

RESEARCH

Open Access



Attention based multi-component spatiotemporal cross-domain neural network model for wireless cellular network traffic prediction

Qingtian Zeng¹, Qiang Sun¹, Geng Chen^{1*}  and Hua Duan²

* Correspondence: gengchen@sdust.edu.cn

¹College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China
Full list of author information is available at the end of the article

Abstract

Wireless cellular traffic prediction is a critical issue for researchers and practitioners in the 5G/B5G field. However, it is very challenging since the wireless cellular traffic usually shows high nonlinearities and complex patterns. Most existing wireless cellular traffic prediction methods lack the abilities of modeling the dynamic spatial-temporal correlations of wireless cellular traffic data, thus cannot yield satisfactory prediction results. In order to improve the accuracy of 5G/B5G cellular network traffic prediction, an attention-based multi-component spatiotemporal cross-domain neural network model (att-MCSTCNet) is proposed, which uses Conv-LSTM or Conv-GRU for neighbor data, daily cycle data, and weekly cycle data modeling, and then assigns different weights to the three kinds of feature data through the attention layer, improves their feature extraction ability, and suppresses the feature information that interferes with the prediction time. Finally, the model is combined with timestamp feature embedding, multiple cross-domain data fusion, and jointly with other models to assist the model in traffic prediction. Experimental results show that compared with the existing models, the prediction performance of the proposed model is better. Among them, the RMSE performance of the att-MCSTCNet (Conv-LSTM) model on Sms, Call, and Internet datasets is improved by 13.70 ~ 54.96%, 10.50 ~ 28.15%, and 35.85 ~ 100.23%, respectively, compared with other existing models. The RMSE performance of the att-MCSTCNet (Conv-GRU) model on Sms, Call, and Internet datasets is about 14.56 ~ 55.82%, 12.24 ~ 29.89%, and 38.79 ~ 103.17% higher than other existing models, respectively.

Keywords: 5G/B5G, Cellular network, Attention, att-MCSTCNet, Traffic prediction

1 Methods/experimental

In this paper, in order to improve the accuracy of 5G/B5G cellular network traffic prediction, a multi-component spatiotemporal cross-domain neural network model based on attention mechanism was proposed. The wireless cellular traffic data were divided into neighborhood data, daily data, and weekly data according to its periodic characteristics. The six-part structure of the model was introduced and explained in detail.



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Secondly, the algorithm of the model training process was given. Finally, under different datasets, different models with different structures were used for experiments. The comparative experiment of att-MCSTCNET model using Conv-GRU structure and Conv-LSTM structure, and the parameter optimization experiment of att-MCSTCNET model were carried out. The results of Experiment 5.2 show that the model can effectively utilize the periodic characteristics of wireless cellular traffic data, save training time and greatly reduce the workload of the model, and further improve the prediction performance of the model. In addition, Experiment 5.3 proves that the iteration time of Conv-GRU is shorter than that of Conv-LSTM, and the convergence speed of Conv-GRU is faster. Finally, Experiment 5.4 gives the specific hyperparameters which are most suitable for att-MCSTCNET model.

2 Introduction

With the rapid development of mobile internet and internet of things services, the demands and challenges brought about by the fifth-generation (5G) and beyond fifth-generation (B5G), the development of wireless communication technology has entered a new stage. Supported by new theoretical technologies such as big data [1, 2] and artificial intelligence [3, 4], wireless communication is characterized by flexible diversification and cross-domain fusion [5]. In this context, wireless service traffic prediction [6] has become a hot issue in 5G wireless communication networks. Accurate prediction of wireless cell traffic is helpful for base station site selection, urban area planning, and regional traffic prediction. However, accurate prediction of wireless service traffic is a very challenging problem, which is mainly due to the following three reasons. First, the source of wireless communication network traffic is mobile users, and the mobility of wireless users makes the traffic between multiple areas spatially dependent. In particular, the emergence of new types of transportation makes it possible for people to get from one end of the city to the other in a short time. This makes the spatial dependency of wireless service traffic not only local, but also a large-scale global dependency. On the other hand, the wireless traffic is also dependent on the time dimension. 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). Second, the spatial constraint of wireless service traffic is caused by multi-source cross-domain data. The causes that affect wireless business traffic in a certain area are diverse. When making wireless cellular traffic prediction, not only should the hidden regular patterns of wireless business traffic be mined from the perspective of historical data, but also the spatial constraints of other cross-domain and cross-source data on traffic should be considered. For example, factors such as base station data in a certain area, point of interest information, and the level of social activities in the area will all have an impact on changes in traffic. Therefore, how to efficiently integrate these multi-source and cross-domain data that do not seem to be directly related to wireless service traffic is a difficult problem to be solved. Third, it is also a difficult problem how to achieve higher prediction accuracy of wireless cellular traffic in the case of considering time and space factors and combining cross-domain data.

The prediction of wireless cellular networks can actually be regarded as the analysis of time series. Cellular traffic is not only related to historical traffic data in the area, but also affected by many external factors. The deep learning technology can accurately

grasp the spatial and temporal correlation of cellular traffic data and accurately predict wireless cellular traffic with neural networks. Therefore, deep learning for the wireless cellular network traffic prediction model is widely studied. Early wireless cellular traffic prediction models used some simple shallow learning algorithms, such as the linear regression (LR) model [7] and support vector regression (SVR) model [8]. In recent years, due to the maturity of deep learning technology, wireless cellular network models based on deep learning are increasing. Wang et al. [9] proposed a new autoencoder-based spatial model for spatial modeling and long short-term memory unit (LSTM) for temporal modeling. Its prediction accuracy is better than traditional models such as the support vector regression (SVR). To further realize the modeling of the space, the neural network based on graph convolution [10] predicts the cellular flow of any shape and size in the city. Qiu et al. [11] also use LSTM for time-dependent capture, but compared with Jing et al. [10] in spatial feature learning, the multi-task learning idea is used to fully integrate business traffic in different regions, and the impact of other cross-domain data is not taken into account. On this basis, Hu et al. [12] used LSTM to model the spatial and temporal dependencies of different scales in the crowd flow problem, and merged a variety of cross-domain data (weather, air quality, holiday information, etc.), which further improved the model prediction accuracy. Zhang et al. [13] and Hu et al. [12] have similar ideas. In wireless cellular network traffic prediction, multiple cross-domain data are added to the prediction model as auxiliary traffic prediction, and the space and time factors of wireless cellular traffic are captured by Conv-LSTM and CNN modules. The results show that the performance of wireless cellular network traffic prediction is better when all factors are combined. However, Qu et al. [14] further proves the importance of cross-domain data to the prediction model in the airport delay prediction model, and the results show that the prediction accuracy of the airport delay model is higher than that of adding only one cross-domain dataset when integrating multiple cross-domain datasets.

In recent years, attention mechanisms have been widely used in various tasks such as natural language processing, image caption, and speech recognition [15, 16]. The goal of the attention mechanism is to select information that is relatively critical to the current task from all input. The neural network is constructed through the attention mechanism to receive attention-related input and pay adaptive attention to the input data features so as to extract features more effectively. In the field of short-term traffic flow prediction, Feng et al. [17] proposed an attention-based space time graph convolutional network (ASTGCN) model, effectively capturing the daily periodicity, weekly periodicity, and nearest neighbors in traffic data. Convolution is used to capture the spatial pattern, and the output of these three components is weighted and fused by the attention mechanism module. The final prediction result shows that the prediction performance is better than other models. In conclusion, the challenges and problems of wireless cellular network traffic prediction mainly include the following three points: firstly, how to make full use of the time and space characteristics of wireless cellular traffic data itself, secondly, how to integrate multiple cross-domain data for prediction, and lastly, which network structure should be adopted to fulfill the above two requirements.

3 Related work

Motivated by the studies mentioned above, considering the temporal and spatial characteristics of wireless cellular traffic and combining with cross-domain data, we simultaneously adopt Conv-LSTM or Conv-GRU and attention mechanism to model the traffic data of network structure. Specifically, the main contributions of our work can be summarized into two folds:

In this paper, we propose an attention-based multi-component spatiotemporal cross-domain neural network model (att-MCSTCNet). The model finely divides historical data and uses the Conv-LSTM or Conv-GRU structure to model the three time characteristics of wireless cellular network traffic, such as proximity, daily periodicity, and weekly periodicity, combined with timestamp feature embedding, multiple cross-domain data fusion, and other modules jointly assist the model to predict traffic. Depending on the internal network structure used, the model can be further divided into att-MCSTCNet (Conv-LSTM) and att-MCSTCNet (Conv-GRU).

We introduce an attention mechanism in the MCSTCNet model. According to the relationship between the three kinds of time feature data (nearest neighbor data, daily cycle data, and weekly cycle data) and the predicted time, the attention mechanism layer will assign different weights to these three types of data, improve their feature extraction ability, suppress interference information, and achieve the effective use of historical wireless cellular traffic data further improving the prediction accuracy of the model. The experiment proves that taking the RMSE as an example, on the Sms dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model increases by about 13.70 ~ 54.96%, and the RMSE of the att-MCSTCNet (Conv-GRU) model increases by about 14.56 ~ 55.82%. On the Call dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model is improved by about 10.50 ~ 28.15%, and the RMSE of the att-MCSTCNet (Conv-GRU) model is improved by about 12.24 ~ 29.89%. On the Internet dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model has increased by approximately 35.85 to 100.23%, and the RMSE of the att-MCSTCNet (Conv-GRU) model has increased by approximately 38.79 to 103.17%.

