular traffic.

Attention: Now, we introduce the attention module to cooperate with the seq2seq module to further improve the performance of modeling bursty cellular traffic. Based on the previous section, the seq2seq module extracts sequential features from the input traffic sequence and outputs the final hidden state as its only representation. However, because of the limited capacity of the recurrent encoder and the long period of traffic sequence, the final hidden state will lose some information of the input. Meanwhile, due to the burstiness of cellular traffic, the adjacent traffic points with extreme values may confuse the neural network to make the wrong prediction. Thus, more information from the long period before the current step is needed to make stable and accurate decisions. From the view of the model, this means the sequential model needs to revisit all the intermediate hidden states of the encoder to discover more valuable and reliable information. To implement this, our attention module first accepts all the intermediate hidden states of the encoder as the candidates. Then it calculates the "correlation" between the current state of the decoder and these candidates. Finally, with this normalized correlation as the weight, we obtain the weighted sum of these candidates as the most related and stable "summary" of the input traffic sequence for the current decoder state. With this knowledge as additional input, the decoder is expected to generate more accurate and stable forecasting results.

In conclusion, our model consists of two specifically designed components for the cellular traffic forecasting task. The first one is a feature extractor to extract spatial-temporal features of mobile cellular traffic and related features of external information. The second one is a sequential module, which is expected to model the complicated and reliable temporal transition from the bursty and long-period mobile cellular traffic. In the next section, we test the performance of our proposed model on traffic forecasting with large-scale real mobile cellular traffic data.

Evaluation on Large-Scale Mobile Cellular Traffic Data

Dataset

The dataset used in our experiments is collected from a large-scale mobile cellular network in Shanghai, a major city in China. The dataset records the data traffic load of approximately 9600 base stations every 10 minutes from August 1 to August 31, 2014. Each entry of the trace contains detailed mobile data usage of 1,000,000 users, including the anonymized IDs of devices, and start-end times of data connection, base station location, and the amount of third generation (3G) or LTE data used in each connection. The traffic logs 1.96 billion tuples of the described information, containing 2.4 PB logs, 77 TB/day. and 8 GB/base station on average. This large-scale and fine-temporal-grained traffic data guarantees the credibility of our mobile cellular traffic modeling and forecasting. Besides, we collect PoIs data of Shanghai from the Open API (application programming interface) of the Baidu map. Based on collected PoIs data, we decide the function of each base station like Xu *et al.* [13] did. The basic information of the datasets is presented in Table 1.

To evaluate the accuracy of our predictive model, we compared the proposed model with several of the most up-to-date methods:

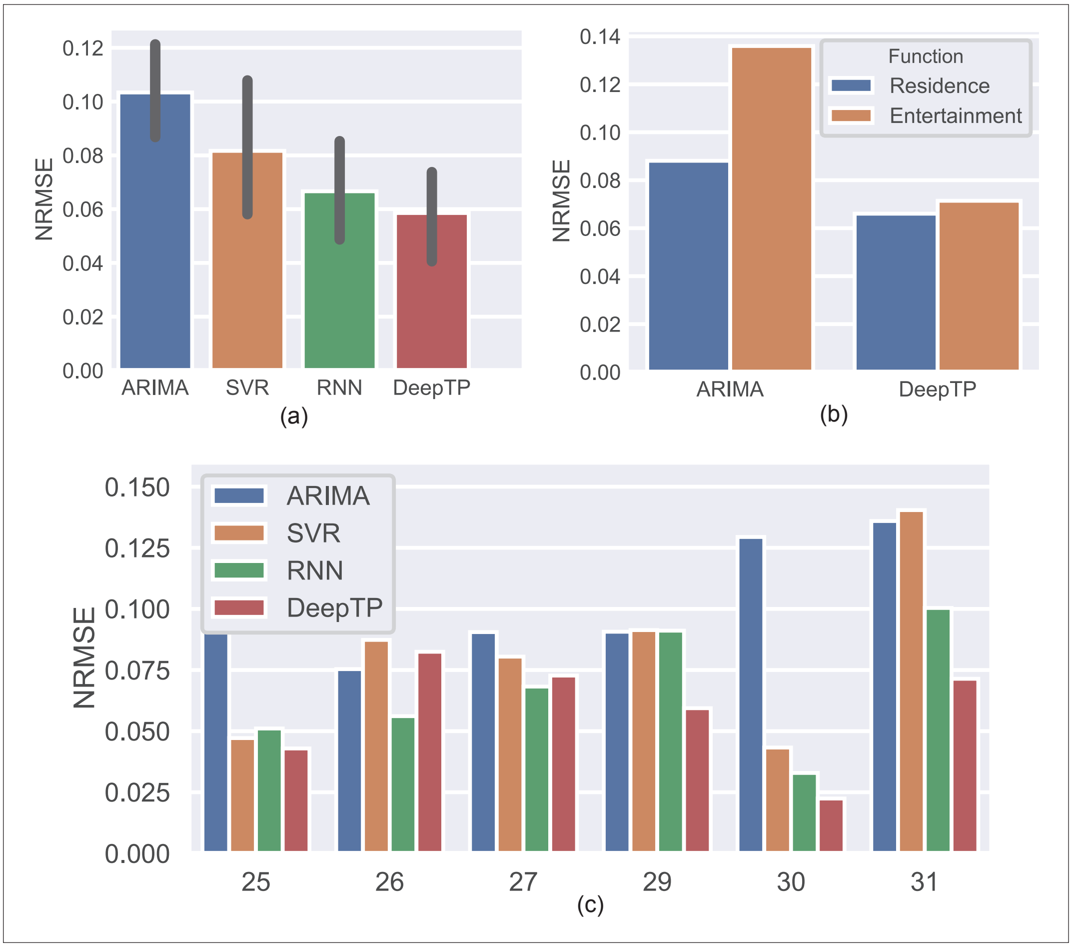

FIGURE 4. Performance comparison between DeepTP and baselines: a) overall forecasting performance on selected base stations; b) forecasting performance varies in different functional zones; c) independent performance on each selected base station.

- ARIMA: As the most widely used time series forecasting method, ARIMA was first applied to traffic load prediction by Shu *et al.* [4]. We use the python implementation from the *statsmodels* library.
- SVR: The regression version of SVM is also a representative classic method for time series prediction. We use the python implementation from the *sklearn* library.
- RNN (GRU): As introduced above, the RNN has been widely used in many time series forecasting problems including traffic prediction [3, 5].

Our model works in an end-to-end manner without requiring handcrafting features. We choose MSELoss as the loss function, while the Adaptive Moment Estimation (Adam) algorithm with default learning rate is utilized to optimize the model. Several widely used tricks, such as dropout, L2 regularization, selu unit, gradient clipping, and learning rate schedule, are used to avoid the overfitting problem. While normalized root mean square error (NRMSE) is used as the basic metric for evaluation, we randomly choose 15 base stations for testing.

Prediction Results

In this section, we first directly compare the prediction results of our model with the actual traffic from two randomly selected base stations. Then we present the performance comparison between DeepTP and three baseline models in terms of NRMSE in Fig. 4.

> In this article, we propose Deep Traffic Predictor (DeepTP), a deep-learning-based model to forecast traffic demands from spatial-dependent and long-period cellular traffic. It works in an end-to-end manner without requiring any handcrafting features

As Fig. 4a shows, the prediction results of two classic methods ARIMA and SVR are more than 0.08 in terms of NRMSE. Specifically, because of the bursty nature and nonlinear relations in our traffic data, SVR performs better than ARIMA. Compared to the classic methods, the RNN-based model reduces the average NRMSE of prediction to 0.06, which demonstrates the effectiveness of the RNN in modeling a complicated sequence. Furthermore, by considering the spatial dependencies and introducing external information, our model is capable of modeling the actual cellular traffic and outperforms the general RNN (GRU) model by more than 12.31 percent in the real system. Further, we find that the forecasting performance varies among different functional zones. For example, as Fig. 4b shows, compared to network traffic from the amusement district, the performance of our model and ARIMA in the residential area is better. The high volatile nature of

