Traffic-Tran: A Parallel Multi-encoder Structure for Cellular Traffic Prediction

Shilong Fan[1], Boyuan Zhang[1], and Xinyu Gu[1,2](✉)

[1] Beijing University of Posts and Telecommunications, Beijing, China
guxinyu@bupt.edu.cn
[2] Purple Mountain Laboratories, Nanjing 211111, China

Abstract. Wireless cellular traffic prediction is a critical research topic for the realization of intelligent communications. The high nonlinearities and mutability of wireless cellular network traffic bring great challenges to accurate prediction. Due to the lack of dynamic spatio-temporal correlation modeling ability and complex network structure, the existing prediction methods cannot meet the requirements of accuracy and complexity in real scenes. By generating time series data for network traffic of a single grid, and spatial series data for network traffic of all grids with the same timestamp, this paper proposes a multi-encoder structure named "Traffic-Tran", which learns sequence correlation independently and in parallel by multiple network units. Meanwhile, in order to improve the recognition ability of multi-encoder feature information, an information supplement method is proposed. In addition, the design of sampling output module realizes the parallel multi-step flow prediction, which enlarges the application range of the model. Experimental results on a large real dataset verify the effectiveness of Traffic-Tran. The model complexity of Traffic-Tran is greatly reduced, with less memory usage and shorter runtime than other models. Under the premise of the same predictive performance, the number of training parameters of Traffic-Tran is reduced by 44.9%.

Keywords: Cellular traffic prediction · Spatio-temporal correlation · Transformer · Multi-step prediction

1 Introduction

In recent years, with the comprehensive coverage and popularization of new information infrastructure such as the Fifth Generation (5G) mobile network, people's demand for wireless cellular network is growing rapidly. According to the "Cisco Annual Internet Report" [1], the total number of global Internet users will reach 5.3 billion (66% of the global population) by 2023, up from 3.9 billion in 2018. How to configure cellular network resources, optimize network management strategy, improve network service quality and reduce energy consumption needs to be further thought and solved in the next generation of mobile communication.

F. Gao et al. (Eds.): ChinaCom 2022, LNICST 500, pp. 337–346, 2023.
https://doi.org/10.1007/978-3-031-34790-0_26

Accurate wireless traffic prediction helps optimize network management strategies to improve service quality and prevent network congestion, while reducing base station power consumption and operating costs. However, accurately predicting wireless cellular network traffic is a very challenging problem for the following two reasons. First, due to the emergence of new types of transportation, people can get from one end of the city to the other in a short time. The mobility of wireless users makes the cellular traffic between regions spatially dependent. Secondly, due to the regularity of the daily life and work of wireless users, the wireless cellular traffic also shows the pattern of regular changes in time. The traffic value at a certain moment is highly correlated with the traffic value at a similar moment (short-term dependence) and a relative moment of a certain day (periodicity). The modeling ability of time dependence and space dependence determines the performance of the final prediction results.

Data-driven artificial intelligent algorithms play an important role in improving network service quality. Among deep learning methods, convolutional neural networks (CNN) [2] and recurrent neural networks (RNN) [3] are the basic structures for modeling spatio-temporal dependence. Among them, spatial features are modeled mainly by convolutional neural networks [4] and graph neural networks (GCN) [5]. Some researchers also use transfer learning [6] and meta-learning [7] to improve the modeling ability. In addition, some researchers tried to enhance the representation of feature information through cross-domain data, so as to obtain more accurate prediction results [8].

Transformer [9] is a parallel encoder-decoder structure with many successful applications, and the ability of its internal attention layer to model sequence data has been proven in many practical scenarios. In terms of wireless cellular network traffic prediction, Transformer is also trying to be used to model the spatial-temporal correlations, and ST-Tran [10] is the first to apply the transformer architecture to predict cellular traffic. However, the combination of four transformer blocks in ST-Tran increases the parameter size, resulting in more memory usage and longer runtime. Therefore, this paper aims to build a simpler and more efficient model based on Transformer.

The sequence modeling ability of Transformer is utilized in this paper to model the space-time sequence of wireless cellular network traffic. We design a multi encoder-single decoder structure named Traffic-Tran to model sequence data by using the attention mechanism in each encoder. Multiple encoders in Traffic-Tran capture the spatio-temporal characteristic in parallel. After decoding this information, the output module will be used for multi-step prediction. Traffic-Tran can achieve effective cellular network traffic prediction with fewer training parameters.