The rest of this article is structured as follows. The fourth part introduces the dataset adopted in this paper. The fifth part introduces three network structures used in the att-MCSTCNet model. The sixth part constructs the att-MCSTCNet model based on attention mechanism and introduces the training process of the model. In the seventh part, the model is verified and analyzed in three datasets, and the parameters of the model are tested. The last part is the summary of this paper.

4 Dataset

4.1 Introduction of dataset

The dataset used in this paper comes from detailed wireless cellular traffic data in Milan [18], and the cross-domain dataset is base station information (BS), point of interest distribution (POI), and social activities (hereinafter called Social) in the area around Milan. The dataset is divided into 100×100 grid areas covering an area of approximately 552 km² in Milan. The wireless cellular traffic data collected by the dataset is from November 1, 2013 solstice to January 1, 2014, and the unit of data statistics is in the hour. Section 4.4 describes timestamps. Table 1 details the Telecom Italia dataset.

Table 1 Telecom Italia dataset

Dataset	Telecom Italia
City	Milan, Italy
Type of datasets	{Sms, Call, Internet}
Span of time	2013/11/01 ~ 2014/01/01
Timestamp	[week, hour, workday, weekend]
Time interval	1 h
Grid size	(100, 100)
Cross-domain data	{BS, POI, Social}

4.2 Preprocessing of dataset

As shown in Fig. 1, the data preprocessing in this paper goes through the following three steps:

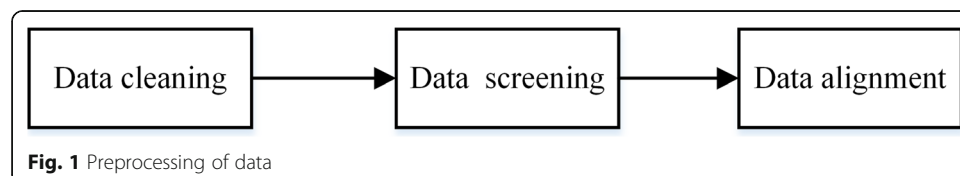
Step 1: Data cleaning. The dataset used in this article is derived from the detailed wireless cellular traffic data of Milan area [19]. The time span is from 0:00 on November 1, 2013 to 23:00 on January 1, 2014. The experiments in this paper extract Sms, Call, and Internet wireless cellular traffic data of three different services. For the missing traffic data of a certain area in a certain period, the average traffic value of the surrounding area or period will be used to fill in.

Step 2: Data screening. Since the recording interval of the original data is 10 min, and most of the recorded data values are 0, this results in sparse data values. The data were divided by hours and min-max normalization was used to process the data to speed up the training process.

Step 3: Data alignment. This article divides the cleaned wireless cellular traffic data, cross-domain data and the city of Milan into a 100×100 grid area one-to-one correspondence. It is convenient to formulate the data below.

4.3 Wireless cellular traffic datasets

The type of wireless traffic data in Milan is represented as k , where $k \in \{\text{Sms, Call, Internet}\}$. Taking the Internet as example, according to the timestamp of the wireless traffic data, the wireless business traffic in Milan can be expressed as a t -dimensional tensor, where T is the total number of time intervals, $t \in \{1, 2, \dots, T\}$, X and Y represent the coordinate points of the city. The urban traffic matrix of the t -th time slot can be expressed as



$$D_t = \begin{bmatrix} d_t^{(1,1)} & d_t^{(1,2)} & \dots & d_t^{(1,Y)} \\ d_t^{(2,1)} & d_t^{(2,2)} & \dots & d_t^{(2,Y)} \\ \vdots & \vdots & \ddots & \vdots \\ d_t^{(X,1)} & d_t^{(X,2)} & \dots & d_t^{(X,Y)} \end{bmatrix} \quad (1)$$

where, t is time point of every data, (X,Y) represents the horizontal and vertical coordinates of each data.

Similarly, formula (1) applies to Sms business and Call business.

Figure 2 shows the temporal dynamics of different kinds of cellular traffic in different areas. The x-axis denotes the time interval index (in hour scale) and y-axis, the number of events of a specific cellular traffic. The black line denotes the most famous university in Milan, Bocconi University, which is the southern suburb of Milan; the red line denotes Navigli, which is the nightlife area of Milan; the blue line denotes the Duomo of Milan, located in the center of Milan. The following can be clearly seen from Fig. 2:

- 1) Data's periodicity. The wireless cellular traffic of different services shows the same periodicity. For instance, in Fig. 2a, b, and c, the traffic of three different business has the same trend in the Bocconi University area. In addition, wireless cellular traffic in different regions also has a similar periodicity. For example, in Fig. 2a, the Sms traffic change tendency of three different areas are similar.
- 2) Differences in regional data. The data volume of wireless cellular traffic in different areas is quite different. Taking the cell Navigli as an example, there is little difference in the data volume of wireless cellular traffic in the region of Navigli, which is the nightlife area of Milan. However, the Bocconi University area is on the outskirts of Milan, so there is relatively little wireless cellular data.
- 3) Differences in business data. The data volume of wireless cellular traffic between different services is also different. For instance, the duration of Internet traffic peaks is shorter than the other two services.

4.4 Timestamp

To make full use of the features of the timestamp (D_{meta}) for auxiliary prediction, four features are extracted from the timestamp. For example, the four characteristic values extracted from 15:00 on December 14, 2013 are as follows: the value of week is 5, the value of hour is 14, the value of working day is 0, and the value of weekend is 1. The four features are processed into a vector m , which is reshaped into a tensor T_s with the same size as the wireless cellular traffic dataset and cross-domain dataset through the fully connected layer. So the vector m goes from dimension 4 to $T \times X \times Y$. The four extracted features are shown in Table 2.

4.5 Cross-domain datasets

The cross-domain (D_{cross}) dataset mainly contains three types of social information (Social), base stations (BS), and points of interest (POI) ($cross \in \{BS, POI, Social\}$) because it

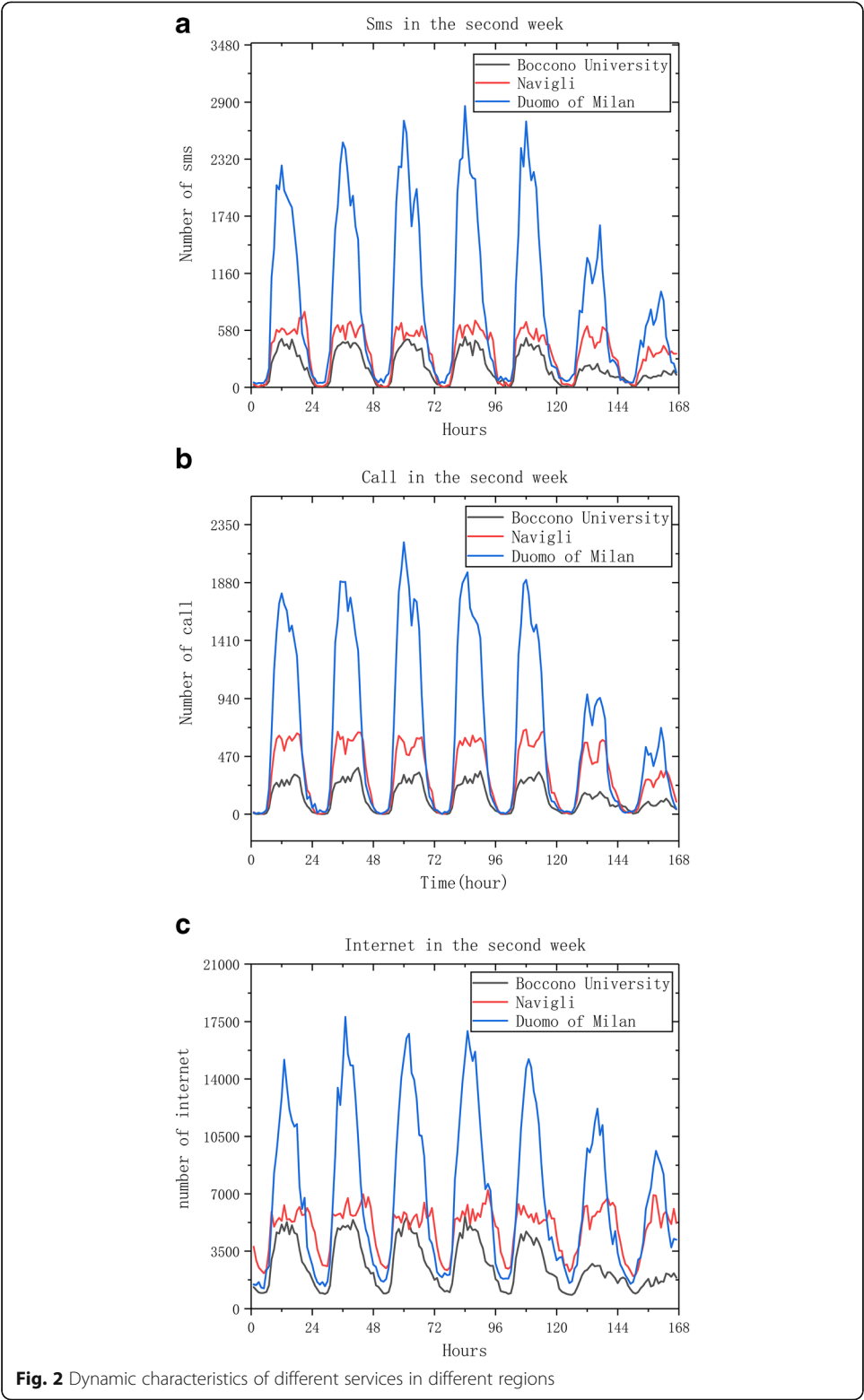


Table 2 Four characteristics of D_{meta}

Feature	Name	Value
1	Week	0, 1, 2, ..., 6
2	Hour	0, 1, ..., 23
3	Weekday	0, 1
4	Weekend	0, 1

can be seen that they are more relevant to wireless service traffic from Fig. 3. Since these three data types have small changes on the time axis, we treat them as static datasets, and then map the data to specific areas based on coordinate information. Referring to Eq. 1, Eq. 2 can be obtained as follows:

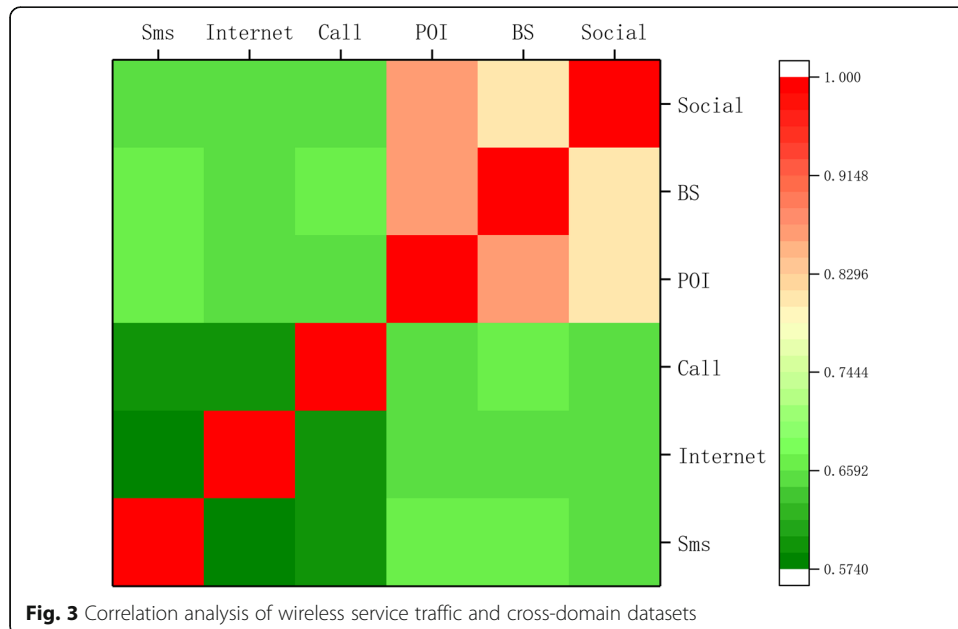
$$D_{cross} = \begin{bmatrix} d_c^{(1,1)} & d_c^{(1,2)} & \dots & d_c^{(1,Y)} \\ d_c^{(2,1)} & d_c^{(2,2)} & \dots & d_c^{(2,Y)} \\ \vdots & \vdots & \ddots & \vdots \\ d_c^{(X,1)} & d_c^{(X,2)} & \dots & d_c^{(X,Y)} \end{bmatrix} \quad (2)$$

where $d_c^{(X,Y)}$ denotes cross-domain data under the x- and y-axes.

In order to analyze the correlation between different business traffic and cross-domain datasets, the Pearson correlation coefficients are calculated as follows:

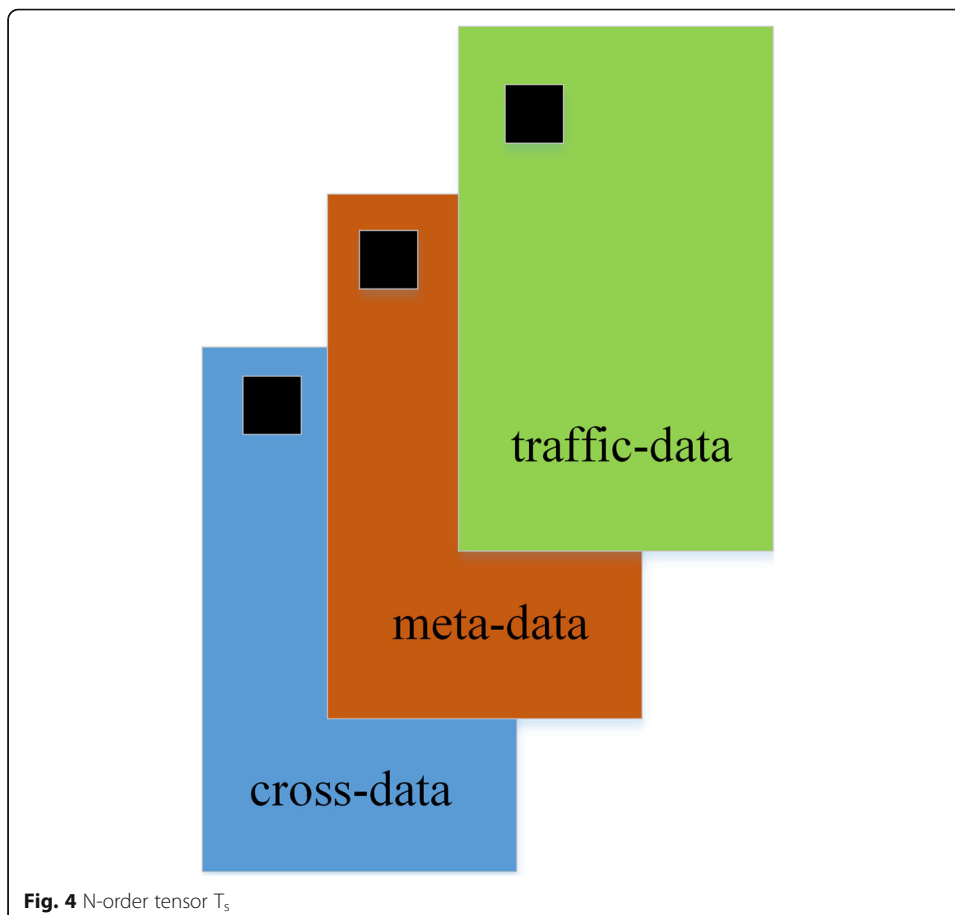
$$\rho = \frac{\text{conv}(d^{(x,y)}, d^{(x',y')})}{\sigma_{d^{(x,y)}} \sigma_{d^{(x',y')}}} \quad (3)$$

where $\text{conv}(\cdot)$ denotes the covariance operator, and σ is the standard deviation.



To further quantify the spatial correlations between cross-domain datasets and cellular traffic, the Pearson correlation coefficients are calculated and shown in Fig. 3. From this Fig. 3, we conclude the following:

- (1) Relevance of data. The correlation between Sms, Call, and Internet is high. If the source domain and target domain data have the same spatial distribution and high similarity, then the transfer learning strategy can be used to transfer the knowledge learned on a certain dataset to the learning of other datasets and tasks, so that the learning of new datasets and tasks does not start from zero, but has a certain a priori basis. Therefore, we can also use the transfer learning strategy across different businesses.
- (2) Similarity of data. The similarity between cross-domain data and wireless business traffic is also relatively high. Therefore, it can be regarded as a constraint on the spatial characteristics of wireless business traffic to make a more accurate prediction of business traffic.
- (3) Relevance of the data. The correlation between POI, BS, and wireless cellular traffic is greater than Social, which shows that the impact of POI and BS on the accurate prediction of business traffic is relatively larger than Social.



Finally, we will get a N-order tensor T_s , it has a dimension of $N \times X \times Y$, which is composed of matrices D_t , D_{meta} and D_{cross} . The data form is shown in Fig. 4. As shown in the black square in Fig. 4, each element in the tensor measures the cellular traffic volume with coordinates (X, Y), timestamp information, and the number of cross-domain data of (X, Y).

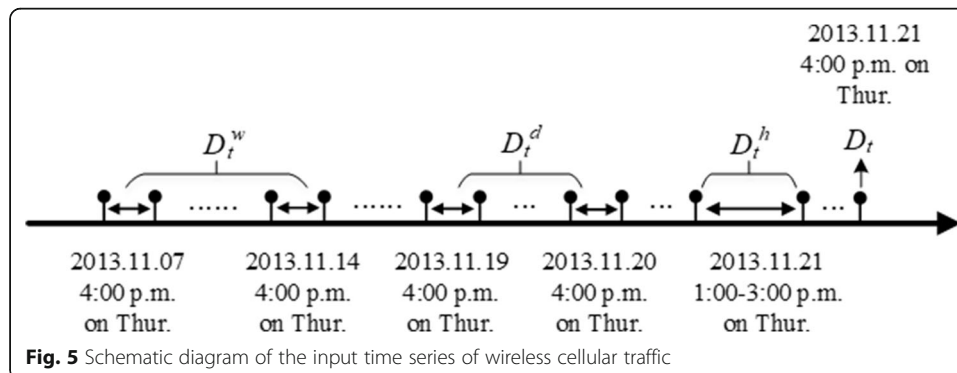
5 Model and network architecture

5.1 Model

Assuming that the predicted cellular traffic time is at 4 pm on Monday, we want to capture the features of the weekly cycle (4 pm Monday) and the nearest neighbor (1 pm Monday to 3 pm Monday) cellular traffic data associated with the target moment, rather than extract the features of the daily cycle (4 pm Sunday) cellular traffic data, because the gap between wireless data traffic on weekdays and weekends is very large, the cellular traffic of daily cycle (4 pm last Sunday) will interfere with the data at the predicted target time. To solve this problem, we introduce the attention mechanism layer and propose an attention-based multi-component spatiotemporal cross-domain neural network model (att-MCST CNet). The model focuses on historical cellular traffic information, which is more critical to the target time, among many input information, reduces the attention to other information, and even filters out irrelevant information. Therefore, the efficiency and accuracy of wireless cellular traffic prediction are improved. The specific structure of the model is shown in Fig. 6. It contains the following 5 parts:

The first part is the modeling of the recent data D_t^h : $D_t^h = [D_{t-\ell_c}, D_{t-(\ell_c-1)}, \dots, D_{t-1}]$, ℓ_c is the time interval in hours. It represents a piece of cellular traffic data sequence segment of a historical time directly adjacent to D_t , as shown in part D_t^h of Fig. 5. Obviously, this type of data will inevitably have a great impact on the current cellular traffic prediction.

