Traffic Prediction of Wireless Cellular Networks Based on Deep Transfer Learning and Cross-Domain Data

QINGTIAN ZENG¹, QIANG SUN¹, GENG CHEN¹, HUA DUAN², CHAO LI¹, AND GE SONG¹

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

Corresponding author: Geng Chen (gengchen@sdust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701284, Grant 31671588, and Grant 61472229; in part by the Innovative Research Foundation of Qingdao under Grant 19-6-2-1-eg; in part by the Science and Technology Development Fund of Shandong Province of China under Grant 2016ZDJ1S02A11 and Grant ZR2017MF027; in part by the Humanities and Social Science Research Project of the Ministry of Education under Grant 18YJAZH017; in part by the Taishan Scholar Climbing Program of Shandong Province under Grant ts20190936; in part by the Shandong University of Science and Technology Research Fund under Grant 2015TDH102; and in part by the Science and Technology Support Plan of Youth Innovation Team of Shandong Higher School under Grant 2019KJN024.

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 show high nonlinearity and complex patterns. Most existing wireless cellular traffic prediction methods, lacking abilities of modeling the dynamic spatial-temporal correlations of wireless cellular traffic data, thus cannot yield satisfactory prediction results. To improve the accuracy of 5G/B5G cellular network traffic prediction, more cross-domain data was considered, a cross-service and regional fusion transfer learning strategy (Fusion-transfer) based on the spatial-temporal cross-domain neural network model (STC-N) was proposed. Multiple cross-domain datasets were integrated. The training accuracy of the target service domain based on the data characteristics of its source service domain according to the similarity between services and the similarity between different region was improved, so the predictive performance of the model was enhanced. The experimental results show that the prediction accuracy of the traffic prediction model is significantly improved after the integration of multiple cross-domain datasets, the RMSE performance of SMS, Call and Internet service can be improved about 8.39%, 13.76% and 5.7% respectively. In addition, compared with the existing transfer strategy, the RMSE of the three services can be improved about 2.48%~13.19%.

INDEX TERMS 5G/B5G, cellular network, fusion-transfer, cross-domain data, traffic prediction.

