Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data

Chuanting Zhang, *Student Member, IEEE*, Haixia Zhang, *Senior Member, IEEE*, Jingping Qiao, *Student Member, IEEE*, Dongfeng Yuan, *Senior Member, IEEE*, and Minggao Zhang

***Abstract*— Machine (deep) learning-enabled accurate traffic modeling and prediction is an indispensable part for future big data-driven intelligent cellular networks, since it can help autonomic network control and management as well as service provisioning. Along this line, this paper proposes a novel deep learning architecture, namely Spatial–Temporal Cross-domain neural Network (STCNet), to effectively capture the complex patterns hidden in cellular data. By adopting a convolutional long short-term memory network as its subcomponent, STCNet has a strong ability in modeling spatial–temporal dependencies. Besides, three kinds of cross-domain datasets are actively collected and modeled by STCNet to capture the external factors that affect traffic generation. As diversity and similarity coexist among cellular traffic from different city functional zones, a clustering algorithm is put forward to segment city areas into different groups, and consequently, a successive inter-cluster transfer learning strategy is designed to enhance knowledge reuse. In addition, the knowledge transferring among different kinds of cellular traffic is also explored with the proposed STCNet model. The effectiveness of STCNet is validated through real-world cellular traffic datasets using three kinds of evaluation metrics. The experimental results demonstrate that STCNet outperforms the state-of-the-art algorithms. In particular, the transfer learning based on STCNet brings about 4%∼13% extra performance improvements.**

***Index Terms*— Cellular traffic prediction, big data, deep learning, intelligent traffic management.**

I. Introduction

As TECHNOLOGICAL innovations gather pace, the smart phone evolution in the past decade has accelerated data generation and explosion, which has sped the era of big data [1]–[4]. Among all the data sources, mobile traffic [5], [6] will represent 20 percent of the total Internet traffic by 2021 and particularly, data traffic produced by smartphones will surpass 86 percent of all the mobile data traffic as the emergence of various mobile applications such as live streaming, virtual reality and Internet of vehicles [7]. To meet the diverse requirements of the mobile users, an emerging consensus on the adoption of machine (deep) learning and artificial intelligence (AI) [8]–[10] to the fifth-generation mobile networks (5G) and beyond has been intensively investigated [11]–[14]. In particular, the International Telecommunication Union (ITU) has recently launched a new focus group to assist AI and machine learning in contributing to the efficiency of the emerging 5G systems. The introduction of AI will enable wireless networks to self-optimize, improve efficiency, and deliver optimal user experiences, and consequently, lead to more stable network connections for individual users and for businesses [15].

On the way to AI-enhanced fully automated network management, one of the essential problems lies in the accurate traffic prediction [11] because many tasks in wireless communications require real-time or non-real time traffic analysis and prediction capabilities. For instance, the efficiency of the demand-aware resource allocation is largely benefited from the accurate prediction of future wireless traffic [16]. Besides, the functional base station (BS) sleeping mechanism also relies heavily on the knowledge of the predicted traffic of specific BSs or areas to achieve the purposes of green communications and critical user requirements [17]. However, it is a very challenging task to simultaneously predict cellular traffic network-widely due to the following reasons. Firstly, mobile users have various needs at different time in different places and this makes the traffic hard to predict. Secondly, user mobility introduces spatial dependencies into cellular traffic among geographically distributed cells [18]. Finally, cellular traffic is influenced by many external factors such as the number of BSs. These factors further complicate the spatiotemporal dependencies among cellular traffic of different BSs.

Researchers in recent years have made great efforts to solve the above mentioned challenges. Inherently, cellular traffic prediction can be treated as a time series forecasting problem. According to the solving methods, existing works can be roughly divided into two categories, i.e., statistical-based methods and machine learning-based methods. For the first category, the cellular traffic is modeled and predicted

Manuscript received July 21, 2018; revised December 21, 2018; accepted March 1, 2019. Date of publication March 14, 2019; date of current version May 15, 2019. This work was supported in part by the National Science Fund of China for Excellent Young Scholars under Grant 61622111 and in part by the National Natural Science Foundation of China under Grant 61671278 and Grant 61860206005. *(Corresponding author: Haixia Zhang.)*

C. Zhang, D. Yuan, and M. Zhang are with the Shandong Provincial Key Laboratory of Wireless Communication Technologies, Shandong University, Jinan 250100, China (e-mail: chuanting.zhang@gmail.com; dfyuan@sdu.edu.cn; zhangmg@cae.cn).

H. Zhang is with the Shandong Provincial Key Laboratory of Wireless Communication Technologies, Shandong University, Jinan 250100, China, and also with the School of Control Science and Engineering, Shandong University, Jinan 250061, China (e-mail: haixia.zhang@sdu.edu.cn).

J. Qiao is with the School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China (e-mail: qiaojingping07@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2019.2904363

based on the statistics or probabilistic distributions, including α-stable distribution [19], AutoRegressive Integrated Moving Average (ARIMA) [20]–[22], covariance function [23] and entropy theory [24]. These works have made a comprehensive exploration and characterization on the patterns and characteristics of cellular traffic. It has been demonstrated that the traffic is temporally self-similar and spatially inhomogeneous but also correlated to each other. In prediction, the spatial and/or temporal dependencies are modeled to improve performance. Generally, most of these methods are linear statistical methods. However, it becomes increasingly clear that linear models are not adapted to many real applications [25]. Though several nonlinear models were proposed such as generalized autoregressive conditional heteroskedasticity, the analytical study of nonlinear forecasting method is still in its infancy compared to linear models.

For the second category, with the accumulation of massive cellular traffic data and the advances in machine learning and AI techniques [26]–[29], data-driven machine learning-based traffic prediction methods have established themselves as strong competitors to classical statistical models and obtained tremendous attentions in wireless communication domain [30]–[33]. In the beginning, several shallow learning methods such as linear regression [34] and support vector regression (SVR) [35], are utilized for traffic prediction. With the fast development and widespread penetration of deep learning [10], how to make an accurate traffic prediction for cellular networks by leveraging the powerful deep learning techniques has become a hot topic. Nie *et al.* [30] proposed a deep learning-based prediction method in which the temporal dependence is captured by the low-pass component of discrete wavelet transform. To further capture the spatial dependence of wireless traffic among geographically distributed cell towers, [31] designed a hybrid deep learning model for spatiotemporal prediction, in which the spatial dependence is modeled by autoencoders and the temporal dependence is captured by Long Short-Term Memory networks (LSTM). Instead of using all neighboring traffic information, the most correlated neighbors that have the highest correlation coefficients with the target BS are selected to provide the spatio-temporal information [36]. In order to simultaneously capture the spatial and temporal dependencies of traffic and predict the traffic in citywide scale, the convolutional neural networks (CNN) are leveraged in [32] and [33]. Specifically, Zhang et al. [32] proposed a novel framework by fusing different kinds of temporal dependencies, i.e., closeness and period, using a parametric-matrix-based fusion strategy, then densely connected CNN [37] is introduced to learn spatial dependence and enhance feature propagation. While in [33], the authors proposed a multi-step prediction framework based on convolutional LSTM (ConvLSTM), which has the ability of modeling temporal and long-distance spatial dependencies.

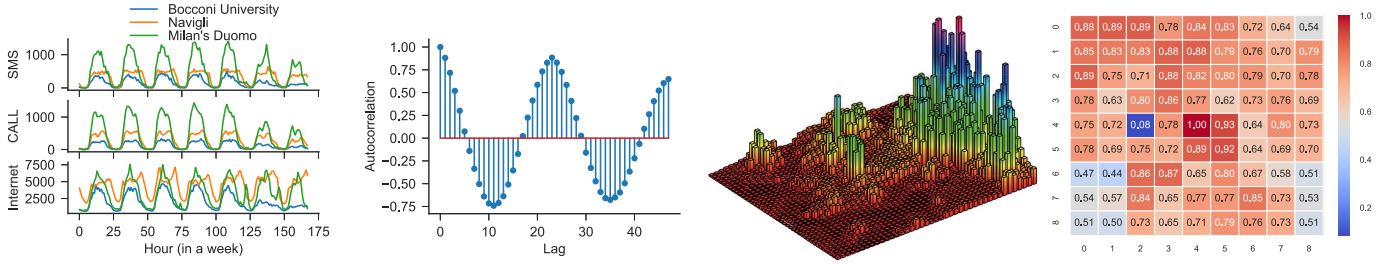
All the above works mainly focus on the cellular traffic dataset itself, and various external factors such as BSs information and POIs distribution, are hardly ever considered. However, it is well understood that these influential factors are directly correlated to the generation of cellular traffic [5], [6]. Besides, current works on network-wide traffic prediction failed to capture the pattern diversity of different city functional zones and the traffic similarity of various services.