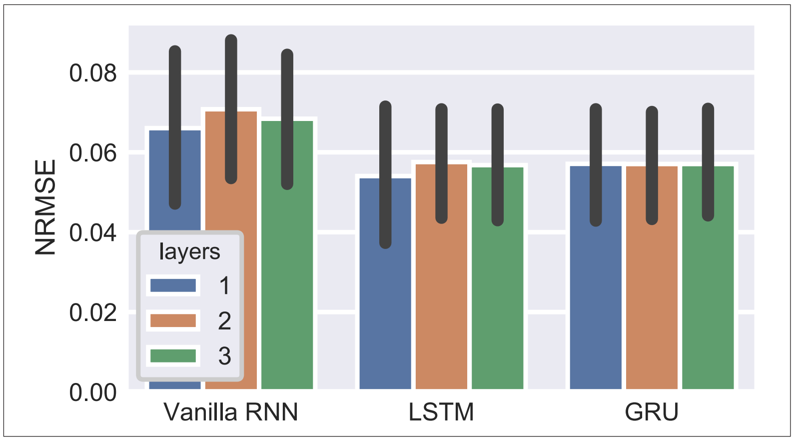

FIGURE 5. Model variations with different parameter settings including different recurrent units and stacking layers.

population and traffic demands in amusement can account for this to a great extent. Our model succeeds in modeling this high volatility and achieves better performance than ARIMA. Figure 4c presents the performance of our model and baselines on some randomly selected base stations. We can see from Fig. 4c that our model is almost better than the RNN (GRU) baseline model in every base station.

In conclusion, by considering the influence of external information and spatial dependencies, our proposed model outperforms traditional methods like ARIMA more than 40 percent and simple RNN-based methods more than 12 percent in real mobile cellular traffic data.

Effects of Hyperparameters

Based on the framework of DeepTP, we further explore the influence of various neural units and hyperparameter settings in the evaluation. Specifically, we try different recurrent units (e.g., GRU, LSTM, and general RNN) and various hidden layers in the evaluation, and the results are presented in Fig. 5. As Fig. 5 shows, the DeepTP model is not sensitive to the hyperparameter settings, demonstrating the robustness of our model design. Besides, the results demonstrate that the shallow recurrent network is enough for time series forecasting, while stacking more layers is not helpful to extract representative features. In the experiments, the model with one-layer LSTM achieve the best results.

Conclusion

In this article, we investigate the problem of mobile cellular traffic forecasting in a large-scale real system. We propose a deep-learning-based end-to-end model DeepTP, which captures two natural characteristics of mobile cellular traffic: a general feature extractor to model spatial dependencies and external information, and a seq2seq model with an attention mechanism as the sequential model to capture reliable temporal information from bursty and long-period traffic data. Extensive experiments on large-scale mobile cellular traffic show that DeepTP outperforms all the baselines. Inspired by our work, we believe that utilizing multiple network traffic and jointly considering more practical factors to do multivariable forecasting is a promising direction. It can not only improve the performance of forecasting but also give us more chances to understand the variations of the network traffic demands.

Except for the promising performance of our model, we would like to discuss some limitation of our model and deep models. The largest weakness of our model is the computing efficiency. Coupled with a modified attention module as our spatial correlation selector, which is known as a time-consuming component, our model can be much slower than the standard recurrent model and traditional methods. Furthermore, even the standard recurrent model itself is slower than traditional methods, because the recurrent model needs to keep the memory during training and inference, which prevents it from utilizing the paralleli power of GPU. To solve this problem, the highly parallel convolutional network can be regarded as a promising potential solution.

Except for the aforementioned paralleling direction, in the future, we plan to improve the efficiency of the feature extractor to enable the application of DeepTP in a real-time system. Considering the high correlations between the traffic demands and human activity [14, 15], we plan to introduce more user behavior [14] into our model and build a multi-task framework to predict traffic demands and human flow simultaneously. Multitask learning may be a good starting point for this.