I. INTRODUCTION

As the age of 5G/B5G comes, the number of mobile devices and the Internet of things is showing an exponential growth worldwide, and people’s demand for wireless mobile data is growing rapidly. According to Cisco’s “VNI Forecast Highlights Tool”, global service mobile data traffic will increase six-fold from 2017 to 2022, and its annual growth rate of 42% [1]. How to allocate and optimize existing cellular network resources scientifically and rationally, improve resource utilization, and reduce the energy consumption of cellular base stations are the problem that the communications industry needs further thinking and solving. The mobile Internet is closely related to people’s lives, and the use of mobile data traffic also reflects people’s lives to a certain extent, which contains a lot of information [2]. 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 caused by multi-source cross-domain data. The causes that affect wireless service traffic in a certain area are diverse. When making wireless cellular traffic prediction, not only should the hidden regular patterns of wireless service 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. Furthermore, 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 accuracy of wireless cell traffic prediction is restricted by many problems, so it is urgent to find new methods to solve them. In recent years, the era of big data has spawned the in-depth application of artificial intelligence technology. Artificial intelligence can solve the problems of learning changes, problem classification, and predicting the future through interaction with the environment [3]. Big data-driven network intelligence plays an important role in reducing wireless network overhead and improving user experience quality [4]. Deep learning is a feature learning algorithm based on data samples in machine learning [5], which is a significant way to achieve artificial intelligence [6]. With the accumulation of massive wireless cellular traffic data and the maturity of machine learning (deep learning) technologies, wireless cellular network prediction methods based on deep learning have been realized. How to use deep learning methods to accurately predict regional traffic changes has become a hot topic in the communication field. Literature [7] proposed a wireless traffic prediction framework based on deep confidence networks. Experiments show that the framework can predict the long-term dependence of wireless traffic, but this solution only considers the time correlation of wireless traffic. In order to further capture the spatial correlation between different regions and combine the results obtained by the long-term and short-term memory (LSTM) network [8] to capture the temporal correlation of deep learning models, but the model only considers time characteristics. The literature [9] proved that the LSTM model performs better than the traditional statistical ARIMA model. The prediction of cellular data traffic is more accurate, this work provides the basic structure for future wireless cell traffic prediction models. Literature [10] developed a deep model of traffic prediction combined with automatic encoder and long-short-term memory network LSTM. It took the lead in using deep learning technology to carry out space-time modeling and solve prediction problems. The experimental results prove that space-time correlation modeling improves the importance of traffic prediction accuracy. However, the model can only be trained separately for each component. Literature [11] designed a spatiotemporal modeling module composed of conv-LSTM and conv-Net two neural network architectures to simultaneously capture the dependence of traffic on the time and space dimensions. The spatiotemporal modeling module can train the structure in an end-to-end manner, and large-scale prediction tasks can be carried out. Nonetheless, the literature [10], [11] does not make wireless cellular traffic prediction based on the actual scenario, and only considers the traffic prediction of a single region. The literature [12] further applies the dependence of the time and space dimensions to the real scene of the city scale, and makes the proposed neural network architecture more robust and work better, but this model can only get good prediction results in a certain time or a certain region. Literature [13] adopts LSTM to capture time correlation and adopts the idea of multi-task learning to fully integrate business traffic in different regions, but it does not consider the impact of other cross-domain data. In the field of crowd flow prediction, to consider the impact of cross-domain data more comprehensively in practical application scenarios, Literature [14] used LSTM to model the spatial and temporal dependencies of different scales, and merged multiple cross-domain data (weather, air quality, holiday information, etc.) to further improve the accuracy of crowd flows prediction. Reference [15] added cross-domain data (point-of-interest distribution data) and wireless service traffic data, and selected auxiliary traffic data based on distance as inputs to the traffic prediction model. Among them, the wireless service traffic data is used as the main feature sequence, and the point-of-interest data is generated by discrete embedded learning networks to generate external auxiliary features, and then the two are feature-fused and gradually predicted by the variant S2S model of the RNN network [16]. This research proves the effectiveness of multi-source data in wireless service traffic prediction. However, the cross-domain data considered in [14]–[16] is relatively single, while [17] considers the impact of multiple cross-domain data on wireless cellular traffic prediction, analyzes the correlation between them and three different traffic services, and compares the forecast Base station (BS) data, point of interest distribution (POI) data and social information (Social) data in the area are added to the prediction model as auxiliary traffic prediction. The Conv-LSTM and CNN modules used in the model can well capture the impact of space, time and various external factors, in order to capture regional differences and similarities from space and time domains. The literature [17] only analyzes the single auxiliary factors related to the prediction of cellular traffic. There is no systematic and comprehensive study on the performance
of models incorporating different numbers of cross-domain datasets. In the field of airport delay prediction, reference [18] systematically and comprehensively added different numbers of cross-domain datasets (weather factors, flight operation data, etc.) to the airport delay prediction model. The results prove that the accuracy of airport delay prediction is higher than that of adding only one cross-domain dataset after fusing the weather information of the airport and related flight information. This work provides ideas for us to investigate the impact of different cross-domain datasets on wireless cellular traffic prediction.

Transfer learning is an optimization method for a model. It can apply the trained model of the first task to the second related task, which can make the model’s progress in modeling the second task faster and improve its prediction performance. Transfer learning is introduced between multiple services at [17], the similarity of different types of cellular services in time and space is fully utilized, and the prediction performance is further improved. Thus, transfer learning can be used between different services to improve the prediction accuracy, but the learned model does not continue to transfer to different regions to achieve fusion transfer and further improve the prediction accuracy on the original basis. On the basis of the above work, we do a further study on transfer learning on wireless cellular traffic prediction.

This paper proposes a deep learning method for regional traffic prediction based on multiple cross-domain big data, and analyzes in detail the gains to the model after adding different numbers of cross-domain big datasets. The results fully illustrate the necessity of adding different numbers of cross-domain datasets to the prediction of cellular traffic. Using the characteristics of transfer learning, a new fusion transfer learning strategy is proposed to further improve the prediction performance on the original model [17].

The main contributions of this article include:

Considering all kinds of influencing factors of wireless cell traffic prediction is not only helpful to the prediction results, but also more satisfying to the needs of the reality. Unfortunately, there is not much work on the analysis and processing of large datasets added to cross-domains, in order to fully analyze the impact of the types and number of large datasets on the prediction of traffic, this paper uses the spatio-temporal cross-domain neural network (STC-N) as a benchmark model, different types of cross-domain big datasets are taken as research objects, and the impact of different amounts of cross-domain big data on the accuracy of traffic prediction is discussed separately. In addition, the reasons for the different influences are analyzed. The experimental results verify that for one more cross-domain dataset considered, the prediction performance will be enhanced to varying degrees.