Motivated by the aforementioned problems, this work focuses itself on deep learning-based accurate traffic prediction in cellular networks under the scenario of cross-domain big data. In order to make a full characterization on external factors that influence cellular traffic volume, three kinds of cross-domain datasets, i.e., BSs information, POIs distribution and social activity level, are actively collected in this paper. The correlations between these datasets and different kinds of cellular traffic are investigated and analyzed in detail to facilitate cellular traffic prediction. The citywide cellular traffic data are then clustered into several groups to model pattern diversity of different functional zones. After the clustering operation, a novel traffic prediction framework, namely, Spatial-Temporal Cross-domain neural Network (*STCNet*), is designed and a successive inter-cluster transfer learning strategy is put forward to enhance prediction performance. To further exploit the similarities of different kinds of cellular traffic, the model-based transfer learning is also explored in this work. Specifically, the main contributions of this paper can be summarized as follows.

- Three kinds of cross-domain datasets are actively collected and their correlations with different types of cellular traffic are investigated in detail. Although it is a preliminary analysis on these datasets, it is of great importance in designing prediction model.
- A novel deep learning based traffic prediction architecture is proposed and it can effectively fuse the cross-domain datasets into a unified representation. The spatial, temporal and various external factors that influence traffic generation can be well captured by ConvLSTM and CNN. The dense connectivity pattern is introduced in the feature learning process and it can enhance the feature propagation and reuse for traffic prediction.
- To capture the pattern diversity and similarity of cellular traffic of different city functional zones, a clustering algorithm is put forward to segment city areas into different clusters. Then, a successive inter-cluster transfer learning strategy is proposed to capture the regional differences and similarities from spatial and temporal domain, respectively.
- The model-based deep transfer learning is also explored to fully utilize the spatiotemporal similarities of different kinds of cellular traffic, thus further improve prediction performance.

Our findings show that the cross-domain datasets have a high correlation with the cellular traffic and the introduction of these datasets greatly benefits the prediction performance. To demonstrate the superiority of the proposed deep learning-based traffic prediction architecture, it is validated on three kinds of real world cellular traffic datasets. The experimental results show that the proposed prediction architecture outperforms baseline methods greatly and the model-based transfer learning can indeed improve prediction performance.

The rest of this paper is organized as follows. Section II gives a detailed description of the cellular traffic dataset and a preliminary analysis associated with the main observations on



(a) Temporal dynamics of cellular traffic. (b) Autocorrelation analysis. (c) Spatial distribution of cellular traffic. (d) Spatial correlation analysis.

Fig. 1. The spatial and temporal dynamics of cellular traffic.

the dataset. Then Section III introduces the overall prediction framework based on deep learning including spatiotemporal and external factors modeling of cellular traffic. The detailed experiment setup and results analysis are shown in Section IV. Finally Section V concludes the work.

II. Dataset Description and Preliminary Analysis

In this section, detailed cellular traffic dataset considered in this work is introduced. The traffic dynamics along with the spatiotemporal dependencies are displayed from spatial and temporal domain, respectively. In addition, the cross-domain datasets that model the external factors to cellular traffic are described and their correlations with different kinds of cellular traffic are also represented.

A. Big Traffic Data From Cellular Networks

The cellular traffic dataset analyzed in this paper comes from a large telephony services provider in Europe, Telecom Italia, as part of the "*Big Data Challenge*" [38]. The dataset is collected from 11/01/2013 to 01/01/2014 with a temporal interval of 10 minutes over the whole city of Milan (62 days, 300 million records, about 19 GB). The area of Milan is divided into a grid overlay of $H \times W (100 \times 100)$ squares and the size of each square is about 235×235 meters and we refer to it as a cell.[1] In each cell, three kinds of cellular traffic are recorded by the service provider, that is, short message service (SMS), call service and Internet service. For a specific service type $s \in \{\text{SMS}, \text{Call}, \text{Internet}\}$, the city wide cellular traffic can be denoted as a spatiotemporal sequence of data points $\mathcal{D}_s = \{\mathbf{D}_{s,t} | t = 1, 2, \cdots, T\}$, where T is the total number of time intervals. $\mathbf{D}_{s,t}$ is the traffic matrix at the t-th time interval in a geographical area represented as $H \times W$ cells and it can be written as

$$\mathbf{D}_{s,t} = \begin{bmatrix} d_{s,t}^{(1,1)} & d_{s,t}^{(1,2)} & \cdots & d_{s,t}^{(1,W)} \\ d_{s,t}^{(2,1)} & d_{s,t}^{(2,2)} & \cdots & d_{s,t}^{(2,W)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{s,t}^{(H,1)} & d_{s,t}^{(H,2)} & \cdots & d_{s,t}^{(H,W)} \end{bmatrix}, \quad (1)$$

where $d_{s,t}^{(h,w)}$ measures the cellular traffic volume in a cell with coordinates (h, w) and the sequence can be regarded as a tensor $\mathcal{D}_s \in \mathbb{R}^{T \times H \times W}$. Note that the next preliminary analysis from spatial and temporal domain is suitable to any kinds of cellular traffic. Thus, for ease of readability, the notation of service type is omitted in the following, that is, $\mathbf{D}_{s,t} = \mathbf{D}_t$ and $d_{s,t}^{(h,w)} = d_t^{(h,w)}$, unless otherwise specified. After carefully exploring the cellular traffic dataset, the spatial and temporal dynamics and the corresponding correlation analysis are demonstrated in Fig.1.

Fig.1a shows the temporal dynamics of different kinds of cellular traffic in different cells. The x-axis denotes the time interval index (in hour scale) and y-axis the number of events of a specific cellular traffic. The upper sub-figure shows the SMS traffic dynamics, the middle one is the CALL and the bottom one shows the Internet. It can be clearly seen from Fig.1a that: 1) The cellular traffic, no matter which type they belong to, show strong daily patterns in all the three different cells; 2) The daily patterns of different cellular traffic are not the same. For example, the Internet service has shorter peak traffic hours compared with SMS service and Call service; 3) For a specific service, the traffic patterns of different cells have considerable differences. Taking the cell Navigli as an example, there is a significant delay in the arriving of the peak traffic hours compared with the other two cells. Fig.1b displays the autocorrelation of SMS in a specific cell.[2] The autocorrelation coefficient [32] at cell (h, w) is computed as follows:

$$r_k = \frac{\sum_{t=1}^{T-k} (d_t^{(h,w)} - \bar{d}^{(h,w)})(d_{t+k}^{(h,w)} - \bar{d}^{(h,w)})}{\sum_{t=1}^{T} (d_t^{(h,w)} - \bar{d}^{(h,w)})^2}, \quad 0 \leqslant k \leqslant T, \quad (2)$$

where $\bar{d}^{(h,w)}$ represents the mean value of the cell over time domain. As depicted in Fig.1b, the cellular traffic exhibits non-zero autocorrelations in time domain and this indicates the future traffic volume can be predicted through historical observations.

Fig.1c shows the spatial distribution of SMS traffic at a specific time interval. We can see from this figure that the traffic are distributed unevenly among the whole city. Accordingly, Fig.1d shows the spatial correlation in terms of Pearson correlation coefficient ρ, which is a widely adopted metric for measuring spatial correlations [31], [32] and its

[1] This is actually the best coverage approximation of a cellular tower from publicly available dataset in such a large scale and fine granularity.

[2] The other traffic of different cells show similar results and we only take SMS traffic for demonstration purpose and omit the other autocorrelation results.

definition is expressed as

$$\rho = \frac{\mathrm{cov}(d^{(h,w)}, d^{(h',w')})}{\sigma_{d^{(h,w)}}\sigma_{d^{(h',w')}}}, \tag{3}$$

where $\mathrm{cov}(\cdot)$ denotes the covariance operator and σ is the standard deviation. It can be clearly told that the spatial correlation indeed exists among different cells. The degree of spatial correlation depends, to a certain extent, not only on the distance between any two cells. For example, though cell $(3, 5)$ and cell $(5, 5)$ are with the same distance to the target cell $(4, 4)$, their correlation values, which are 0.62 and 0.92, respectively, are not the same.