The second part is the modeling of the daily periodic data D_t^d : $D_t^d = [D_{t-\ell_m \times m}, D_{t-(\ell_m-1) \times m}, \dots, D_{t-m}]$, $m = 24$. It consists of the same cellular traffic data sequence segment in the previous n days as the predicted target time, as shown in the D_t^d part in Fig. 5. Due to factors such as morning and evening peak cycles, people's



daily work and sleep patterns, cellular traffic data often have a strong similarity at the same time every day. The purpose of the daily cycle module is to model the cycle characteristics of cellular traffic in units of days in wireless cellular data.

The third part is the modeling of the weekly periodic data D_t^w : . It consists of segments of the cellular traffic data sequence with the same properties and the same time in the previous n weeks of the predicted moment and the predicted target week, as shown in the D_t^w part of Fig. 5. Similar to the daily periodicity, wireless cellular traffic data also has obvious weekly cycle characteristics. For example, the wireless cellular traffic pattern at 4 pm on Thursday has similarities to the wireless cellular traffic pattern at 4 pm on Thursday in previous weeks. The weekly periodic module mainly captures the changing rules of wireless cellular traffic with a weekly cycle.

The three parts of the feature input are imported into two layers of the ConvLSTM or Conv-GRU structure, after passing through an attention layer. Then, increase the weight of historical cellular traffic information that is more critical to the target moment, reduce the weight of other information, achieve the purpose of filtering irrelevant information, and further improve the efficiency and accuracy of wireless cellular traffic prediction. In this way, the weight of historical cellular traffic information that is more critical to the target moment can be increased, while the weight of other interference information can be reduced. The irrelevant information is filtered, so the efficiency and accuracy of wireless cell traffic prediction are further improved.

The fourth part is the modeling of display time features. The input is a matrix with timestamps as features. The feature matrix D_{meta} is put into two layers of fully connected neural network for training.

The fifth part is cross-domain data modeling. The input is the cross-domain dataset D_{cross} . The cross-domain dataset we used mainly includes BS, Social, and POI in this region, where D_{cross} is a collection of three cross-domain data. The fused cross-domain dataset D_{cross} is imported into two layers of convolutional neural network to process such data and assist the prediction of wireless cellular traffic.

The sixth part is the feature fusion layer. The above six preliminary feature outputs are spliced into a new tensor according to the specified dimensions, and the tensor is input to a densely connected convolutional network (DenseNet). The network contains a total of L layers, and each layer implements a composite function transformation. The operations in the feature learning of cross-domain data are the same, including batch regularization (BN), activation function (Relu), and convolution operation (Conv).

The Frobenius norm is calculated for the final output:

$$\ell(\theta) = \arg \min_{\theta} \|\hat{D}_t - D_t\|_F \quad (3)$$

where θ is the set of all parameters of STC-N, \hat{D}_t represents the predicted value of traffic data, D_t represents the true value of traffic data.

The following is the algorithm of the att-MCSTCNet model (Fig. 6) training process. First, build a training example from the original sequence (lines 1–5), and then train with Adam through backward propagation (lines 6–11).

Algorithm : training process of att-MCSTCNet model

Input : historical data: $\{D_0, \dots, D_{t-1}\}$

 timestamp data: D_{meta}

 Cross-domain data: D_{cross}

 recent, periodic sequence length : ℓ_c, ℓ_m

 recent length: $c=1$; cycle length : $m=24$;

Output : the trained att-MCSTCNet model M

// Construct training examples

 $T \leftarrow \emptyset$
while all available time intervals t ($1 \leq t \leq n$)

 $D_t^h = [D_{t-\ell_c}, D_{t-(\ell_c-1)}, \dots, D_{t-1}]$
 $D_t^d = [D_{t-\ell_m \times m}, D_{t-(\ell_m-1) \times m}, \dots, D_{t-m}]$
 $D_t^w = [D_{t-\ell_m \times m \times 7}, D_{t-(\ell_m-1) \times m \times 7}, \dots, D_{t-m \times 7}]$ put training examples

 $(\{D_t^h, D_t^d, D_t^w, D_{meta}, D_{cross}\}, D_t)$ into T

 // D_t is the true value at time t .

end

// training model

 initialize parameter ϑ
repeat

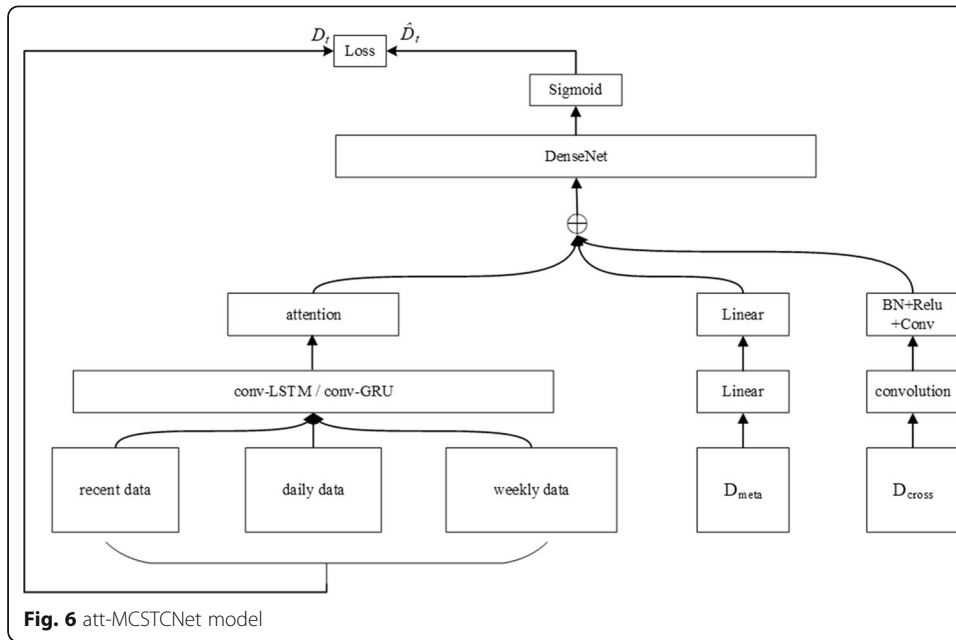
 randomly select a batch of instances T_b from T

 find the optimal parameter ϑ through T_b and Adam optimization algorithm

until Stop training after meeting the conditions

5.2 Conv-LSTM structure

The Conv-LSTM structure is shown in Fig. 7. Each cell of the Conv-LSTM network layer has a storage unit C for storing state information. Cell C deletes and adds data information through three gates, which are input gates i_g and f_g and output gate o_g , respectively. Among them, the input gate i_g selectively stores the required data information, and the forget gate f_g also selectively “forgets” the redundant information. The final hidden state is controlled by the output gate o_g and determines the importance of the output data information. The key operation of Conv-LSTM is as formula (4):

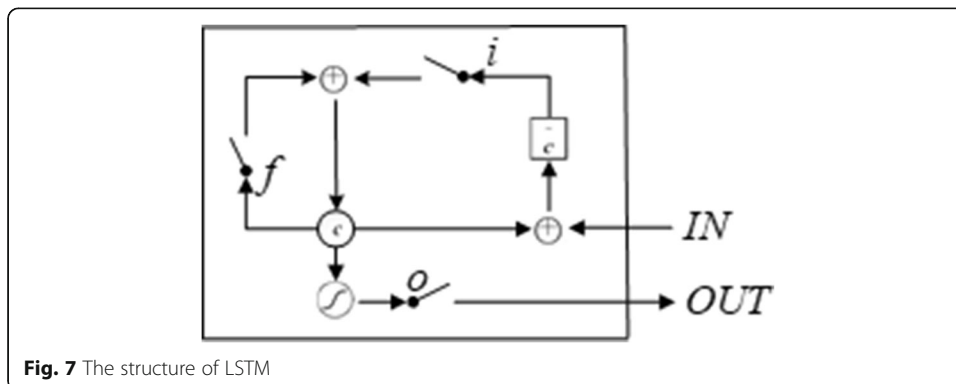


$$\begin{aligned}
 i_g^r &= \sigma(W_{di} * D_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \\
 f_g^r &= \sigma(W_{df} * D_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \\
 c_t &= f_g^r \odot C_{t-1} + i_g^r \odot \tanh(W_{dc} * D_t + W_{hc} * H_{t-1} + b_c) \\
 o_g^r &= \sigma(W_{do} * D_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \\
 H_t &= o_g^r \odot \tanh(c_t)
 \end{aligned} \quad (4)$$

where $\sigma(\cdot)$ is the activation function; $*$ is the convolution operation; \odot is the Hadamard product operation; $W(\cdot)$ is the training weight; $b(\cdot)$ is the training bias; $\tanh(\cdot)$ is the hyperbolic tangent function; and i_g^r, f_g^r, c_t, o , and H_t are all a three-dimensional tensor. The output is $o_t, o_t \in \mathbb{R}^{H \times X \times Y}$. H is the number of feature maps.