Since the traffic changes between different services and between different regions have a high similarity, the idea of transfer learning can be introduced into the regional wireless traffic prediction. We fuse transfer learning between different services and successive transfer learning between classes to get a new transfer learning strategy, namely: cross-service and regional fusion transfer strategy (Fusion-transfer for short), and apply the new transfer learning strategy to training in STC-N model. Finally, the trained model is used in the wireless service traffic dataset. The learning of the new datasets and tasks do not start from scratch, but has a certain a priori basis. The source domain and the target domain can share the model parameters, namely the model trained in the source domain through a large amount of data is applied to the target domain for prediction. The experimental results show that the newly added Fusion-transfer has better performance than the No-transfer and Part-transfer models, which fully shows that the new fusion migration strategy is better than other migration strategies. In this way, we apply a trained model of a service in one region to the other regions of that service and the other regions of other services. This transfer learning method greatly reduces the training data, computing power and generalization ability required to build a deep learning model.

The rest of this article is structured as follows. The second part introduces the dataset and preprocessing. The third part introduces the architecture of the neural network model and analyzes the network layer by layer. The fourth part shows the experimental results of the training model and evaluates it in detail. The last part is the summary of this paper.

II. DATASET
A. DATASET PREPROCESSING

The dataset used in this article is derived from the detailed wireless cellular traffic data of Milan area [19]. As shown in Fig.1, the data preprocessing in this paper goes through the following three steps.

Step 1: Data cleaning. The time span of wireless cellular traffic data in Milan is from 0:00 on November 1, 2013 to 23:00 on January 1, 2014. The wireless cellular traffic data of Sms, Call and Internet are extracted from the experiment. 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 minutes, and most of the recorded data values are 0, this results in sparse data values. Data are divided by hours and normalized by min-max to speed up the training process.

Step 3: Data alignment. To facilitate the formulation of its data below, the cleaned wireless cellular traffic data, cross-domain data and Milan city are divided into 100 × 100 grid area for one-to-one correspondence.
B. WIRELESS CELLULAR TRAFFIC DATASETS

The wireless traffic in a certain region can be expressed as $D_{k,t}$. It is a spatiotemporal series composed of a large number of data points, as follows.

$$D_{k,t} = \begin{bmatrix} d_{k,t}^{(1,1)} & d_{k,t}^{(1,2)} & \cdots & d_{k,t}^{(1,Y)} \\ d_{k,t}^{(2,1)} & d_{k,t}^{(2,2)} & \cdots & d_{k,t}^{(2,Y)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k,t}^{(X,1)} & d_{k,t}^{(X,2)} & \cdots & d_{k,t}^{(X,Y)} \end{bmatrix}$$ (1)

Similarly, Eq. (1) applies to Sms service and Call service, $d_{k,t}^{(X,Y)}$ represents the amount of data in the wireless cellular traffic at time $t$ in the coordinates $(X, Y)$ for $k$ service, where $k \in \{\text{Sms, Call, Internet}\}$.

Fig.2 shows the temporal dynamics of different kinds of wireless cellular traffic in different areas. The x-axis denotes the time interval index (in hour scale) and y-axis represents the value of wireless cellular traffic for a particular service. 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. It can be clearly seen from Fig.2 that:

1) Data’s periodicity. The wireless cellular traffic of different services shows the same periodicity. For instance, in fig.2(a), (b), (c), the traffic of three different service 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.2(a), the SMS traffic changed 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. For example, in fig.2(a), the Bocconi University area is a suburb of Milan, so the wireless cellular data volume is the least compared to other areas. And Navigli is Milan’s nightlife area, the amount of wireless cellular data in this area is not much different within a week, and the amount of wireless cellular data is also not high. Milan’s Duomo is the central city of Milan, with the highest amount of wireless cellular traffic data, and the amount of wireless cellular data on weekdays in the area is very different from weekends.

3) Differences in service 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.

C. TIME STAMP

To make full use of the features of the timestamp ($D_{meta}$) for auxiliary prediction, four features are extracted from the timestamp, and the four features are processed into a vector $m$, which is then processed by the full connection layer into a tensor $T$ consistent in size with the wireless cellular traffic dataset and cross-domain dataset. The four extracted features are shown in Table 1.

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

where $d_c^{(X,Y)}$ is the amount of c cross-domain data in the coordinates $(X, Y)$, $c \in \{\text{BS, POI, Social}\}$.
In order to analyze the correlation between different service traffic and cross-domain datasets, the Pearson correlation coefficients $\rho$ are calculated as follows:

$$\rho = \frac{\text{cov}(d^{(X,Y)}, d^{(X',Y')})}{\sigma_{d^{(X,Y)}} \sigma_{d^{(X',Y')}}} \quad (3)$$

where $\text{cov}(\cdot)$ denotes the covariance operator and $\sigma$ is the standard deviation, $d^{(X,Y)}$ is wireless cellular traffic in $(X, Y)$ coordinates, $d^{(X',Y')}$ is wireless cellular traffic in $(X', Y')$ coordinates.