B. Cross-Domain Datasets Description

The cellular traffic volume is influenced not only by the spatiotemporal factors, but also by other external factors such as the number of BSs and POIs of a cell. For example, the number of BSs of a cell decides how much traffic load can be carried. When the traffic load reaches its peak, the observed traffic volume is fixed no matter how many extra users enter this cell. In order to achieve accurate cellular traffic prediction, multiple influencing factors must be considered based on cross-domain datasets, since different datasets can characterize the cellular traffic from different perspectives. Intuitively, the number of BSs and POIs as well as social activities of a cell can directly reflect user's requesting ability for telecommunication services, therefore three kinds of cross-domain datasets are considered in this work,[3] i.e., the BSs information, the POIs distribution and the social activity level.

The dataset about BSs information is obtained from OpenCellID [39], which is an open-source project collecting data about mobile cells all over the world. This dataset contains many types of information about the BS such as the location (longitude and latitude), the mobile country code and the estimated coverage range of each BS. With the geolocation information of each cell, we can map the location of each BS to the cell which the BS is located in after simple preprocessing. Then the number of BSs of each cell $d_{\mathrm{BS}}^{(h,w)}$ can be calculated. This BSs information is denoted as

$$\mathbf{D}_{\mathrm{BS}} = \begin{bmatrix} d_{\mathrm{BS}}^{(1,1)} & d_{\mathrm{BS}}^{(1,2)} & \cdots & d_{\mathrm{BS}}^{(1,W)} \\ d_{\mathrm{BS}}^{(2,1)} & d_{\mathrm{BS}}^{(2,2)} & \cdots & d_{\mathrm{BS}}^{(2,W)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{\mathrm{BS}}^{(H,1)} & d_{\mathrm{BS}}^{(H,2)} & \cdots & d_{\mathrm{BS}}^{(H,W)} \end{bmatrix}. \tag{4}$$

For the POIs distribution information, which can be crawled using Google Places API [40]. Specifically, 13 kinds of POIs are collected including subway station, store, restaurant, etc. The detailed description on each category is displayed in Table.I. The number of each category is added together to form the final representation. The matrix generated through

[3]The cellular traffic may also influenced by other factors, but they are not considered in this work due to the availability of datasets.

TABLE I
DETAILED STATISTICS OF THE DATASETS

Dataset	Type	# of records
Cellular traffic	SMS / Call / Internet	$\approx$ 300 million
POI	Subway station	104658
	Store	19748
	Church	512
	Cafe	995
	Park	765
	Library	188
	Bank	882
	Bar	3192
	Parking	392
	Restaurant	4666
	School	1284
	Lodging	2922
	Hospital	1585
BSs	GSM / CDMA / LTE	69909
Social activity	Twitter	269290

POI dataset is expressed as

$$\mathbf{D}_{\mathrm{POI}} = \begin{bmatrix} d_{\mathrm{POI}}^{(1,1)} & d_{\mathrm{POI}}^{(1,2)} & \cdots & d_{\mathrm{POI}}^{(1,W)} \\ d_{\mathrm{POI}}^{(2,1)} & d_{\mathrm{POI}}^{(2,2)} & \cdots & d_{\mathrm{POI}}^{(2,W)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{\mathrm{POI}}^{(H,1)} & d_{\mathrm{POI}}^{(H,2)} & \cdots & d_{\mathrm{POI}}^{(H,W)} \end{bmatrix}. \tag{5}$$

The social activities of a cell reflect the overall user demand degree for network services. The dataset on social activity level is obtained through Dandelion API [41]. The obtained data contains the information a user generated when using twitter, such as the location and keywords. After preprocessing, the social activity level can be obtained and we denote this matrix as

$$\mathbf{D}_{\mathrm{Social}} = \begin{bmatrix} d_{\mathrm{Social}}^{(1,1)} & d_{\mathrm{Social}}^{(1,2)} & \cdots & d_{\mathrm{Social}}^{(1,W)} \\ d_{\mathrm{Social}}^{(2,1)} & d_{\mathrm{Social}}^{(2,2)} & \cdots & d_{\mathrm{Social}}^{(2,W)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{\mathrm{Social}}^{(H,1)} & d_{\mathrm{Social}}^{(H,2)} & \cdots & d_{\mathrm{Social}}^{(H,W)} \end{bmatrix}, \tag{6}$$

where $d_{\mathrm{social}}^{(h,w)}$ represents the number of social activities of cell (h, w).

The complexity of obtaining the above three kinds of cross-domain datasets is relatively low as there exists standard APIs for these data. These datasets are treated as static in our work, this is because they will not frequently change over a period of time. So once obtained, these datasets can be used in model training. The heatmaps generated by the above three kinds of cross-domain datasets along with the city topology and cell partition of Milan are displayed in Fig.2, which contains 5 layers. From the bottom layer to the top one, these 5 layers are city topology of Milan, illustration of different cells, BSs distribution, social activity level and POIs distribution. It can be seen from this figure and Fig.1c that the cross-domain datasets have similar spatial distribution compared with cellular traffic of different services. The city center has more facilities than rural areas, thus more traffic is generated in these places.

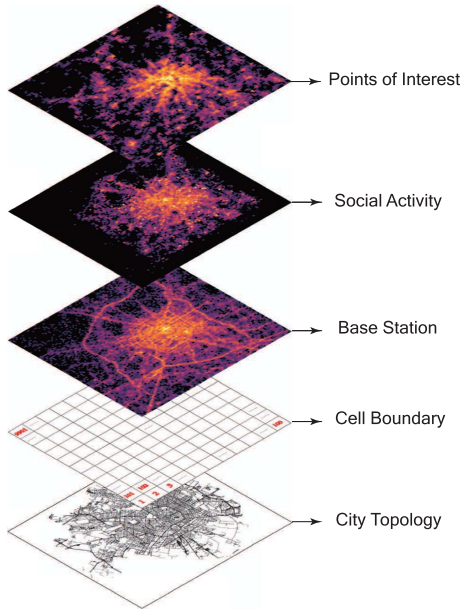


Fig. 2. Heatmap of cross-domain datasets along with the city topology and cell division of Milan.

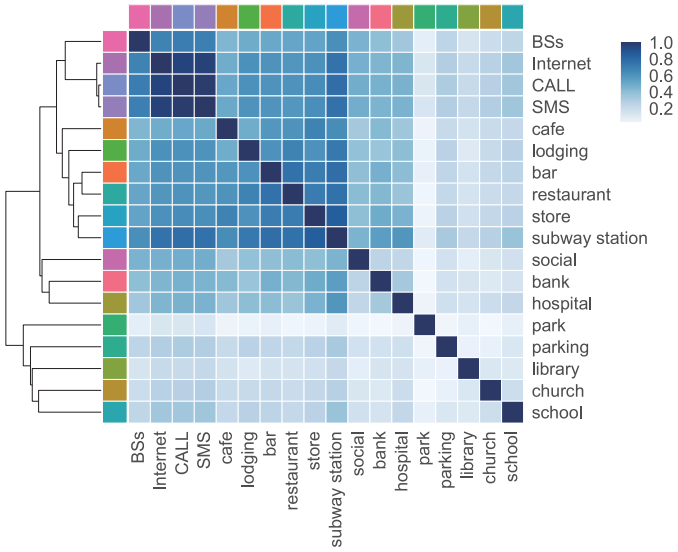


Fig. 3. Correlation analysis between different kinds of cellular traffic and external datasets.

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 figure, several important observations can be concluded in the following. 1) The spatial correlations, which can be calculated using equation (3), among three kinds of cellular traffic are the highest. This indicates that different kinds of cellular traffic have a certain similarity, thus the pattern knowledge learned from one type of traffic data can be transferred to another type of traffic data. 2) The BSs distribution is most correlated to cellular traffic compared with the other datasets. This shows that the number of BSs in a cell indeed influences the traffic generation thus it can be used as features to facilitate traffic prediction. 3) In terms of POI categories, social gathering places like cafe, bar and restaurant have high correlation

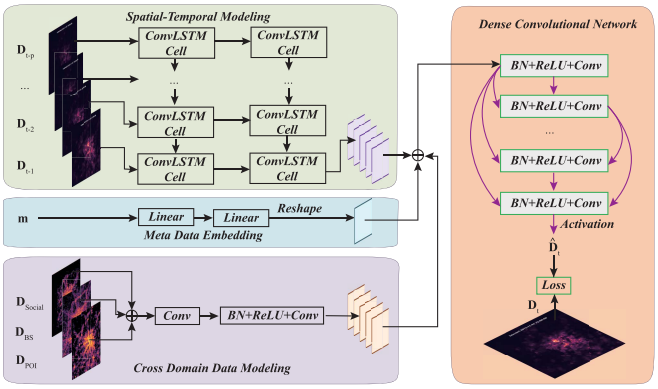


Fig. 4. *STCNet* framework.

coefficients with cellular traffic. This indicates that the number of POIs of a cell can influence the number of users, thus further influence the traffic volume. 4) The social activity level, along with the other kinds of POIs have a less correlation with cellular traffic, but the values are also non-zeros.