2 System Model

2.1 Milan Data Description

The wireless cellular traffic data used in this paper comes from the Milan Telecom Data set [11], which is jointly initiated by Telecom Italia and European Institute of Innovation and Technology (EIT), and records the real wireless traffic data of Milan, Italy. Milan has a total population of about 1.3 million and covers an area of about 522 square kilometers. To collect wireless traffic data, Milan is evenly divided into $H \times W$ grids, each with a size of about (235 m $\times$ 235 m), where H and W are both 100. This data set records the traffic data of wireless mobile users in Milan, including five traffic services: receiving and dialing call services(Call-in, Call-out), receiving and sending short message services(SMS-in, SMS-out) and mobile internet services(Internet).

In the data set, wireless service traffic was recorded for 62 days from November 1, 2013 to January 1, 2014. Five kinds of wireless traffic of each grid are recorded and stored every ten minutes, that is, every ten minutes is a time granularity. The data at each time granularity is called a timestamp. Since there are many blank values in the data set collected by the 10-minute time granularity, which will lead to experimental errors. Therefore, the 10-minute time granularity is further aggregated into the one-hour time granularity before the experiment. Therefore, we finally got a time series data set of hourly time granularity in 62 days with a sequence length of 1488.

2.2 Problem Statement

The prediction of cellular traffic is further introduced as follows. This paper will finally realize the multi-step prediction of cellular flow, and the time step of the prediction result is expressed by T_{target}. As mentioned in the previous chapter, the predicted traffic value is highly correlated with similar times on the same day and relative times in the preceding days. In this paper, as shown in Fig. 1, the similar times on the same day is called close data $X_{close} \in \mathbb{R}^{N \times T_{close}}$, and its time step is T_{close}. In the same way, the relative times in the preceding days is called period data $X_{period} \in \mathbb{R}^{N \times T_{period} \times p}$, its time step is T_{period} for each of the previous p days, where N is the number of grids.

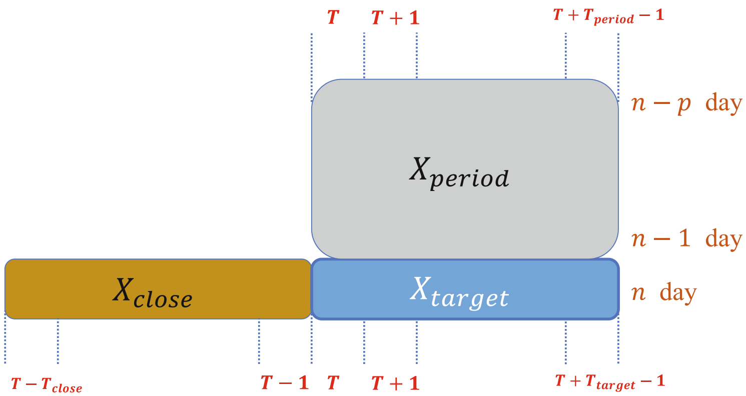

Fig. 1. Construction of X_{close}, X_{period} and X_{target}.

In order to capture the correlation between different grids, we use each grid data under the same timestamp to generate spatial sequence data. However, due to the large number of grids N, the correlation between all grids cannot be constructed. Therefore, we use (1) to quantify the correlation between grids and select K grids most related to grids as sequence data $X_{spatial}$.

$$R_{X,Y} = \frac{cov(X,Y)}{\sigma_x \cdot \sigma_y}, \tag{1}$$

where X and Y represent two grids.

The system model of wireless cellular network traffic prediction is roughly expressed as:

$$X_{target} = f(X_{close}, X_{period}, X_{spatial}), \tag{2}$$

where X_{close} and X_{period} are time sequence data, $X_{spatial}$ is spatial sequence data, and $f(*)$ is a nonlinear function constructed by deep learning algorithm.

The objective of the entire network is to reduce the error between the predicted value and the real value. We quantify the final performance using the following three evaluation indexes:

$$MAE = \frac{1}{n}\sum_{1}^{n}\left|\hat{y}_i - y_i\right|, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}\left(\hat{y}_i - y_i\right)^2}, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{1}^{n}\left(y_i - \bar{y}\right)^2}, \tag{5}$$

where n is the number of the sample, $\hat{y}_i$ is the predicted value of the sample, y_i is the label of the sample and $\bar{y}$ is the mean of the sample.