To further quantify the spatial correlations between cross-domain datasets and different cellular traffic, the Pearson correlation coefficients are calculated and shown in Fig.3.

Finally, we will get a multi-dimensional tensor $T$, which is composed of matrices $D_t$, $D_{meta}$ and $D_{cross}$. The data form is shown in Fig.3. As shown in the black square in Fig.4, takes Sms as an example, it represents the value of the cross-domain data of the Sms service, time stamp information and wireless cellular traffic value at $(x, y)$ coordinates.

III. MODEL AND FUSION TRANSFER STRATEGY

A. MODEL

Fig.5 presents the overall framework of the spatial-temporal cross-domain neural network model (STC-N) used in this paper. It consists of four components as follows:

1. Spatio-temporal correlation modeling
   Input: $(D_{t-3}, D_{t-2}, D_{t-1})$. three historical moments directly adjacent to the predicting period. It imports a two-layer conv-LSTM network to model spatio-temporal dependencies and data sequence information.

2. Timestamp modeling
   Input: $D_{meta}$. It imports a two-layer fully connected neural network for embedded learning.

3. Cross-domain data modeling
   Input: $D_{cross}$. It is a collection of one to three cross-domain data. It imports a two-layer convolutional neural network to process these data.

4. Feature fusion layer
   Input: A new tensor, which is composed of the above four feature outputs stitched according to the specified dimension. It imports a densely connected convolutional network (DenseNet), which contains a total of L layers, and each layer implements a composite function transformation. This composite function is the same as the operation in cross-domain data feature learning, which includes batch regularization (BN), activation function (Relu) and convolution operation (Conv).

The conv-LSTM layer realizes the protection and control of information through three self-parameterized controlling “gates”, i.e., input gate $i_g$, forget gate $f_g$ and output gate $o_g$. $i_g$ selectively stores the required data information, $f_g$ also selectively “forgets” the redundant information, and the final hidden state is controlled by the $o_g$ and determines the important data information required for output. The key operations of conv-LSTM are as follows:

$$i_g = \sigma(W_{di} * D_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i)$$

$$f_g = \sigma(W_{df} * D_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f)$$

$$c_t = f_g \odot C_{t-1} + i_g \odot \tanh(W_{dc} * D_t + W_{hc} * H_{t-1} + b_c)$$

$$o_g = \sigma(W_{do} * D_t + W_{ho} * H_{t-1} + b_o)$$

$$H_t = o_g \odot \tanh(c_t)$$
$$o_g^t = \sigma(W_{do} \ast D_t + W_{ho} \ast H_{t-1} + W_{co} \odot C_t + b_o)$$

$$H_t = o_g^t \odot \tanh(c_t)$$

where $\sigma(\cdot)$ denotes the activation function, $\ast$ denotes the convolution operation, $\odot$ is the Hadamard product, $W(\cdot)$ and $b(\cdot)$ are the weights and biases to be trained, respectively. And $\tanh(\cdot)$ refers to the hyperbolic tangent function. Note that the $H^t_{1}, T^t_{1}, c_t, o_g^t, H^t_t$ in the conv-LSTM unit are all three-dimensional tensors. The output of conv-LSTM network is denoted as $o_t \in \mathbb{R}^{H \times X \times Y}$, where $H$ is the number of feature maps.

The preliminary characteristics of $D_{meta}$ are denoted as $O_{meta}$, and the processing is as follows:

$$o_{meta} = \sigma(w_{meta}^2 \sigma(w_{meta}^1m + b_{meta}^1) + b_{meta}^2)$$

where $w_{meta}^1, b_{meta}^1$ are learnable parameters, $o_{meta} \in \mathbb{R}^{H \times X \times Y}$.

The preliminary characteristics of $D_{cross}$ are denoted as $O_{cross}$, and the processing is as follows:

$$o_{cross} = f(W_{cross} \ast D_{cross})$$

$$D_{cross} = D_{BS} \oplus D_{POI} \oplus D_{Social}$$

where $\oplus$ is the concatenation operation, $W_{cross}$ is the weights that will be trained. $f(\cdot)$ represents a composite function that contains the Batch Normalization (BN), rectified linear units (Relu) and convolution operation (Conv).