Based on analysis of the above spatial correlations, it can be concluded that the cross-domain datasets considered, i.e., BSs information, POIs distribution and social activity level, have high correlation coefficients with different kinds of cellular traffic. They should be considered in the design of cellular traffic prediction model.

III. Cellular Traffic Prediction Framework: STCNet

In this section, the proposed deep learning prediction architecture, *STCNet*, is introduced and the diagram is displayed in Fig.4. Specifically, a comprehensive description on the main body of *STCNet* is given first. To segment city areas into different functional zones, the clustering algorithm is designed. Then, a successive inter-cluster transfer learning strategy is put forward to enhance knowledge reuse and prediction performance.

A. Prediction Model

There are three inputs for the proposed *STCNet*. As shown in Fig.4, the first input $(\mathbf{D}_{t-1}, \cdots, \mathbf{D}_{t-p}, p \in \mathcal{N})$ is a sequence of data traffic matrix before the target time interval. The second input $(\mathbf{m})$ refers to the metadata of datetime corresponding to the time interval, such as day of week and hour of day. The third input represents the cross-domain datasets including BSs information $(\mathbf{D}_{\mathrm{BS}})$, POIs distribution $(\mathbf{D}_{\mathrm{POI}})$ and social activity $(\mathbf{D}_{\mathrm{Social}})$. To handle these three different inputs according to their data format and characteristics, three kinds of neural networks are designed as follows.

1) Spatiotemporal Modeling: The first input can be seen as a video-like data which has p frames[4] and each frame is a one-channel image. It is well know that CNN has strong abilities to model spatial dependence as it can effectively fuse local area information and automatically extract features for specific tasks. But time sequence information is not precisely

[4]A frame denotes a traffic snapshot at one time interval.

captured by CNN. While LSTM networks can well model the sequences information of cellular traffic. Thus by combining CNN and LSTM, a two-layer ConvLSTM network is designed to simultaneously model the spatial-temporal dependencies and the sequence information, as shown in Fig.4.

Each unit in the ConvLSTM layer has one memory cell $\mathcal{C}$ to accumulate state information. This memory cell can be accessed and modified through three self-parameterized controlling "gates", i.e., input gate i_g, forget gate f_g and output gate o_g. Specifically, whenever a new input comes to the ConvLSTM unit, the information it carries can be stored to $\mathcal{C}$ if the input gate i_g is activated. Similarly, the past cell status can also be forgotten in this process if the forget gate f_g is on. The final hidden state $\mathcal{H}$ is controlled by the output gate o_g, which decides whether the cell output $\mathcal{C}$ should be propagated to the final state or not. We specify the key operations of ConvLSTM unit on the frame $\mathbf{D}_{t-n}$ in below, where $n \in \{1, 2, \cdots, p\}$. $\sigma(\cdot)$ denotes the activation function, $*$ denotes the convolution operation and $\odot$ is the Hadamard product:

$$\begin{aligned} i_g^\tau &= \sigma(\mathbf{W}_{di} * \mathbf{D}_\tau + \mathbf{W}_{hi} * \mathcal{H}_{\tau-1} + \mathbf{W}_{ci} \odot \mathcal{C}_{\tau-1} + b_i), \\ f_g^\tau &= \sigma(\mathbf{W}_{df} * \mathbf{D}_\tau + \mathbf{W}_{hf} * \mathcal{H}_{\tau-1} + \mathbf{W}_{cf} \odot \mathcal{C}_{\tau-1} + b_f), \\ \mathcal{C}_\tau &= f_g^\tau \odot \mathcal{C}_{\tau-1} + i_g^\tau \odot \tanh(\mathbf{W}_{dc} * \mathbf{D}_\tau + \mathbf{W}_{hc} * \mathcal{H}_{\tau-1} + b_c), \\ o_g^\tau &= \sigma(\mathbf{W}_{do} * \mathbf{D}_\tau + \mathbf{W}_{ho} * \mathcal{H}_{\tau-1} + \mathbf{W}_{co} \odot \mathcal{C}_\tau + b_o), \\ \mathcal{H}_\tau &= o_g^\tau \odot \tanh(\mathcal{C}_\tau). \end{aligned} \tag{7}$$

In the above equation, $\mathbf{W}_{(\cdot)}$ and $b_{(\cdot)}$ are the weights and biases to be learned, respectively. Besides, $\tanh(\cdot)$ refers to the hyperbolic tangent function which acts as nonlinear transformation of input. Note that the i_g^τ, f_g^τ, o_g^τ, $\mathcal{C}_\tau$ and $\mathcal{H}_\tau$ in the ConvLSTM unit are all three-dimensional tensors. The output of ConvLSTM network is denoted as $\mathbf{O}_{\text{st}} \in \mathbb{R}^{F \times H \times W}$, where F is the number of feature maps.

2) Meta-Data Embedding: As the date and time information of the cellular traffic is recorded when mobile users asking for services, we extract the meta data and treat them as features. For example, if the datetime of t-th time interval is 13:00:00 11/19/2014, then four kinds of meta data are extracted, i.e., Day_of_Week (Tuesday), Hour_of_Day (13), is_Weekday (Yes), is_Weekend (No), and form a feature vector $\mathbf{m}$. This feature vector is fed into a two-layer neural network, in which the dimensionality of $\mathbf{m}$ is increased from 4 to $F \times H \times W$. The mathematical expression of $\mathbf{o}_{\text{meta}}$ is denoted as

$$\mathbf{o}_{\text{meta}} = \sigma(\mathbf{w}^2_{\text{meta}} \sigma(\mathbf{w}^1_{\text{meta}} \mathbf{m} + b^1_{\text{meta}}) + b^2_{\text{meta}}), \tag{8}$$

where $\mathbf{w}^l_{\text{meta}}$ and b^l_{meta} are learnable parameters at l-th layer, $l \in \{1, 2\}$. Then after a reshape operation, the output of this component, $\mathbf{O}_{\text{meta}}$, can be obtained.

$$\mathbf{O}_{\text{meta}} = \text{Reshape}(\mathbf{o}_{\text{meta}}), \tag{9}$$

where $\mathbf{o}_{\text{meta}} \in \mathbb{R}^{FHW \times 1}$ and $\mathbf{O}_{\text{meta}} \in \mathbb{R}^{F \times H \times W}$.

3) Cross-Domain Data Modeling: To model external influencing factors of traffic generation and learn feature representations contained in the cross-domain datasets, a two-layer CNN architecture is designed. In this architecture, the datasets $\mathbf{D}_{\text{BS}}$, $\mathbf{D}_{\text{POI}}$ and $\mathbf{D}_{\text{Social}}$ are processed into a tensor $\mathbf{D}_{\text{cross}}$ through concatenation operation. After performing nonlinear transformation on $\mathbf{D}_{\text{cross}}$, the initial feature representations of cross-domain datasets $\mathbf{O}_{\text{cross}}$ can be obtained and written as

$$\mathbf{O}_{\text{cross}} = f(\mathbf{W}_{\text{cross}} * \mathbf{D}_{\text{cross}}), \tag{10}$$

$$\mathbf{D}_{\text{cross}} = \mathbf{D}_{\text{BS}} \oplus \mathbf{D}_{\text{POI}} \oplus \mathbf{D}_{\text{Social}}, \tag{11}$$

where $\oplus$ is the concatenation operation and similarly, $\mathbf{W}_{\text{cross}}$ is the weights that will be learned through optimization. $f(\cdot)$ represents a composite function that implements the Batch Normalization (*BN*),[5] rectified linear units (*ReLU*) and convolution operation (*Conv*) sequentially.

4) Feature Learning With Dense Convolutional Network: The three outputs are fused together by the concatenation operation which can be expressed as

$$\mathbf{O} = \mathbf{O}_{\text{st}} \oplus \mathbf{O}_{\text{meta}} \oplus \mathbf{O}_{\text{cross}}, \tag{12}$$

where $\mathbf{O}$ refers to the overall representation of the initial feature map and is also the input to the dense convolutional network. Note that the addition operation is not recommended in this process since it mixes different kinds of information together and does not benefit the effective feature learning. This component consists of L layers and each layer implements a composite function $f_l(\cdot)$, which is the same as the one in cross-domain data modeling, that is, $f_l(\cdot) = f(\cdot)$ except that l indexes the layer.