3 Proposed Model—Traffic-Tran

As shown in Fig. 2, the proposed model is designed based on the basic structure of Transformer. As mentioned in the previous chapter, the traffic data is preprocessed to obtain the close time series, periodic time series and spatial series. Three independent encoder structures are used to capture the temporal and spatial characteristic information of cellular traffic in parallel, then the characteristic information is fused in the decoder structure. In order to enable the decoder structure to identify the feature information of different encoders, the outputs of the different encoders are further encoded by the designed position function. In addition, an initial sequence is generated according to the spatio-temporal node information specific to each prediction timestamp as the input sequence of the decoder structure. After the feature information fusion through the decoder, the final cellular traffic prediction is carried out through the output module.

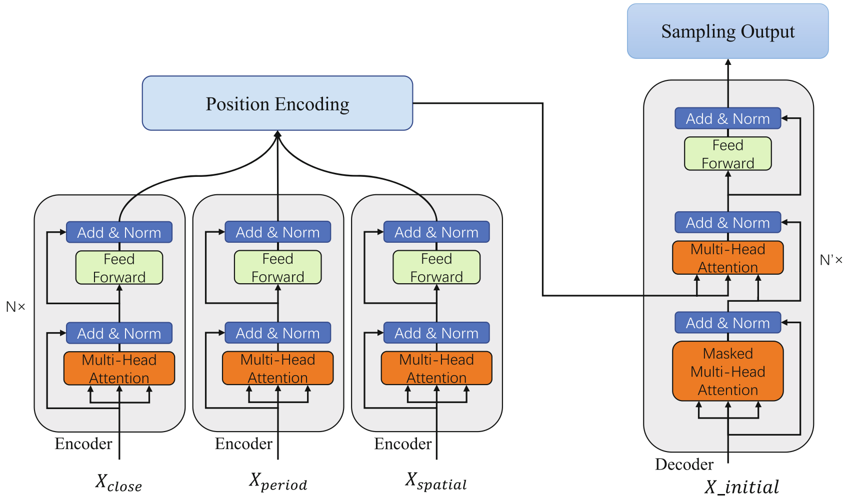

Fig. 2. Overview of Traffic-Tran.

3.1 Multi Encoder-Single Decoder Structure

The close data, periodic data and spatial data have different correlations, so it is necessary to construct different nonlinear functions to represent the correlations between the data and predict values. In this paper, multiple independent encoders structures are proposed to capture the feature information carried by different data in parallel. In addition, a single decoder is used to fuse the information captured by different encoders in order to balance the dependence between predicted values and different feature information. The core operation in encoders and decoders is the multi-head attention layer. The multi-head attention layer can effectively process sequence data. Note that the decoder needs to input an initial sequence to process the information from the encoder. In order to better realize the multi-step prediction of cellular flow, the predicted value of each step will generate a unique initial sequence called $X_{initial}$ according to its spatio-temporal position. Specifically, the initial sequence of the predicted values of each step carries the close information of T_{close} step and the periodic information of p step.

3.2 Position Encoding

Since recurrence and convolution are not included in our model, in order to recognize order of the sequences based on our model, we must inject some information about the relative or absolute positions of sequences. Among Traffic-Tran, two modules use position encoding to inject order information. One is to inject internal relative position information into the input sequence of each encoder structure, the other is to inject external relative position information into the output sequence of all encoders. Among them, the effectiveness of infusing information into the encoder input sequence has been proved in many scenarios, and this paper creatively proposes to infuse information into the output sequence of encoder. Different from the previous single-encoder structure, Traffic-Tran uses the design of multi-encoder structure to capture more feature information in parallel. In this paper, we use sine and cosine functions of different frequencies [9] as supplementary information for relative positions:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right), \tag{6}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right). \tag{7}$$

3.3 Sampling Output

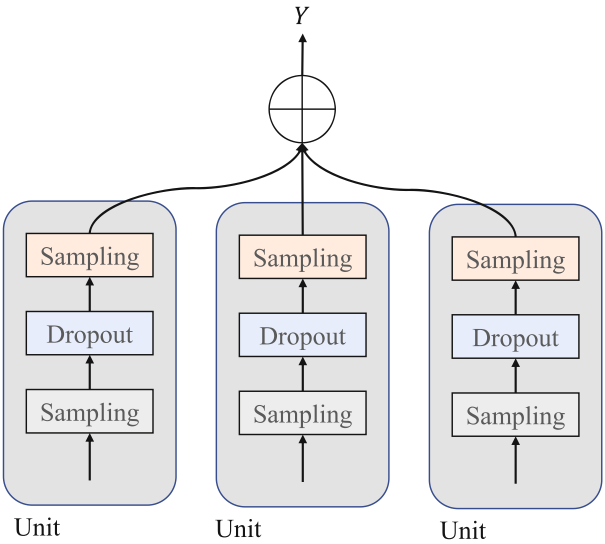

Fig. 3. Sampling output structure.