The objective function of STC-N is to minimize the Frobenius norm of error matrix between prediction and truth over all areas. It can be expressed as

$$\ell(\hat{\theta}) = \arg \min_\theta \left\| \hat{D_t} - D_t \right\|$$

where $\hat{\theta}$ is the parameter of STC-N, which can be obtained by training through optimization techniques, $\hat{D_t}$ is the predicted value of the wireless cell traffic, $D_t$ is the true value of the wireless cell traffic.

The STC-N is optimized by stochastic gradient optimization technique, and trained for 300 epochs with batch size 32. An adaptive learning rate is used to speed up the training. The initial learning rate is set to 0.01, and then the learning rate is attenuated by 10 times at 150 epochs and 225 epochs. In the convolutional layer, the number of feature maps is 16, the size of the convolution kernel is $3 \times 3$, 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 $1 \times 1$. 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 $P = 3$.

### B. FUSION TRANSFER STRATEGY

Traditional machine learning trains different datasets separately, and different model parameters are learned. Multiple learning systems of models are relatively independent. And transfer learning is to learn knowledge on a certain dataset, and transfer this knowledge to the learning of other datasets and tasks, so that the learning of the new datasets and tasks do not start from scratch, but has a certain a priori basis. As shown in Fig. 6, assuming that the source domain (dataset 1) and the target domain (dataset 2) have some common features, and the source domain data and the target domain data in this space have the same data distribution and high similarity, the source domain and the target domain can share the model parameters, namely The model trained in the source domain through a large amount of data is applied to the target domain for prediction. Therefore, we first train a basic network model on the basic dataset and task, and then readjust or transfer the learned features to the second target network model to train another target dataset and tasks.

Fig. 2 and Fig. 3 prove the similarity between different services on multiple datasets. The wireless traffic in different regions also has certain similarities, such as periodicity. However, the wireless service traffic data in different regions have both differences and similarities. To capture the differences in different regions, K-means clustering method is used to group similar regions together. The results of clustering, all regions are divided into three cluster [18]. On the basis of clustering results, the traffic pattern knowledge learned from one cluster could be transferred to other clusters. According to the similarity between services and the similarity of clustered regions, a cross-service and regional fusion transfer (Fusion-transfer) learning strategy was proposed, which fusion different kinds of cellular traffic transfer learning strategy and successive inter-cluster transfer learning strategy. Fusion-transfer strategy is illustrated in Fig. 7.

Take Sms service and Call service as examples, a model $M_{S1}$ can be obtained after training STC-N using Call service data from the source domain S. Then $M_{S1}$ can be transferred to cluster 1 of Sms service target domain by
means of parameter initialization. The parameters of STC-N are continually trained using data from target domain and the model $M_2$ can be learned. Repeat the above steps until the parameters are transferred to cluster 3 of the Sms service target domain. Then, the model $M_T$ can be learned and the prediction is carried out on the test data of the target domain. Finally, the results can be obtained. Through this fusion-transfer strategy, it can not only make full use of the similarity of wireless service traffic of different services, but also make use of the similarity of wireless service traffic of different regions, so the prediction performance of the model can be improved. Similarly, due to the similarity between different services, this strategy is also applicable to Call and Internet services.

### IV. EXPERIMENT

Four metrics are adopted in this work for the sake of a comprehensive evaluation of different prediction algorithms. They are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared ($R^2$) and Explain the variance score (Explained variance score, hereinafter Called Variance).

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T} \sum_{X=1}^{X} \sum_{Y=1}^{Y} (\hat{d}_t^{(X,Y)} - d_t^{(X,Y)})^2}{T \times X \times Y}}$$  \hspace{1cm} (10)

$$\text{MAE} = \frac{\sum_{t=1}^{T} \sum_{X=1}^{X} \sum_{Y=1}^{Y} |\hat{d}_t^{(X,Y)} - d_t^{(X,Y)}|}{T \times X \times Y}$$  \hspace{1cm} (11)

$$R^2 = 1 - \frac{\sum_{t=1}^{T} \sum_{X=1}^{X} \sum_{Y=1}^{Y} (\hat{d}_t^{(X,Y)} - d_t^{(X,Y)})^2}{\sum_{t=1}^{T} \sum_{X=1}^{X} \sum_{Y=1}^{Y} (\hat{d}_t^{(X,Y)} - d_t^{(X,Y)})^2}$$  \hspace{1cm} (12)

$$\text{Explained variance score} = 1 - \frac{\text{variance}_{\hat{d}_t^{(X,Y)} - d_t^{(X,Y)}}}{\text{variance}_{d_t^{(X,Y)}}}$$  \hspace{1cm} (13)

where $T$ is the time point, $X$ and $Y$ are the coordinate information of the time point, $\hat{d}_t^{(X,Y)}$ represents the values predicted by the wireless cell traffic by a model at the time $t$ and the coordinate is $(X, Y)$, $d_t^{(X,Y)}$ represents the values of ground truth of the wireless cell traffic at the time $t$ and the coordinate is $(X, Y)$, $\bar{d}_t^{(X,Y)}$ represents the average value of wireless cell traffic at time $t$, coordinate is $(X, Y)$.