In order to fully capture the spatial-temporal dependencies and many other external influencing factors that affect cellular traffic volume, the dense connectivity pattern is designed in this component. This kind of connectivity denotes that there exist direct connections from any layer to all subsequent layers and the resulting layout of this component is illustrated in Fig.4. Consequently, the l-th layer receives the feature maps of all preceding layers, $\mathbf{O}_0, \mathbf{O}_1, \cdots, \mathbf{O}_{l-1}$, as input:

$$\mathbf{O}_l = f_l(\mathbf{O}_0 \oplus \cdots \oplus \mathbf{O}_{l-1}), \tag{13}$$

where $\mathbf{O}_0 = \mathbf{O}$. The output at the last layer of this component can be expressed as $\mathbf{O}_L \in \mathbb{R}^{H \times W}$. After an activation operation, the final prediction is obtained.

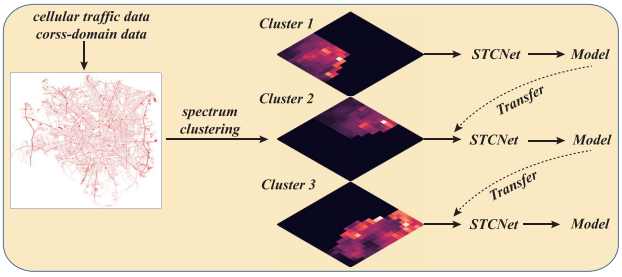
$$\hat{\mathbf{Y}} = \sigma(\mathbf{O}_L). \tag{14}$$

Thus the objective function of *STCNet* is to minimize the Frobenius norm of error matrix between prediction and ground truth over all cells. It can be expressed as

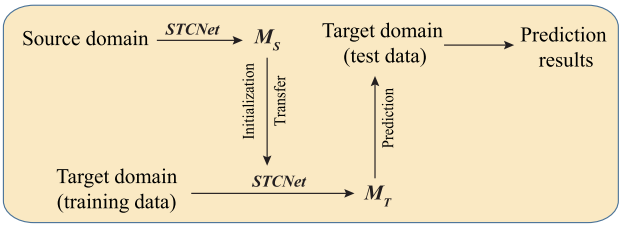
$$\mathcal{L}(\theta) = \arg\min_\theta \|\hat{\mathbf{Y}} - \mathbf{Y}\|_F, \tag{15}$$

where θ refers to the parameters of *STCNet* and can be easily trained through optimization techniques.

[5]BN is used to accelerate deep neural network training by reducing internal covariate shift. Training a deep network model is complicated by the fact that the distribution of each layers inputs changes during training, as the parameters of the previous layers change. This slows down the training by requiring lower learning rates and careful parameter initialization, and makes it hard to train models with saturating nonlinearities



(a) City segmentation and successive inter-cluster transfer learning.



(b) Knowledge transfer between source domain and target domain.

Fig. 5. Transfer learning strategy for traffic prediction.

B. Successive Inter-Cluster Transfer Learning

The pattern of cellular traffic for different areas of cellular networks is very complex since diversity and similarity coexist. These coupled influence factors will lead to serious performance degradation of prediction methods, if they are not properly captured and modeled. Hence, aiming to capture the pattern diversity and similarity of different areas in generating cellular traffic, a clustering algorithm and a transfer learning strategy are proposed in this work, as shown in Fig.5a.

For the clustering algorithm, by incorporating cellular traffic dataset $\mathcal{D}_s$ and cross-domain datasets $\mathbf{D}_{\text{BS}}$, $\mathbf{D}_{\text{POI}}$ and $\mathbf{D}_{\text{Social}}$, the overall feature representation $\mathbf{D}$ is first formed. Then a graph can be obtained, in which vertices are the cells and edges denote the adjacency information of different cells, that is, if two cells are horizontally or vertically adjacent, then they form an edge in the graph. Based on the descriptions of vertices and edges, the corresponding adjacency matrix $\mathbf{A}$ can be calculated whose weights are the gradients of $\mathbf{D}$. Next, the Laplacian matrix can be constructed as

$$\mathbf{L} = \mathbf{P}^{-1/2}\mathbf{A}\mathbf{P}^{-1/2}, \tag{16}$$

where $\mathbf{P}$ is a diagonal matrix whose (i, i)-element is the sum of $\mathbf{A}$'s i-th row. After that, the k largest eigenvectors of $\mathbf{L}$, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$, can be achieved and form the matrix $\mathbf{X} = [\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_k] \in \mathbb{R}^{n\times k}$ by stacking the eigenvectors in columns. Note that the rows of $\mathbf{X}$ should be renormalized to have unit length. The rows of $\mathbf{X}$ can be treated as features of cells, then we perform K-Means algorithm on $\mathbf{X}$ and thus k clusters are found. Finally, the cluster label of each cell is obtained.

On the basis of clustering results, an inter-cluster transfer learning strategy is designed to learn the pattern similarity of different areas. This kind of similarity indicates that the traffic pattern knowledge learned from one cluster could be transferred to other clusters. Transfer learning can make the prediction model avoid learning from scratch thus accelerate the optimization process. As shown in Fig.5a, the strategy is described as follows: the dataset of the first cluster is trained through *STCNet* and the parameters can be learned. Then these learned parameters are treated as prior knowledge of the second cluster (parameters' initialization) and continually trained using the dataset of the second cluster. This operation is repeated on all the clusters and we can get all the models. The obtained models can well capture both the pattern diversity and similarity of different areas, since the operation of successive transferring of knowledge.

C. Transfer Learning Among Different Kinds of Cellular Traffic

From the preliminary data analysis in Section II-B, it can be told that there exist high similarities among datasets of SMS, Call, and Internet, which indicate the possibility of knowledge transferring. Thus, similar to the successive inter-cluster knowledge reuse, we further propose a model-based transfer learning strategy to fully exploit the similarities among cellular traffic. The workflow of this strategy is illustrated in Fig.5b. Specifically, a model M_S can be obtained after training *STCNet* using data from the source domain S. Then M_S can be transferred to the target domain T by means of parameter initialization. The parameters of *STCNet* are continually trained using data from target domain and the model M_T can be learned. Finally the prediction is carried out on the test data of the target domain and the results can be obtained.

IV. Experimental Results and Analysis

In this section, extensive experiments are conducted to demonstrate the effectiveness of *STCNet* for cellular traffic prediction. The experimental settings and key parameters are given first. Then the performance metrics and baseline methods which we compared with are described in detail. Next we give the overall prediction performance of *STCNet* and baseline methods on various kinds of cellular traffic and particularly, give the predictions and ground truth comparisons for two randomly selected cells in the city. The transfer learning results on different kinds of cellular traffic are also given.

A. Data Preprocessing and Experiment Settings

As traffic values of some cells at certain time intervals are missing due to data storage error or not properly transmitted. So data completion needs to be done before proceeding to the next stage of cellular traffic prediction. We use a standard way [42] to fill missing values of cell (h, w), that is, these values are represented by the mean traffic volume values of its surrounding cells. The operation is expressed as

$$d_t^{(h,w)} = \sum_{i\in[-1,1]} \sum_{j\in[-1,1]} \frac{d_t^{(h+i,w+j)}}{8}. \tag{17}$$

The granularity of the original cellular traffic is 10-minute and we first group it into hour scale due to the following reasons: 1) Most of the cells have traffic volume of zero

in 10 minutes, thus the data are very sparse and not conducive to traffic prediction. 2) Resource planning, such as cell zooming, in 10-minute level is a very challenging task for network operators and may result in unstable networks or excessive overhead. Then the traffic volume is scaled into range of [0, 1] using Min-Max normalization to accelerate the training process. When the prediction is finished and the evaluation is performed, the value is rescaled back to its normal scale. Data from the first seven weeks are utilized to construct training dataset, and consequently, data from the last week are used to construct test dataset, on which the performance of various prediction algorithms will be tested. Both training and test datasets are constructed using sliding window method with window size $p = 3$.

The *STCNet* is optimized using a stochastic gradient-based optimization technique, Adam [43] which is widely used in current deep learning domain. Besides, the model is trained for 500 epochs with batch size 32. An adaptive learning rate α is adopted in this work, whose initial value is set to be 0.01 and will be divided by 10 and 100 at 50% and 75% of the total number of training epochs accordingly.[6] The initial convolution layers have 16 filters ($F = 16$) with kernel size (3×3) and *ReLU* activation function except for the last layer, which has 1 filter with kernel size (1×1) and *sigmoid* activation function. The values of these parameters are determined based on experiment requirements and other values can also be chosen, but the optimization of these parameters are not focused in this work.