The prediction of each time step corresponds to the initial sequence of the decoder input one by one, and the prediction of each time step does not share parameters in the output module. This paper presents a multi-step prediction of cellular traffic, so this output module contains multiple independent network units. This section takes the prediction of three time steps as an example to illustrate the output module. The prediction of three timestamps is shown in Fig. 3, thus containing three network units. Each network unit is composed of the dropout layer, which is to prevent overfitting, and two sampling layers, where the sampling layer downsample the feature information. After the initial sequence of each prediction time step is fused with feature information in the decoder, the network unit in the output module is used to predict the output of this time step. Multiple network units in the output module are parallel to achieve multi-step prediction.

Most of the research works is to predict the output directly through fully connected layer after the feature information fusion. The output module uses a more advanced network structure to get more accurate prediction results than the output directly through the fully connected layer. In addition, because the output module can independently use the sequence vector output by the decoder structure, the design of output module realizes multi-step prediction of traffic data using different learnable parameters in parallel.

4 Experimental Results and Analysis

4.1 Experiment Setup

Experiments were conducted on Call-in data from Telecommunication activities mentioned in Sect. 2.1. We chose 400 from all 10000 grids. The time dimension T_{close}, T_{period} and T_{target} of the sequence data are all 3, and the period days $p = 3$. In other words, each sample contains a total of 15 timestamps, and each timestamp contains 400 grid Call-in data, among which the sequence data of twelve timestamps will be used as the input of the model, and the remaining three timestamps will be used as labels. In the data set containing 1488 time stamps, a total of 1413 samples are generated, among which 168 samples are used as test dataset D_{test} and the remaining 1245 samples are used as training dataset D_{train}.

In the same experimental environment (GeForce RTX 2080Ti), the comparison experiments of Traffic-Tran and other strategy are conducted. The network models are trained with the widely used optimization technique, Adam [12] with 500 epochs. The size of mini-batch is determined according to the complexity of the model. The initial learning rate is set to 0.001, and the effective learning rate follows a polynomial decay.

4.2 The Depth of Decoder

The decoder of Transformer is composed of a stack of $N = 6$ identical layers. In order to improve the ability of Traffic-Tran to fuse feature information, we conducted an experiment on the structural depth of the decoder. The effect of decoder depth is shown in Table 1. MAE and $RMSE$ are the lower the better, and R^2 is the closer to 1 the better. When N is 8, the values of MAE and $RMSE$ are the smallest, and R^2 is the closest to 1, achieving the best prediction performance. So the decoder of our model is composed of a stack of $N = 8$ identical layers.

Table 1. Effect of decoder depth.

N	MAE	$RMSE$	R^2	$Params$
6	10.29568	0.56877	0.80760	3593219
7	10.08085	0.56832	0.80790	3791235
8	10.04836	0.55609	0.81608	3989251
9	10.13335	0.56944	0.80714	4187267
10	10.16580	0.56759	0.80840	4385283

4.3 Performance Evaluation

We compared the results of Traffic-Tran with two widely used methods:

- STDenseNet: STDenseNet [4] learns spatial-temporal features using densely connected CNNs and fuse feature information by fully connected layer.
- ST-Tran: ST-Tran [10] is the first to apply the transformer architecture to predict cellular traffic. The combination of four transformer blocks is used to model temporal and spatial correlations.

Table 2. Comparison on Traffic-Tran and two widely used methods.

$Model$	MAE	$RMSE$	R^2	$Params$
STDenseNet	12.74558	0.62273	0.76935	249816
ST-Tran	10.00348	0.53643	0.82885	7239844
Traffic-Tran	10.04836	0.55609	0.81608	3989251

The effectiveness of Traffic-Tran compared to other methods is shown in Table 2. Compared with STDenseNet, Traffic-Tran has a significant advantage in predicting performance, but it also uses more trainable parameters. Compared with ST-Tran, Traffic-Tran's prediction performance is similar to that of ST-Tran, but the training parameters are greatly reduced, which achieves significant advantage in model complexity. Figure 4 shows the results of Traffic prediction using Traffic-Tran for one of the grids.

In general, Traffic-Tran can complete multi-step Traffic prediction, and its prediction performance is almost the same as that of the best prediction scheme. Moreover, the model complexity of Traffic-Tran is greatly reduced, with less memory usage and shorter runtime than other models.