RMSE is used to measure the deviation between the predicted value of the model and the true value. MAE could better reflect the actual situation of the model prediction value error. The value range of $R^2$ is $[0, 1]$, the closer its value is to 1, the better the model’s effect will be; otherwise, the worse the model’s effect will be. Variance means a larger value implies a better fitting to the data, thus, a better performance. For RMSE and MAE, the smaller the value, the better the performance. On the contrary, for $R^2$ and Variance metric, a larger value implies a better fitting to the data, thus, a better performance.

### A. PERFORMANCE OF ADDING MULTIPLE CROSS-DOMAIN DATASETS

To more intuitively prove the improvement effect of multiple cross-domain datasets on the prediction performance of wireless service traffic, we conducted a comparative experiment based on the STC-N model with No-cross, One-cross, Two-cross, and Three-cross as independent variables. The results of evaluation metrics on three different kinds of cellular traffic are plotted in Fig.7. No-cross in the x-axis of Fig.6 represents that we do not use the cross-domain datasets when training the model. One-cross means that model only uses one cross-domain dataset ($D_{social}$). Two-cross indicates that two cross-domain datasets are considered when training prediction model ($D_{social} \oplus D_{BS}$). Three-cross denotes that model adds the three cross-domain datasets ($D_{social} \oplus D_{BS} \oplus D_{POI}$). The y-axis represents the performance index.

Fig.8, Fig.9 and Fig.10 has shown the benefits of introducing different numbers of cross-domain datasets into three cellular traffic prediction. From these figures we can draw the following two conclusions:

1) Different Numbers of cross-domain datasets have different effects on performance. For example, for Sms dataset, the RMSE of Three-cross is 4.84 lower than No-corss, the RMSE performance improvement was about 8.39%. At the same time, for Call dataset, the RMSE of Three-cross is 5.29 lower than No-corss, the RMSE performance improvement was about 13.76%. In addition, for Internet dataset, the RMSE of Three-cross is 10.16 lower than No-corss, the RMSE performance improvement was about 5.7%. From the perspective of overall results, the more
comprehensive the types of cross-domain datasets considered, the better the training performance of the model performance, which is consistent with the actual situation.

2) Taking RMSE and MAE as examples, on the Sms service, BS data can bring more performance gains than others cross-domain data. RMSE decreased by about 2.52, MAE decreased by about 1.21. On the Call service, the performance gain of the model is the largest after adding the cross-domain dataset BS, RMSE and MAE were reduced by about 2.12 and 1.29, respectively. This shows that BS cross-domain datasets have a better performance on Sms and Call services than the other two cross-domain datasets Social and POI, which may be due to the greater dependence of the Sms and Call services on BS data. But on the Internet service, the performance gain of the model is the largest after adding the POI dataset, RMSE and MAE were reduced by about 4.77 and 2.85, respectively. This indicates that POI has a greater impact on the prediction of Internet service than the other two cross-domain datasets, which may be due to the high demand for network in public places.

**B. COMPARISON OF DIFFERENT TRANSFER LEARNING STRATEGIES**

To illustrate the performance of the fusion-transfer learning strategy, we consider three types of services (Sms, Call, Internet) respectively, and compare the experiments using the fusion migration strategy (Fusion-transfer) and the model without the migration strategy (No-transfer). The valid loss value obtained by model training is shown in Fig. 11.
As described in Fig. 11, the red dotted line indicates that the model without adding the transfer strategy (No-transfer) obtained the valid loss value, the black dotted line represents the value of valid loss obtained by the model after adding fusion-transfer strategy (Fusion-transfer), and the x-axis represents epochs of model training. y-axis represents valid loss of model training. From the overall trend of the curve in Figure 11, the red and black lines both drop rapidly first and then stabilize. The red lines stabilize after about 180 epochs, the black lines stabilize after about 150 epochs. In the three figures, the black lines always drop faster than the red lines, and the stable value of valid loss is lower. This shows that on Sms, Call and Internet services, the model based on the fusion-transfer learning strategy is trained faster, and the fusion-transfer learning strategy has better prediction performance than the model without transfer learning strategy. This is mainly because transfer learning does not need to train the model from scratch, and can get shared parameters from other models. To comprehensively prove the gain of fusion-transfer strategy on the predicted performance of the model, we conducted comparative experiments on models using fusion-transfer strategy, no-transfer strategy and part-transfer strategy. The experiment will show the gain of the Fusion-transfer strategy to the training model (STC-N). From the four performance indicators of RMSE, MAE, $R^2$ and Variance. The performance evaluation index data measured in the experiment is shown in Table 2.