B. Evaluation Metrics and Baseline Methods

Three metrics are adopted in this work for the sake of a comprehensive evaluation of different prediction algorithms.

The first one is Root Mean Square Error (RMSE). This is a frequently used measure of difference between values predicted by a model and the values of ground truth.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}\sum_{h=1}^{H}\sum_{w=1}^{W}(\hat{d}_t^{(h,w)} - d_t^{(h,w)})^2}{T \times H \times W}} \tag{18}$$

The second one is Mean Absolute Error (MAE). MAE measures the average of the absolute differences between prediction and ground truth where all individual differences have equal weight.

$$\text{MAE} = \frac{\sum_{t=1}^{T}\sum_{h=1}^{H}\sum_{w=1}^{W}|\hat{d}_t^{(h,w)} - d_t^{(h,w)})|}{T \times H \times W} \tag{19}$$

The last metric is R-squared (R2), which represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

$$\text{R2} = 1 - \frac{\sum_{t=1}^{T}\sum_{h=1}^{H}\sum_{w=1}^{W}(\hat{d}_t^{(h,w)} - d_t^{(h,w)})^2}{\sum_{t=1}^{T}\sum_{h=1}^{H}\sum_{w=1}^{W}(\bar{d}^{(h,w)} - d_t^{(h,w)})^2} \tag{20}$$

For RMSE and MAE, the smaller the value, the better the performance. On the contrary, for R2 metric, a larger value implies a better fitting to the data, thus, a better performance.

[6]It should be noted that the choices of batch size and learning rate have indeed great influence on the prediction performance and we follow the choices of previous work in deep learning domain. The strategies that finding the best values of these hyper-parameters are beyond the scope of this work.

In order to show the superiority of our proposed cellular traffic prediction model, the performance of *STCNet* is compared with several classical baseline methods that are widely used in time series prediction. The reference approaches including Linear Regression (LR) [34], Support Vector Regression (SVR) [35], LSTM networks [36] and Spatial-Temporal Densely Connected CNN (DenseNet) [32]. The LR and SVR are representatives of shallow machine learning methods and they have a wide range of applications in various fields. Different from LR and SVM, LSTM and DenseNet are two deep learning-based prediction methods that achieve the state-of-the-art in cellular traffic prediction.

C. Prediction Performance on Different Kinds of Traffic

The results of evaluation metrics on three different kinds of cellular traffic are plotted in Fig.6. Among all the sub-figures, Fig.6(a)-(c) represent the results on SMS dataset in terms of RMSE, MAE and R2, respectively. Accordingly, Fig.6(d)-(f) are the results on the Call dataset and Fig.6(g)-(i) the results on the Internet dataset.

As can be clearly seen from Fig.6, our proposed *STCNet* achieves the best prediction results in terms of RMSE, MAE and R2 score on all the three kinds of cellular traffic. The reasons behind the success of our model can be attributed into threefold. First, the spatial and temporal dependencies are simultaneously captured by the *STCNet*, particularly by the ConvLSTM component. Second, the cross-domain datasets, such as BSs information and POIs distribution, put a spatial constraint on the cellular traffic generation. They can be used as features to enhance traffic prediction. Third, the traffic pattern diversity and similarity of different areas are fully exploited through our successive inter-cluster transfer learning strategy. Compared with prediction methods with similar network architecture but using different inputs, *ST-Net* and *STM-Net*,[7] *STCNet* presents the best overall results, which validates the benefits of including meta data and cross-domain datasets.

As cellular traffic dynamics are highly nonlinear from both temporal and spatial dimensions, this makes the prediction of future traffic volume a very challenging task and beyond the ability of linear models. Thus the LR method performs the worst among all the methods. The SVR method, which is a nonlinear model, can deal with the nonlinearities of cellular traffic, thus obtains better results than LR method. For the prediction method based on LSTM network, the performance is not as good as *STCNet* since only temporal dependence is considered and the spatial dependence of cellular traffic from different cells is ignored. Both the spatial and temporal dependencies are captured by the DenseNet-based method, hence it improves prediction performance greatly but still inferior to our proposed *STCNet* model. This is because the sequence information of traffic frames is not modeled. In addition, DenseNet prediction method relies only on the cellular traffic itself and the various external factors that influencing traffic generation are not considered.

[7]Specifically, *ST-Net* takes only cellular traffic as its input, while *STM-Net* has two inputs, i.e., cellular traffic and meta-data feature.

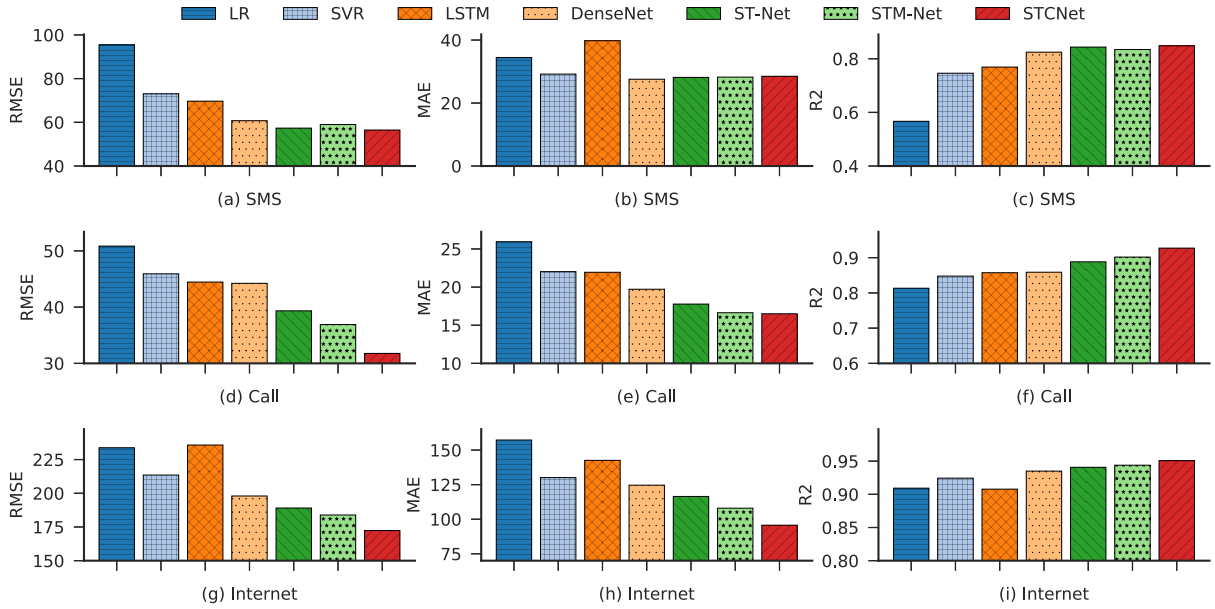


Fig. 6. Comparisons of prediction performance on three different kinds of cellular traffic in terms of various evaluation metrics.

TABLE II
THE IMPACTS OF EACH KIND OF CROSS-DOMAIN DATASETS ON THE PREDICTION PERFORMANCE

Dataset	Type	RMSE	MAE	R2
SMS	No cross-domain	57.7105	32.6001	0.8284
	+ Social	55.5904	27.3112	0.8501
	+ BSs	54.5176	28.7362	0.8576
	+ POIs	52.8764	28.1744	0.8596
Call	No cross-domain	40.1073	17.7448	0.8962
	+ Social	37.0430	17.9974	0.9094
	+ BSs	33.8271	17.6993	0.9252
	+ POIs	33.3415	15.8556	0.9226
Internet	No cross-domain	172.7059	94.1415	0.9488
	+ Social	166.1807	91.8873	0.9504
	+ BSs	167.7501	93.8525	0.9501
	+ POIs	164.3131	90.9473	0.9529

Fig.6 has shown the benefits of introducing cross-domain datasets into cellular traffic prediction, but the contribution of each kind of data is not clear.

Thus the impacts of social, BSs, and POIs datasets on prediction performance are further explored and summarized in Table II. In this table, "No cross-domain" represents the case that we do not use the cross-domain datasets when training the model, while "+A" denotes that dataset A is incorporated into training. For instance, "+ Social" indicates that social activity data is considered when training prediction model. From Table II, it is noticeable that the performance can be considerably improved by introducing cross-domain datasets. Different kinds of cross-domain datasets have different influences on the performance. For SMS dataset, social data brings about 3.7% improvements in terms of RMSE, while BSs 5.5% and POIs 8.4%. At the same time, social data brings about 16.2% improvements with respect to MAE, while BSs 11.9% and POIs 13.6%. So each type of cross-domain datasets has different impacts on the prediction performance. But from the perspective of overall results, BSs and POIs data can bring more performance gains than social data.