5.3 Conv-GRU structure

The Gate Recurrent Unit (GRU) is a type of recurrent neural network [20] and is also a variant of LSTM. Compared with LSTM, GRU can achieve the same effect, and is easier to train, which can greatly improve training efficiency, so we use the Conv-GRU in



the model. As shown in Fig. 8, r_t controls a reset gate. The reset gate is used to control the degree of ignoring the state information at the previous moment. z_t is the update gate, and it is used to control the degree to which the state information of the previous moment is brought into the current state. Compared with the three gates of LSTM, the parameters are reduced, and few parameters save resources and converge faster.

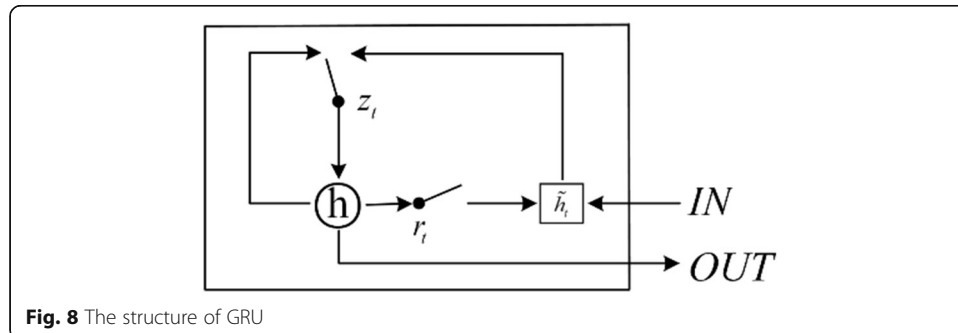
Formula (5) includes the calculation process of resetting gate r_t and updating gate z_t . Among them, \tilde{h}_t mainly contains the currently input x_t data, and the \tilde{h}_t is added to the current hidden state in a targeted manner, which is equivalent to memorizing the state at the current time. $(1-z_t) \odot h_{t-1}$ indicates selective “forgetting” of the originally hidden state, and $1-z_t$ can be regarded as a forgetting gate which can forget unimportant information in h_{t-1} dimensions. $z_t \odot \tilde{h}_t$ means to selectively memorize containing the current node information, which can be regarded as selecting some information in the \tilde{h}_t dimension. Therefore, a gated z_t can perform both forgetting and selective memory, which is also the advantage of the GRU structure.

$$\begin{aligned} z_t &= \sigma(W_z * [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r * [h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W * [r_t \odot h_{t-1}, x_t]) \\ h_t &= \tanh((1-z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t) \end{aligned} \quad (5)$$

where h_{t-1} is the hidden state of the previous node, which contains information about the previous node. x_t is the current input.

5.4 Structure of attention mechanism

Attention mechanism is a solution proposed by imitating human attention, that is, a mechanism that aligns internal experience with external sensations to increase the fineness of observation in some areas. For example, when looking at a picture, the human eye will quickly scan the global image to obtain the target area that needs to be focused. This is the focus of attention. By devoting more attention to this area, we can obtain more detailed information about the targets we need to pay attention to and suppress other useless information. For the wireless cellular traffic time series in this paper, for the output y at a certain time, the attention layer assigns different attention to the hidden layer h corresponding to the input x , that is, different weights are given to features of different importance levels, and associate it with the output to achieve the purpose of information filtering. The structure development of the attention model is shown in Fig. 9. The attention model is roughly divided into three layers: input layer, hidden



layer, and attention layer. We take the data of three sections of D_t^h , D_t^d , and D_t^w as the input part of the network structure of attention mechanism.

The hidden layer state (h_1, h_2, \dots, h_t) is obtained by Conv-GRU.

- (1) The influence of each current input position on the i position can be calculated, as shown in Formula (6).
- (2) Soft-max normalization is performed on e_t to obtain the attention weight distribution, as shown in Formula (7).
- (3) Vector c_t can be obtained by weighted sum of α_t , as shown in Formula (8)

$$e_t = v_a^T \text{relu}(W_a s_{i-1} + U_a h_t) \quad (6)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)} \quad (7)$$

$$c_t = \sum_{i=1}^T \alpha_i h_i \quad (8)$$

Where, V_a , W_a and U_a are the weight values of the attention network, $\text{relu}(\cdot)$ is the activation function, T is the total number of time intervals, and S is the current input state, $\exp(\cdot)$ is an exponential function based on the natural constant e .

6 Results and discussion

6.1 Assessment method

In this paper, root mean square error (RMSE), mean absolute error (MAE), determination coefficient, and three evaluation indexes are adopted. The formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T \sum_{x=1}^X \sum_{y=1}^Y (\hat{d}_t^{(x,y)} - d_t^{(x,y)})^2}{T \times X \times Y}} \quad (9)$$

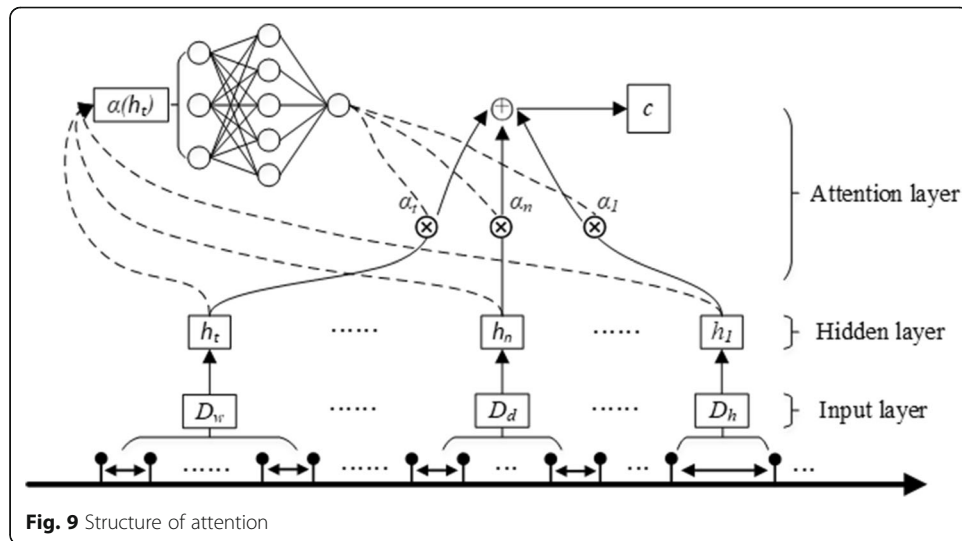


Fig. 9 Structure of attention

$$MAE = \frac{\sum_{t=1}^T \sum_{x=1}^X \sum_{y=1}^Y |\hat{d}_t^{(x,y)} - d_t^{(x,y)}|}{T \times X \times Y} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{t=1}^T \sum_{x=1}^X \sum_{y=1}^Y (\hat{d}_t^{(x,y)} - d_t^{(x,y)})^2}{\sum_{t=1}^T \sum_{x=1}^X \sum_{y=1}^Y (\bar{d}_t^{(x,y)} - d_t^{(x,y)})^2} \quad (11)$$

where T is the time point, X and Y are the coordinate information of the time point respectively, represents the cellular traffic predicted value at time T with coordinates of (X, Y) , and represents the cellular traffic actual value of at time T with coordinates of (X, Y) .

RMSE is used to measure the deviation between the predicted value of the model and the true value. MAE can better reflect the actual situation of the error of the predicted value of the model. For both of them, the smaller they are, the better the model fitting effect will be; otherwise, the worse the effect will be. The value range of R^2 is $[0, 1]$; the closer its value is to 1, the more independent variables can explain the variance of the dependent variable, the better the model's effect; otherwise, the worse the model's effect.

6.2 Comparative experiment of multiple models on different datasets

In order to illustrate the advantages of the att-MCSTCNet (Conv-LSTM) model and att-MCSTCNet (Conv-GRU) model, this paper selects several classical wireless cellular traffic prediction methods for performance comparison.

The benchmark methods are shallow machine learning methods and deep learning methods. Among them, shallow machine learning methods include LR [7] and SVR [8], while deep learning methods include LSTM [9], STDenseNet [19], STNet [18], STMNet [18], and STCNet [18]. On different datasets, RMSE, MAE, R^2 of different models are shown in Tables 3, 4, and 5. In the Tables 3, 4, and 5, F_o , F_r , F_d , F_w , F_m , F_s , and F_c respectively represent temporal characteristics, recently characteristics, daily cycle characteristics, weekly cycle characteristics, timestamp characteristics, spatial characteristics, and three cross-domain data characteristics. "√" in the table indicates that the model uses this characteristic.

As can be seen from Tables 3, 4 and 5, the two models we proposed in this paper have better performance in RMSE, MAE, and R^2 than other models in three different business datasets. Taking the RMSE as an example, for Sms dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model has increased by about 13.70 ~ 54.96%, and the RMSE of the att-MCSTCNet (Conv-GRU) model has increased by about 14.56 ~ 55.82%. For Call dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model is improved by about 10.50 ~ 28.15%, and the RMSE of the att-MCSTCNet (Conv-GRU) model is improved by about 12.24 ~ 29.89%. For Internet dataset, the RMSE of the att-MCSTCNet (Conv-LSTM) model has increased by approximately 35.85 to 100.23%, and the RMSE of the att-MCSTCNet (Conv-GRU) model has increased by approximately 38.79 to 103.17%. And for three datasets, the att-MCSTCNet model with Conv-GRU structure has better prediction performance than the att-MCSTCNet model with Conv-LSTM structure. The RMSE increased by about 0.85 ~ 2.94%. The reasons for the best performance of the att-MCSTCNet (Conv-LSTM) and att-MCSTCNet (Conv-GRU) models are the following two points: firstly, the spatiotemporal correlation of

wireless cellular traffic data was captured by Conv-LSTM and Conv-GRU structures; secondly, attention mechanism structure was added in the att-MCSTCNet model, useful information of wireless cellular network traffic was seized, and useless information was suppressed, so the training performance of this model was further improved.