| Dataset | Type of transfer | RMSE   | MAE    | $R^2$  | Variance |
|---------|-----------------|--------|--------|--------|----------|
| Sms     | No-transfer     | 60.598 | 32.7456| 0.8253 | 0.8404   |
|         | Part-transfer   | 57.4472| 30.5285| 0.843  | 0.8495   |
|         | Fusion-transfer | 52.7486| 28.3276| 0.8677 | 0.8744   |
|         | No-transfer     | 38.9517| 19.8315| 0.8904 | 0.9016   |
| Call    | Part-transfer   | 36.1898| 17.3746| 0.9054 | 0.9108   |
|         | Fusion-transfer | 33.8146| 16.5374| 0.9174 | 0.922    |
|         | No-transfer     | 176.1658| 103.9554| 0.9485 | 0.9486   |
| Internet| Part-transfer   | 174.4239| 102.1173| 0.9495 | 0.9515   |
|         | Fusion-transfer | 171.7885| 99.6702| 0.951  | 0.9523   |

The results are shown in Fig. 12. After using the Fusion-transfer strategy in Fig. 12, compared with the model without the transfer strategy (No-transfer), the RMSE of the Sms service is reduced by about 7.85, MAE is reduced by about 4.42, and the $R^2$ is increased by about 0.04, performance improvements for RMSE, MAE, and $R^2$ were 12.95%, 13.49% and 5.14%, respectively. The RMSE of the Call service decreased by about 5.14, MAE decreased by about 3.29, $R^2$ is increased by about 0.03, performance improvements for RMSE, MAE, and $R^2$ were 13.19%, 17.11% and 2.62%, respectively. The RMSE and MAE of Internet services decreased by approximately 4.38 and 4.29, respectively, and $R^2$ increased by approximately 0.003, performance improvements for RMSE, MAE, and $R^2$ were 2.48%, 4.12% and 0.39%, respectively. Compared with the three different strategies, it shows that for the same training model (STC-N), the performance of fusion-transfer strategy is the best, the performance of part-transfer strategy is the second, and the performance of no-transfer strategy is the worst. The experimental results further prove that the fusion-Transfer strategy can further improve the performance of all aspects of STC-N model, and it is also applicable to three kinds of services (Sms, Call and Internet). The performance of STC-N model training is better than that of STC-N model training without transfer strategy and part-transfer strategy.

V. CONCLUSION AND FUTURE WORK
In this paper, the influences of the types and quantities of cross-domain big datasets on the traffic prediction accuracy...
are fully analyzed. Taking the spatio-temporal cross-domain neural network (STC-N) as the benchmark model, different types of cross-domain big datasets are taken as the research objects to discuss the influences of different numbers of cross-domain big data on the traffic prediction accuracy. The experimental results verify that the prediction performance of the neural network will be enhanced to different degrees when considering one more cross-domain dataset mentioned.

In addition, a fusion-transfer strategy is proposed which fuses between different traffic transfer learning strategy and inter-cluster transfer learning strategy. The high similarity of traffic changes between different services and between different regions is fully exploited. This transfer learning strategy is to learn knowledge on source domain dataset, and transfer this knowledge to the learning of target domain datasets and tasks, so that the learning of the new datasets and tasks do not start from scratch, but has a certain a priori basis. The source domain and the target domain can share the model parameters, namely the model trained in the source domain through a large amount of data is applied to the target domain for prediction. Therefore, we first train a basic network model on the basic dataset and task, and then readjust or transfer the learned features to the second target network model to train another target dataset and tasks. This transfer learning method greatly reduces the training data, computing power, and generalization ability required to build a deep learning model. Experimental results show that the performance of the Fusion-transfer strategy is better than that of the model without no-transfer strategy and that using part-transfer strategy, which fully indicates that the new Fusion-transfer strategy is better than other transfer strategies.

Actually, the structure of the model is complex and the training time of the model is longer. In the future, we will take using a simpler and more efficient model architecture to improve the training accuracy of the model while reducing the training time.