D. Comparisons of STCNet and Baselines

To demonstrate the prediction performance of our proposed *STCNet*, the comparison between predicted values and ground truth along with the corresponding error analysis are displayed in Fig.7 and Fig.8. Specifically, Fig.7 shows the results of cell (50, 60), which belongs to the area of Milan's Duomo, a very famous tourist attraction that is located in the city center of Milan. Fig.8 shows the results of another totally different cell (44, 56), which is located in the Navigli District, one of the most famous nightlife places in Milan.

The three sub-figures on the left side of Fig.7 show the comparisons between predicted values and real values on datasets of SMS, Call and Internet traffic, respectively. The x-axises denote the time interval index of the test dataset and the y-axises are the traffic volume. We can see from these sub-figures that our proposed *STCNet* model can accurately predict the traffic values for all the three kinds of cellular traffic. Though the scales of these traffic differ a lot, the peak traffic volume can still be accurately predicted by *STCNet*.

The three sub-figures of bar plot on the middle side of Fig.7 are the corresponding prediction errors in terms of MAE for traffic volume of each time interval. It can be observed from the bar plot that the prediction errors are relatively small. Several large errors appear around the time index of 145, at which the traffic has a sudden increase for all the three kinds of services. This time index actually corresponds to the New Year's Eve and the abnormal traffic volume is very hard to predict, therefore large error occurs. The overall prediction error can be more quantitatively measured by the cumulative distribution function (CDF) as a function of prediction error and the results are plotted in the sub-figures on the right side of Fig.7. Results reflect that about 75% prediction errors

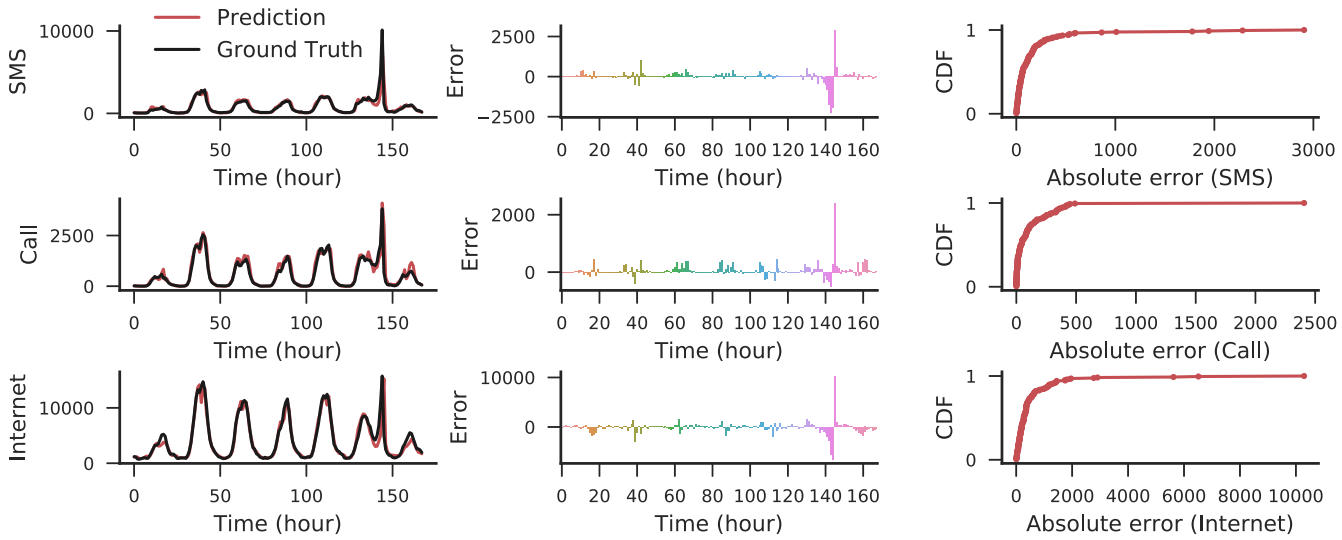


Fig. 7. Prediction results of the first cell (50, 60), Milan's Duomo.

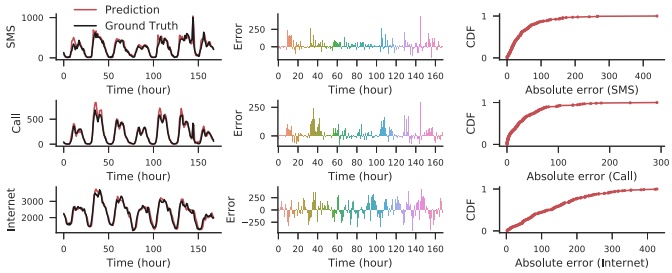


Fig. 8. Prediction results of the second cell (44, 56), Navigli district.

are less than 100 for the SMS traffic. For Call and Internet traffic, the values are 68 and 453, respectively. Based on the comprehensive analysis of Fig.7, we can conclude that the proposed *STCNet* can make a high prediction accuracy for all the three kinds of traffic. Similar conclusions can be made from Fig.8 and for simplicity's sake, we omit them here. From the above analysis, it can be concluded that our proposed *STCNet* remains robust for different cells.

Fig.9 shows the comparisons of the predicted values versus the ground truth for all involved methods and their corresponding CDF of the prediction errors based on Internet dataset of cell (44, 56). Results show that for methods of LR and SVR, there exist relatively big gaps between the predictions and the ground truth. So the performances of shallow learning algorithms are not so good since such algorithms have limited parameter space in modeling complex cellular traffic dynamics. This is because the increased parameter capacity can help well capture the spatial and temporal dependencies of the cellular traffic generated by geographically distributed BSs. *STCNet* achieves the best overall performance especially for peak traffic prediction, which can be intuitively seen from Fig.9. The reasons can be attributed to the strong abilities of *STCNet* in modeling both spatiotemporal dependencies and spatial constraints of cross-domain datasets on cellular traffic generation. To intuitionisticly show the performance difference, we give the CDF of the prediction errors in the last sub-figure of Fig.9. It is shown that 75% prediction errors of

TABLE III
TRANSFER LEARNING PERFORMANCE ON THREE KINDS OF DATASETS

Dataset	Transfer or Not	RMSE	MAE	R2
SMS	No Transferring	55.0727	28.3204	0.8593
	Transferring with Call	50.9684	25.9039	0.8714
	Transferring with Internet	52.7757	25.4138	0.8593
Call	No Transferring	35.4332	16.8691	0.9163
	Transferring with SMS	33.4663	15.7211	0.9240
	Transferring with Internet	30.8529	14.4174	0.9312
Internet	No Transferring	186.1173	111.7783	0.9411
	Transferring with SMS	168.8695	97.8216	0.9511
	Transferring with Call	169.5268	94.3403	0.9503

LR, SVR, LSTM, DenseNet and *STCNet* are approximately lower than 484, 458, 267, 227 and 179, respectively. For the traffic data of this cell, *STCNet* achieves about 21% improvements compared with DenseNet, which is the state-of-the-art prediction method for cellular traffic. Compared with methods that only consider cellular traffic itself, *STCNet* indeed introduces more parameters since the inputs of meta data embedding and cross-domain modeling, but the performance gains are quit significant.

E. Performance of Transfer Learning Between Various Kinds of Cellular Traffic

As described in Fig.3, the correlation coefficients among different kinds of cellular traffic are very high. This indicates the possibility of transferring knowledge from one kind of cellular traffic to another one. So, in this subsection, we report the performance of transfer learning among different kinds of cellular traffic and the obtained results in terms of three evaluation metrics are summarized in Table III.