To compare the superiority of the att-MCSTCNet model more intuitively, the experimental results are shown in Figs. 10, 11, and 12; as can be clearly seen from Figs. 10, 11 and 12, the proposed att-MCSTCNet (Conv-LSTM) model and att-MCSTCNet (The Conv-GRU) model has better prediction performance than other models, and att-MCSTCNet (Conv-GRU) has better prediction performance than att-MCSTCNet (Conv-LSTM).

6.3 Comparative experiment of different structures in the att-MCSTCNet model

In order to further analyze the difference between the Conv-GRU structure and Conv-LSTM in the att-MCSTCNet model, we conducted comparative experiments on the number of training parameters, training time, and changes in model training loss of different structures.

Table 6 shows the amount of training parameters for the two structures under the att-MCSTCNet model. It can be clearly seen that the Conv-GRU structure has fewer training parameters than the Conv-LSTM structure, which shows that the Conv-GRU structure is better than the Conv-LSTM structure. The amount of training is less, and the training is faster.

To fully explain the advantages of the Conv-GRU structure, we analyze the training time and the training_loss and valid_loss of the model training. Train loss is the loss on the training data, which measures the fitting ability of the model on the training set. Valid loss is the loss on the validation set, which measures the fitting ability on unseen data, which can also be said to be the generalization ability. Taking the Sms dataset as an example, the experimental results are shown in Figs. 13, 14, and 15.

As can be seen from Fig. 13, the iteration time of Conv-GRU structure is less than that of Conv-LSTM structure, so in the case of more iterations, the Conv-GRU structure saves a lot of time than the Conv-LSTM structure.

Table 3 Performance comparison of various models and other models on the Sms dataset

Model	Multiple factors							RMSE	MAE	R^2
	F_0	F_r	F_d	F_w	F_m	F_s	F_c			
LR	√							94.1253	35.6864	0.5765
SVR	√							72.6544	30.5419	0.7571
LSTM	√							69.0916	40.1357	0.7715
DenseNet	√							62.7586	27.512	0.8057
STMNet	√					√		58.6207	29.6203	0.8191
STNet	√				√	√		57.8182	28.132	0.8276
STCNet	√				√	√	√	52.8665	28.1034	0.8559
att-MCSTCNet (Conv-LSTM)	√	√	√	√	√	√	√	39.1589	22.6542	0.9274
att-MCSTCNet (Conv-GRU)	√	√	√	√	√	√	√	38.3087	21.8829	0.9303

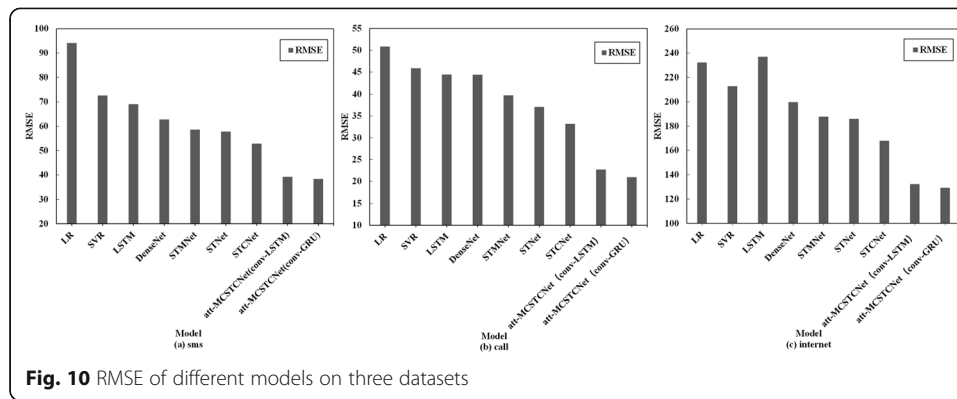
Table 4 Performance comparison of various models and other models on the call dataset

Model	Multiple factors							RMSE	MAE	R^2
	F_o	F_r	F_d	F_w	F_m	F_s	F_c			
LR	√							50.8571	25.903	0.8102
SVR	√							45.8714	21.9142	0.8438
LSTM	√							44.4286	21.2136	0.8521
DenseNet	√							44.3864	19.144	0.8602
STMNet	√					√		39.7341	17.8913	0.8896
STNet	√				√	√		37.0857	16.8159	0.9129
STCNet	√				√	√	√	33.207	16.2393	0.9204
att-MCSTCNet (Conv-LSTM)	√	√	√	√	√	√	√	22.7051	14.2622	0.9628
att-MCSTCNet (Conv-GRU)	√	√	√	√	√	√	√	20.9656	13.1964	0.9683

Figs. 14 and 15 are a comparison of the train_loss and valid_loss of three different structures. They are respectively the convolution LSTM structure (Conv-LSTM), convolution LSTM structure based on attention mechanism (att_Conv-LSTM), and convolution GRU structure based on attention mechanism (att_Conv-GRU). Experimental results show that compared with train_loss and valid_loss of other two structures, the att_Conv-GRU structure converges faster, and the loss value after stabilization is smaller; this shows that the train_loss or valid_loss function has obtained a local optimal solution, thus the fitting effect of the att_Conv-GRU structure model is better. The main reason is that the Conv-GRU structure has one less gating unit than the Conv-LSTM structure, which means that the GRU parameter calculation is less, so the Conv-GRU structure training requires fewer parameters than Conv-LSTM, the iteration time is less than conv-LSTM, and Conv-GRU has a faster convergence rate. In particular, train_loss and valid_loss of Conv-GRU with the addition of attention decrease faster, and the loss value after stabilization is lower, thus the attention mechanism can further improve the fitting effect of the model.

Table 5 Performance comparison of various models and other models on the internet dataset

Model	Multiple factors							RMSE	MAE	R^2
	F_o	F_r	F_d	F_w	F_m	F_s	F_c			
LR	√							232.3171	156.1111	0.9085
SVR	√							212.9878	130.6113	0.9245
LSTM	√							237.0122	141.6667	0.907
DenseNet	√							199.8171	122.5536	0.9362
STMNet	√					√		187.8049	116.6111	0.9448
STNet	√				√	√		186.0366	105.6131	0.9489
STCNet	√				√	√	√	167.9386	96.0912	0.9532
att-MCSTCNet (Conv-LSTM)	√	√	√	√	√	√	√	132.0883	85.244	0.9711
att-MCSTCNet (Conv-GRU)	√	√	√	√	√	√	√	129.1509	84.8507	0.9723



6.4 Model parameter optimization experiment

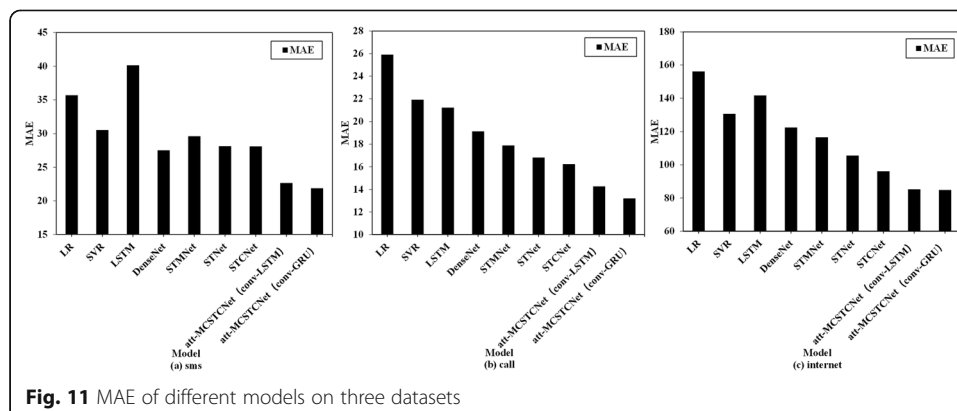
6.4.1 Model depth selection

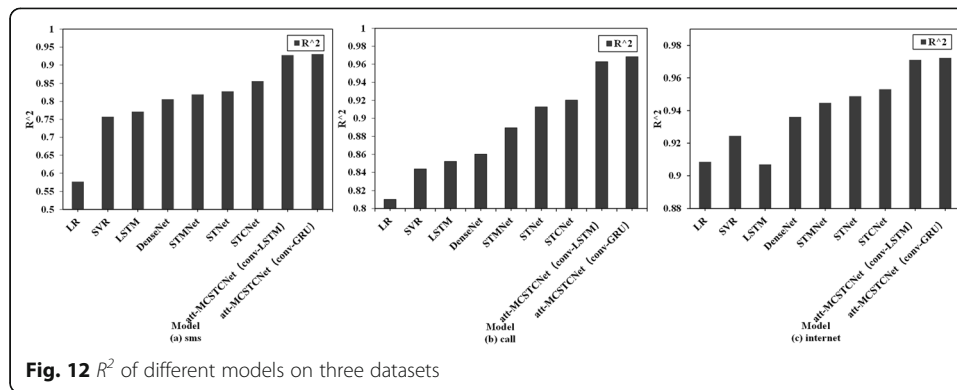
Different network layer depths under the same model will have different effects on the prediction performance of the model. A suitable network layer depth can maximize the prediction effect of the model, so this experiment analyzes five experiments based on the att-MCSTCNet model. As a result, models of different depths are shown in Table 7.

The experimental results are shown in Fig. 16. The prediction performance of the model with different depths is good under the three datasets, but as the network depth increases, the predictive performance of the 3-layer network depth model is the best. Moreover, when the network depth of the model increases to 4 and 5, the RMSE of the model will increase significantly, because the increase of network depth will cause the model parameters to increase greatly, which is not conducive to model training. Therefore, after comprehensive consideration, the att-MCSTCNet model selects the most appropriate three-layer network depth for training.