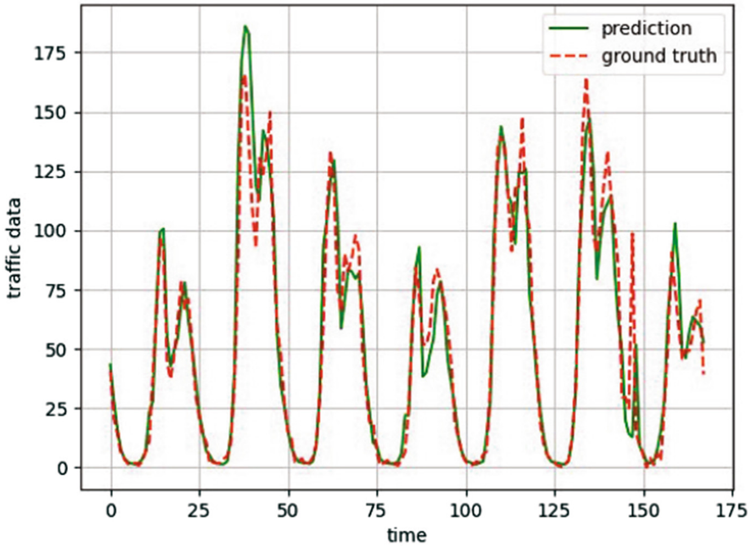

Fig. 4. Results of traffic prediction using Traffic-Tran.

5 Conclusion

Traffic-Tran can complete multi-step Traffic prediction, and the model complexity of Traffic-Tran is greatly reduced, with less memory usage and shorter runtime than other models. Traffic-Tran can construct complex nonlinear data caused by user movement, service randomness and regional limitations in wireless services, and obtain a prediction model with strong generalization performance. Experimental results on a large real dataset verify the effectiveness of Traffic Tran. Compared with existing prediction schemes, Traffic-Tran can realize parallel multi-step prediction, and its model architecture is simpler and more efficient. The number of parameters to be trained in the model decreases by 44.9%. However, its prediction accuracy has not improved, and more advanced architectures will be developed to further improve performance in the future. When the data set is limited, the combination of Traffic-Tran and transfer learning can be considered, so that only parameters in the output module can be trained in the target domain to improve the prediction performance.

Acknowledgment. This work was supported in part by the State Major Science and Technology Special Projects (Grant No. 2018ZX03001024) and in part by the National Key Research and Development Program (Grant No. 2022YFF0610303).

References

1. Cisco annual internet report (2018–2023) white paper. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-Internet-report/white-paper-c11-741490.html. Accessed 9 Mar 2020

2. Mozo, A., Ordozgoiti, B., Gómez-Canaval, S.: Forecasting short-term data center network traffic load with convolutional neural networks. PLOS One **13**(2), 1–31 (2018). https://doi.org/10.1371/journal.pone.0191939
3. A. Dalgkitsis, Louta, M., Karetsos, G.: Traffic forecasting in cellular networks using the LSTM RNN, pp. 28–33 (2018)
4. Zhang, C., Zhang, H., Yuan, D., Zhang, M.: Citywide cellular traffic prediction based on densely connected convolutional neural networks. IEEE Commun. Lett. **22**(8), 1656–1659 (2018)
5. Zhao, N., Wu, A., Pei, Y., Liang, Y.-C., Niyato, D.: Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction. IEEE Commun. Lett. **26**(3), 587–591 (2022)
6. Zhang, C., Zhang, H., Qiao, J., Yuan, D., Zhang, M.: Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data. IEEE J. Sel. Areas Commun. **37**(6), 1389–1401 (2019)
7. He, Q., Moayyedi, A., Dán, G., Koudouridis, G.P., Tengkvist, P.: A meta-learning scheme for adaptive short-term network traffic prediction. IEEE J. Sel. Areas Commun. **38**(10), 2271–2283 (2020)
8. Zeng, Q., Sun, Q., Chen, G., Duan, H., Li, C., Song, G.: Traffic prediction of wireless cellular networks based on deep transfer learning and cross-domain data. IEEE Access **8**, 172 387–172 397 (2020)
9. Vaswani, A., et al.: Attention is all you need. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates Inc. (2017)
10. Liu, Q., Li, J., Lu, Z.: ST-Tran: Spatial-temporal transformer for cellular traffic prediction. IEEE Commun. Lett. **25**(10), 3325–3329 (2021)
11. Barlacchi, G., et al.: A multi-source dataset of urban life in the city of Milan and the Province of Trentino. Sci. Data **2**, 150055 (2015)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR, vol. abs/1412.6980 (2015)