REFERENCES

[1] (2020). Cisco Annual Internet Report (2018-2023) White Paper. Accessed: Mar. 9, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] D. Nabolisi, M. Fiore, S. Ribot, and R. Stanica, “Large-scale mobile traffic analysis: A survey,” IEEE Commun. Surveys Tuts., vol. 18, no. 1, pp. 124–161, 1st Quart., 2016.

[3] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, “Intelligent 5G: When cellular networks meet artificial intelligence,” IEEE Wireless Commun., vol. 24, no. 5, pp. 175–183, Oct. 2017.

[4] P. Agypang, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” IEEE Commun. Mag., vol. 52, no. 11, pp. 65–75, Nov. 2014.

[5] C. M Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). New York, NY, USA: Springer-Verlag, Inc. 2006.

[6] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[7] L. Nie, D. Jiang, S. Yu, and H. Song, “Network traffic prediction based on deep belief network in wireless mesh backbone networks,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Mar. 2017, pp. 1–5.

[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[9] S. Jaffry, “Cellular traffic prediction with recurrent neural network,” 2020, arXiv:2003.02807. [Online]. Available: http://arxiv.org/abs/2003.02807

[10] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, “Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach,” in Proc. IEEE Conf. Comput. Commun., May, 2017, pp. 1–9.

[11] C. Zhang and P. Patras, “Long-term mobile traffic forecasting using deep spatio-temporal neural networks,” in Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.. Jun. 2018, pp. 231–240.

[12] X. Chen, Y. Jin, S. Qiang, W. Hu, and K. Jiang, “Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale,” in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp. 3585–3591.

[13] C. Qiu, Y. Zhang, Z. Feng, P. Zhang, and S. Cui, “Spatio-temporal wireless traffic prediction with recurrent neural network,” IEEE Wireless Commun. Lett., vol. 7, no. 4, pp. 554–557, Aug. 2018.

[14] Z. Hu, H. Yuan, X. Zhu, and W. Ni, “Research on crowd flows prediction model for 5G demand,” J. Commun. China, vol. 40, no. 2, pp. 1–10, 2019.

[15] J. Feng, X. Chen, R. Gao, M. Zeng, and Y. Li, “Deep TP: An end-to-end neural network for mobile cellular traffic prediction,” in Proc. IEEE Netw., 2018, pp. 108–115.

[16] I. Sutskever, O. Vinyals, and V. Le, “Sequence to sequence learning with neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2014, p. 3215.

[17] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, “Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data,” IEEE J. Sel. Areas Commun., vol. 37, no. 6, pp. 1389–1401, Jun. 2019.

[18] J. Qu, M. Ye, and X. Qu, “Airport delay prediction model based on regional residual and LSTM network,” J. Commun. China, vol. 40, no. 4, pp. 149–159, 2019.

[19] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, “A multi-source dataset of urban life in the city of milan and the province of trentino,” Sci. Data, vol. 2, no. 1, Dec. 2015, Art. no. 150055.

[20] Dandelion. Accessed: Jun. 29, 2018. [Online]. Available: https://dandelion.eu

[21] OpenCellID. The World’s Largest Open Database of Cell Towers. Accessed: Jun. 29, 2018. [Online]. Available: https://opencellid.org/

[22] Google. Google Places API. Accessed: Jun. 29, 2018. [Online]. Available: https://developers.google.com/places/

QINGTIAN ZENG received the B.S. and M.S. degrees in computer science from the 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 with the Shandong University of Science and Technology, Qingdao, China. His research interests include petri nets, process mining, and knowledge management.

QIANG SUN received the B.S. degree in communication engineering from Shandong Liaocheng University, Shijiazhuang, China, in 2017. His current research interests include the 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 the 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 with the College of Electronic and Information Engineering, Shandong University of Science and Technology. His current research interests include 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, Taian, China, in 1999 and 2002, respectively, and the Ph.D. degree in applied mathematics from Shanghai Jiao Tong University, in 2008. She is currently a Professor with the Shandong University of Science and Technology. Her research interests include process mining and machine learning.

**CHAO LI** received the B.S. and M.S. degrees from the Shandong University of Science and Technology and the Ph.D. degree from the Chinese Academy of Sciences, in 2014. From 2014 to 2015, he was a Visiting Scholar with The Hong Kong University of Science and Technology. He is currently a Lecturer with the Shandong University of Science and Technology. His research interests include social media, natural language processing, data mining, and network embedding learning. He is a member of CCF.

**GE SONG** received the master’s degree in computer science and technology from the Shandong University of Science and Technology, Shandong, China, in 2004. He is currently teaching with the College of Electronic and Information Engineering. His current research interests include the Internet of Things, wireless sensor networks, robot path planning, and deep learning.

***