6.4.2 Setting of batch_size

A suitable batch_size can find a relative balance between stability and model calculation overhead. Because the GPU can play better performance on the nth power of 2 batch_size, we set the batch_size of the model to 32, 64, and 128 and tested





the effect of this parameter on the three datasets. The experimental results are shown in Fig. 17. It can be seen from Fig. 17 that under the same training times, the model with batch_size of 32 shows better performance than other values, so the att-MCSTCNet model selects batch_size as 32 for training.

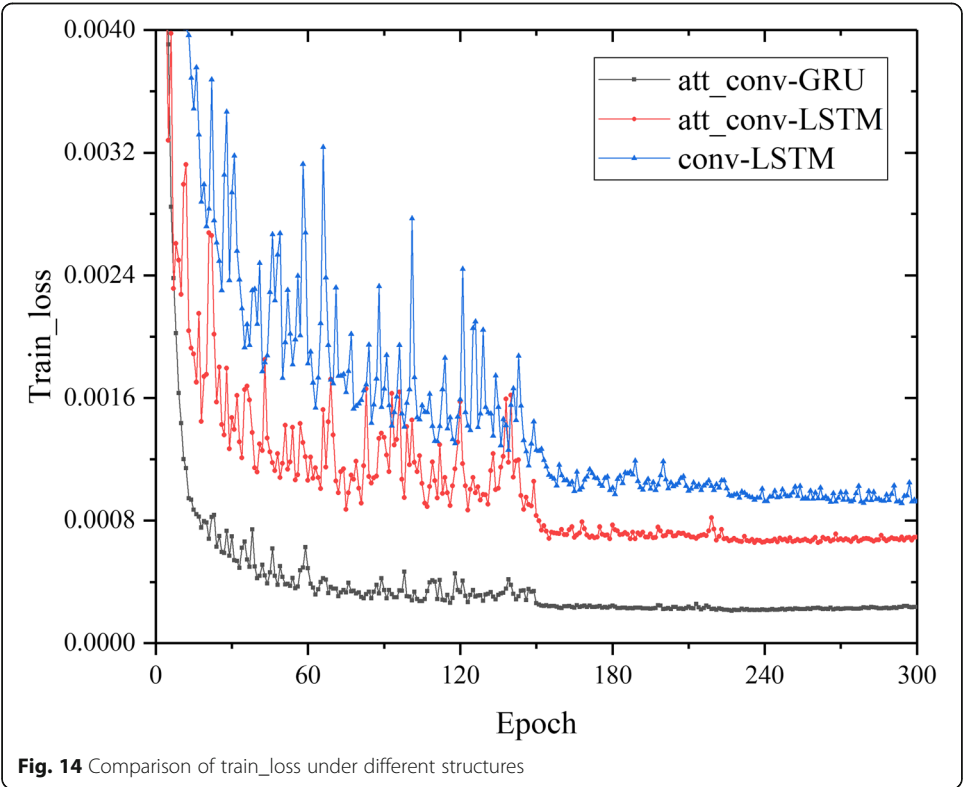
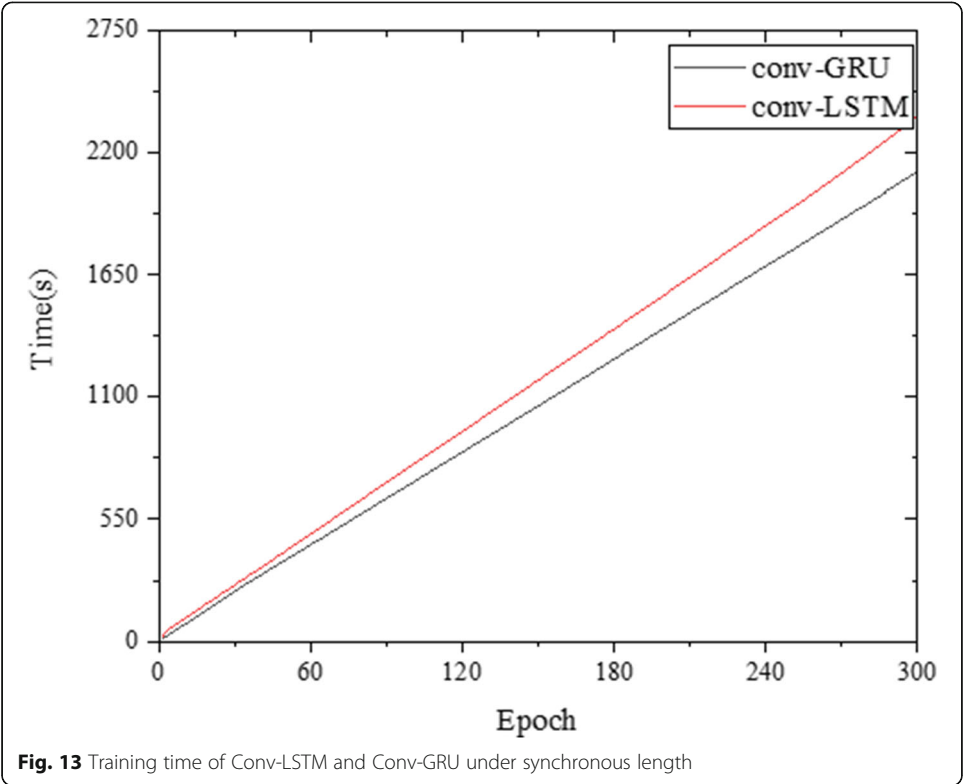
In addition, through repeated experimental verifications, the att-MCSTCNet is optimized using a stochastic gradient-based optimization technique, the model is trained for 300 epochs. An adaptive learning rate (lr) is adopted in this work, whose initial value is set to be 0.01 and will be divided by 10 and 100 at 150 epochs and 225 epochs. In the convolutional layer, the number of feature maps is 16, the size of the convolution kernel is 3×3 , and Relu is used as the activation function. The feature map of the output layer is 1, and the size of the convolution kernel is 1×1 . During training, the first seven weeks of the entire dataset are used as the training set, and the last week's data is used as the test set. Both the training set and the test dataset are constructed using a sliding window method with a window size of $P = 3$. The summary of model training parameters is shown in Table 8.

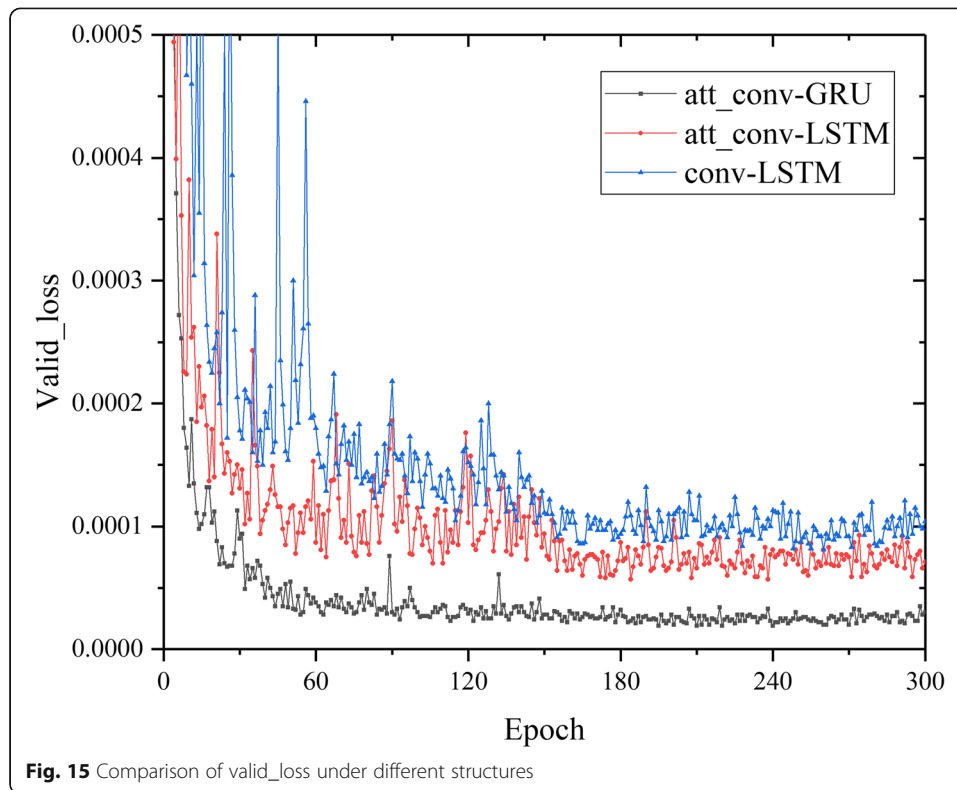
7 Conclusions

We propose an attention-based multi-component spatiotemporal cross-domain neural network model (att-MCSTCNet) to predict wireless cellular network traffic. The model uses the conv-LSTM or conv-GRU structure to model three temporal properties of wireless cellular network traffic (i.e., recent, daily periodic, and weekly periodic dependencies) combined with timestamp feature embedding, multiple

Table 6 Training parameters of different structures

Dataset	Structure of the model	Amount of training parameters
Sms	att-Conv-LSTM	312948
	att- Conv-GRU	298284
Call	att- Conv-LSTM	312948
	att- Conv-GRU	298284
Internet	att- Conv-LSTM	312948
	att- Conv-GRU	298284