Acknowledgement

This work was supported in part by The National Key Research and Development Program of China under grant 2017YFE0112300, the National Nature Science Foundation of China under 61861136003, 61621091 and 61673237, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016-2021 White Paper, 2017.
[2] Y. Zhang *et al.*, "Home M2M Networks: Architectures, Standards, and QoS Improvement," *IEEE Commun. Mag.*, vol. 49, no. 4, Apr. 2011, pp. 44–52.
[3] Y. Hua *et al.*, "Traffic Prediction Based on Random Connectivity in Deep Learning with Long Short-Term Memory," CoRR, vol. abs/1711.02833, 2017.
[4] Y. Shu *et al.*, "Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models," *Proc. IEEE ICC*, 2003.
[5] J. Wang *et al.*, "Spatiotemporal Modeling and Prediction in Cellular Networks: A Big Data Enabled Deep Learning Approach," *Proc. IEEE INFOCOM*, 2017, pp. 1–9.
[6] W. Hong, "Application of Seasonal SVR with Chaotic Immune Algorithm in Traffic Flow Forecasting," *Neural Computing & Application*, vol. 21, 2012, pp. 583–93.
[7] B. Zhou, D. He, and Z. Sun, "Traffic Modeling and Prediction Using ARIMA/GARCH Model," 2005.
[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 98, 1997, pp. 1735–80.
[9] Y. Bengio, "Learning Deep Architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, 2009, pp. 1–127.
[10] H. Wang *et al.*, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *AMC IMC*, 2015, pp. 225–38.
[11] F. Xu *et al.*, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *IEEE/ACM Trans. Networking*, vol. 25, 2015, pp. 1147–61.
[12] F. Xu *et al.*, "Mobile Cellular Big Data: Linking Cyberspace and the Physical World with Social Ecology," *IEEE Network*, vol. 30, 2016, pp. 6–12.
[13] F. Xu *et al.*,"Context-Aware Real-Time Population Estimation for Metropolis," *Proc. Ubicomp*, 2016.

[14] X. Zhang *et al.*, "Enhancing Spectral-Energy Efficiency for LTE-Advanced Heterogeneous Networks: A User's Social Pattern Perspective," *IEEE Wireless Commun.*, vol. 21, no. 2, Apr. 2014, pp. 10-17.

[15] F. Xiao *et al.*, "Artificial Intelligence Empowered Mobile Sensing for Human Flow Detection," to appear, *IEEE Network*.

Biographies

Jie Feng is a Ph.D. candidate student at the Department of Electronics Engineering of Tsinghua University, Beijing, China, advised by Prof. Yong Li. He reccived his B.E. degree in electrical engineering from Tsinghua University in 2016. His research interest falls in the area of spatial-temporal data mining and deep learning. He currently works on applying deep learning methods in spatial-temporal data mining to improve the performance of practical models in many challenging practical tasks like trajectory forecasting and crowd flow forecasting.

Xinlei Chen received his B.E. and M.S. degrees in electrical engineering from Tsinghua University, China, in 2009 and 2012, respectively. He received his Ph.D. degree from the Department of Electric and Computer Engineering, Carnegie Mellon University in 2018. His research interests are in the areas of networking and communications, mobile embedded systems, big data, and so on.

Rundong Gao is an undergraduate at the Department of Electrical Engineering, Tsinghua University. In 2017, he started to work on deep learning. In 2018, he joined THU EE Lab, guided by Professor Li, to work on data mining. His research interests are in representation learning, natural language processing, and data mining.

Ming Zeng received her B.E. and Ph.D degrees in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, in 2009 and 2016, respectively. She is currently a postdoctoal researcher at Tsinghua University. Her research interests include data-driven network optimization, intelligent transportation systems, fog computing, and spatial-temporal data mining.

Yong Li (M'09, SM'16) received his B.S. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and his Ph.D. degree in electronic engineering from Tsinghua University in 2012. He is currently a faculty member in the Department of Electronic Engineering, Tsinghua University. He has served as General Chair, TPC Chair, and TPC member for several international workshops and conferences, and he is on the Editorial Boards of two IEEE journals. His papers have more than 4600 total citations. Among them, 10 are ESI Highly Cited Papers in Computer Science, and four received conference Best Paper (runner-up) Awards. He received the IEEE 2016 ComSoc Asia-Pacific Outstanding Young Researchers Award and the Young Talent Program of China Association for Science and Technology Award.