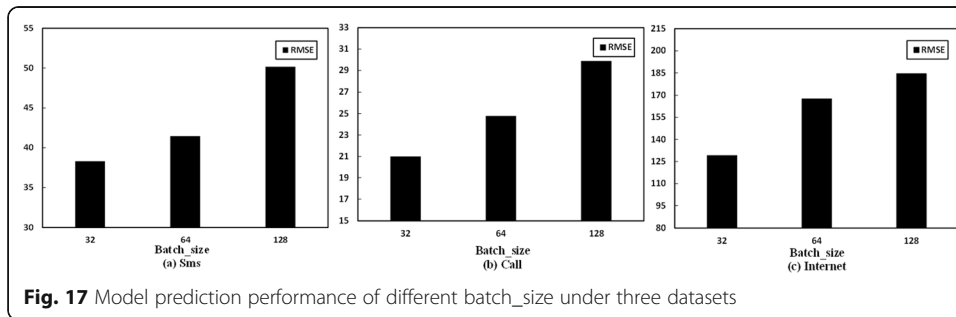
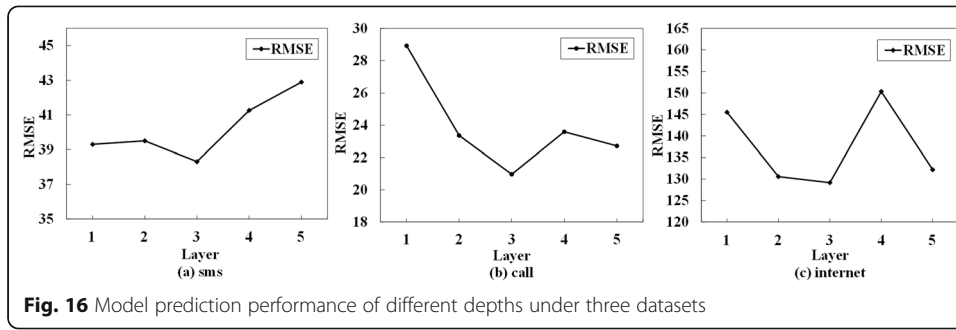


cross-domain data fusion, and other modules to assist the model in traffic prediction. Experiments prove that the proposed model is better than the existing model, and the att-MCSTCNet model with the conv-GRU structure has better prediction effect than the att-MCSTCNet model with the conv-LSTM structure. The model training time is reduced, the workload is greatly reduced, and the prediction performance of the model is further improved.

Due to the complex adoption framework of the model proposed in this paper, the overall training time of the model is still long. The next step will consider adopting a simpler and more efficient model architecture in order to improve the training accuracy of the model while reducing the training time.

Table 7 Model structure of different depths

Depth	Structure of the model
1	input→Conv-GRU layer→attention layer→DenseNet layer→output
2	input→Conv-GRU layer→Conv-GRU layer→attention layer→DenseNet layer→output
3	input→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→attention layer→DenseNet layer→output
4	input→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→attention layer→DenseNet layer→output
5	input→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→Conv-GRU layer→attention layer→DenseNet layer→output

**Table 8** Hyperparameter setting of att-MCSTCNET model

Hyperparameters	Value
Loss	mse
Optimizer	Adam
Batch_size	32
Learning rate	0.01
Epoch	300

Abbreviations

5G/B5G: 5th-Generation/beyond 5th-generation; att-MCSTCNet: Attention based multi-component spatiotemporal cross-domain neural network; ASTGCN: Attention-based space time graph convolutional network; SVR: Support vector regression; CNN: Convolutional Neural Networks; LSTM: Long short-term memory; GRU: Gated Recurrent Unit Recurrent Neural Networks; DenseNet: Densely connected convolutional network; BN: Batch regularization; Relu: Rectified linear unit; Conv: Convolution operation; BS: Base station; POI: Point of interest; Social: Social activity; RMSE: Root mean squared error; MAE: Mean absolute error; R^2 : R squared; LR: Logistic regression; STCNet: Spatial-temporal cross-domain convolutional neural network; STNet: Spatial-temporal convolutional neural network; STMNet: Spatial-temporal meta data convolutional neural network

Acknowledgements

Not applicable.

Authors' contributions

QTZ, QS and GC conceived and designed the experiments. QTZ and QS performed the experiments. GC and HD analyzed the data. QS, GC, and QTZ wrote the paper. The authors have contributed to this research work and read and approved the final manuscript.

Authors' information

Qingtian Zeng received the B.S. degree and the M.S. degree in computer science from Shandong University of Science and Technology, Taian, China, in 1998 and 2001 respectively, and the Ph.D. degree in computer software and theory from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a Professor at Shandong University of Science and Technology, Qingdao, China. His research interests are in the areas of Petri nets, process mining, and knowledge management.

Qiang Sun received the B.S. degree in Communication Engineering from Liaocheng University, Liaocheng, China, in 2017. He is currently pursuing the master's degree in the College of Electronic and Information Engineering of Shandong University of Science and Technology. His current research interests include Internet of Things and Cellular traffic forecasting.

Geng Chen received the B.S. degree in electronic information engineering and the M.S. degree in communication and information system from Shandong University of Science and Technology, Qingdao, China, in 2007 and 2010, respectively, and the Ph.D. degree in information and communications engineering from Southeast University, Nanjing, China, in 2015. He is currently an Associate Professor at the College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China. His current research interests are in the areas of heterogeneous networks, ubiquitous networks, and software-defined mobile networks, with emphasis on wireless resource management and optimization algorithms, and precoding algorithms in large-scale MIMO.

Hua Duan received the B.S. and M.S. degrees in applied mathematics from the Shandong University of Science and Technology, Tai'an, China, in 1999 and 2002, and the Ph.D. degree in applied mathematics from Shanghai Jiao Tong University, in 2008. She is currently a Professor at the Shandong University of Science and Technology. Her research interests include process mining and machine learning.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 61701284, the Innovative Research Foundation of Qingdao under Grant No. 19-6-2-1-cg, the Application Research Project for Postdoctoral Researchers of Qingdao, the Sci. & Tech. Development Fund of Shandong Province of China under Grant Nos. 2016ZDJS02A11 and ZR2017MF027, the Humanities and Social Science Research Project of the Ministry of Education under Grant No. 18YJAZH017, the Taishan Scholar Climbing Program of Shandong Province under Grant No. ts2090936, SDUST Research Fund under Grant No. 2015TDJH102, and the Science and Technology Support Plan of Youth Innovation Team of Shandong higher School under Grant No. 2019KJN024.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China. ²College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao 266590, China.

Received: 13 April 2021 Accepted: 28 June 2021

Published online: 26 July 2021

References

1. J. Huang, J. Tan, Y. Liang, Wireless big data: transforming heterogeneous networks to smart networks. *J. Commun. Inf. Netw.* **2**(1), 19–32 (2017)
2. X. Wu, X. Zhu, G. Wu, Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **26**(1), 97–107 (2014)
3. Y. Lecun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**(7553), 436–444 (2015)
4. M. Jordan, T. Mitchell, Machine learning: trends, perspectives, and prospects. *Science* **349**(6245), 255–260 (2015)
5. J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C. Soong, J.C. Zhang, What will 5G be. *IEEE J. Selected Areas Commun.* **32**(6), 1065–1082 (2014)
6. Y. Shu, M. Yu, J. Yang, Wireless traffic modeling and prediction using seasonal ARIMA models. *IEEE Int. Conf. Commun.* **88**(10), 3992–3999 (2005)
7. H. Sun, H. Liu, H. Xiao, Use of local linear regression model for short-term traffic forecasting. *Trans. Res. Record J. Trans. Res. Board* **1836**, 143–150 (2003)
8. N. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey. *IEEE Comput. Intell. Mag.* **4**(2), 24–38 (2009)
9. J. Wang, J. Tang, Z. Xu, in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Spatiotemporal Modeling and Prediction in Cellular Networks: a Big Data Enabled Deep Learning Approach (2017), pp. 1–9
10. C. Jing, K. Qian, X. Wang, in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Passenger Demand Prediction with Cellular Footprints (2018), pp. 1–9
11. C. Qiu, Y. Zhang, Z. Feng, Spatio-temporal wireless traffic prediction with recurrent neural network. *IEEE Wireless Comm. Lett.* **7**(4), 554–557 (2018)
12. Z. Hu, H. Hao, X. Zhu, Research on crowd flows prediction model for 5G demand. *J. Commun.* **40**(2), 1–10 (2019)
13. C. Zhang, H. Zhang, J. Qiao, Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data. *IEEE J. Selected Areas Commun.* **37**(6), 1389–1401 (2019)
14. J. Qu, M. Ye, X. Qu, Airport delay prediction model based on regional residual and LSTM network. *J. Commun.* **40**(4), 149–159 (2019)
15. V. Mnih, N. Heess, A. Graves, *Recurrent Models of Visual Attention*. *Advances in Neural Information Processing Systems* (2014), pp. 2204–2212
16. Y. Cheng, S. Shen, Z. He, in *Proceeding UCAFL6 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, Agreement-Based Joint Training for Bidirectional Attention-Based Neural Machine Translation (2016), pp. 2761–2767
17. N. Feng, S. Guo, C. Song, Multi-component spatial-temporal graph convolution networks for traffic flow forecasting. *J. Softw.* **30**(3), 759–769 (2019)
18. C. Zhang, H. Zhang, J. Qiao, Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data. *IEEE J. Selected Areas Commun.* **37**(6), 759–769 (2019)
19. C. Zhang, D. Yuan, Citywide cellular traffic prediction based on densely connected convolutional neural networks. *IEEE Commun. Lett.* **22**(8), 1656–1659 (2018)
20. T. Nussbaum, J. Cui, G. Ramabhadran, Acoustic modeling using bidirectional gated recurrent convolutional units. *Interspeech* **2016**, 390–394 (2016)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)