The "No Transferring" in Table III means the results are achieved using only the single dataset. "Transferring with $\star$" denotes that our results are obtained with the aid of knowledge

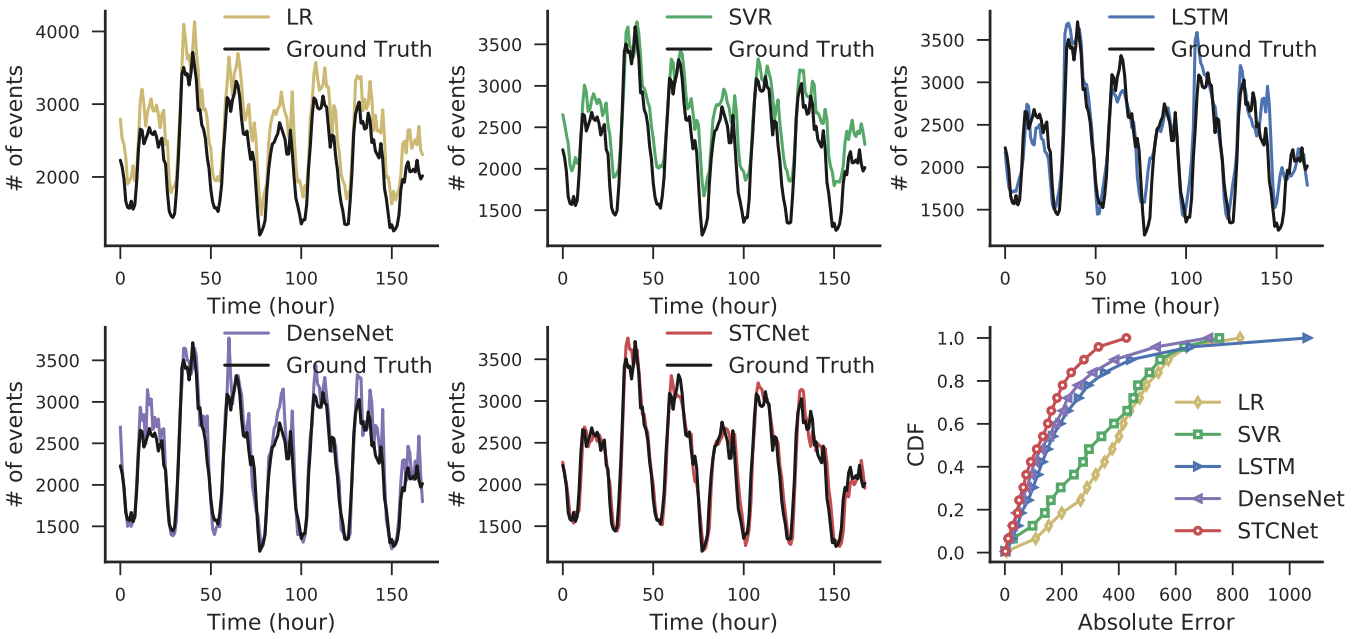


Fig. 9. RMSE comparisons of prediction versus ground truth for all methods on Internet dataset of cell (44, 56).

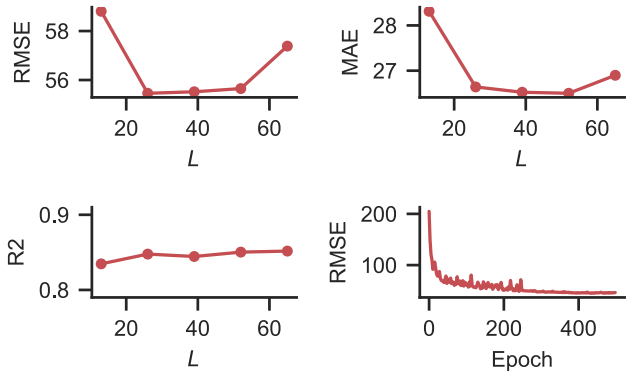


Fig. 10. Impacts of model complexity on performances and the convergence speed illustration.

transferred from the traffic of $\star$. To take the experiments on SMS traffic dataset as an example, when $\star$ represents the Call traffic, the results are obtained with knowledge transferred from Call dataset. That is, the *STCNet* model is trained for SMS traffic dataset but with parameters initialized by those learned from Call dataset.

We can see from Table III that transfer learning can indeed bring performance gains for all the three kinds of cellular traffic. Taking the performance in terms of RMSE as an example, the transfer learning brings about 7.45% and 4.17% improvements when transferring with Call and Internet cellular traffic, respectively. The improvements for Call cellular traffic are 5.55% and 12.9% when transferring with SMS and Internet cellular traffic. Similarly, the improvements for Internet traffic are 9.27% and 8.91% when transferring with SMS and Call cellular traffic. For the other two performance metrics, MAE and R2, the gains can also clearly observed from Table III. The results of Table III validate the effectiveness of transfer learning when performing cellular traffic prediction.

F. Complexity and Convergence Analysis

The computational complexity of *STCNet* is $\mathcal{O}(\sum_{l=1}^{L}(H \cdot W \cdot K_l^2 \cdot F_{l-1} \cdot F_l))$, where H and W are the height and width of the input, K_l and F_l denote the kernel size and filter size of the l-th layer. Since the depth of *STCNet*, L, dominates the complexity and affects the final performance greatly, we investigate the relationship of the parameter complexity and the prediction performance. The results are represented in Fig.10, in which the convergence speed of *STCNet* is also displayed.

Results demonstrate that with the increase of model complexity (L), the RMSE and MAE get better first, and then degrade substantially, while the R2 score performs relatively stable. This is because the representation ability of the model is enhanced as the complexity increases, but after a certain degree, the model is too complex and overfits the data, thus causes performances degradation. For convergence speed, as shown in the last sub-figure of Fig.10, after 250 epochs, the RMSE performance is largely improved and approaches stable, reflecting the effectiveness of our adaptive learning rate strategy. The RMSE performance tends to be stable after 300 epochs, indicating that *STCNet* can converge and the training process is time efficient.

V. Conclusion

This work investigated the intelligent traffic prediction based on deep learning techniques for future cellular networks. To fully characterize various factors (spatial, temporal and external) that affect cellular traffic generation, three kinds of cross-domain datasets, i.e., BSs information, POIs distribution and social activity level, were crawled and their correlations with the cellular traffic were comprehensively analyzed. Based on these datasets, a novel deep neural network architecture, *STCNet*, was proposed to predict the cellular traffic. The ConvLSTM unit is incorporated into *STCNet* to simultaneously capture the spatial and temporal

dependencies of cellular traffic. Various cross-domain datasets were processed as a multi-channel tensor and treated as spatial constraints among different cells to capture the external influencing factors. Besides, the *STCNet* adopted the dense connectivity pattern to ensure maximum information flow between convolution layers, that is, each layer of *STCNet* is connected to every other in a feed-forward fashion. In addition, aiming to model the pattern diversity and similarity of different city areas, a clustering method was proposed to segment the city into different functional zones and a successive inter-cluster transfer learning strategy was put forward to achieve this purpose. Transfer learning between different kinds of cellular traffic was also explored. Experiments have been conducted using real world cellular traffic datasets and the results have demonstrated the effectiveness of our proposed *STCNet* model. The prediction performances on various evaluation metrics have shown the necessary of introducing of cross-domain datasets to enhance traffic prediction. Experimental results have also revealed that deep transfer learning can well capture the similarities between different kinds of cellular traffic thus has great potentials for intelligent traffic prediction.

One possible extension of this work would be exploring other types of cross-domain datasets for cellular traffic prediction and transferring knowledge between different cities. Besides, designing more effective loss functions to deal with the inherent drawbacks of l_p loss would be an interesting direction of future research. Furthermore, introducing noise and sparsity to the dataset and designing robust prediction algorithms based on transfer learning are also worth exploring. The source code of this work is available at *https://github.com/zctzzy/STCNet.*

REFERENCES

[1] C. Lynch, "Big data: How do your data grow?" *Nature*, vol. 455, no. 7209, pp. 28–29, Sep. 2008.

[2] S. J. Walker, "Big data: A revolution that will transform how we live, work, and think," *Int. J. Advertising*, vol. 33, no. 1, pp. 181–183, Jan. 2014.

[3] Y. Huang, J. Tan, and Y.-C. Liang, "Wireless big data: Transforming heterogeneous networks to smart networks," *J. Commun. Inf. Netw.*, vol. 2, no. 1, pp. 19–32, Mar. 2017.

[4] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.

[5] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 882–890.

[6] F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, "Understanding mobile traffic patterns of large scale cellular towers in urban environment," *IEEE/ACM Trans. Netw.*, vol. 25, no. 2, pp. 1147–1161, Apr. 2017.

[7] VNI Cisco, "Global mobile data traffic forecast update, 2016–2021," White Paper 1454457600805266, Mar. 2017.

[8] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[11] P. K. Agyapong, 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.

[12] X. X. Wang Li and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.

[13] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[14] R. Li *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[15] ITUNews. *ITU Launches New Focus Group to Study Machine Learning in 5G Systems*. Accessed: Jun. 29, 2018. [Online]. Available: https://news.itu.int/itu-launches-new-focus-group-study-machine-learning-5g-systems

[16] C. Yao, C. Yang, and C.-L. I, "Data-driven resource allocation with traffic load prediction," *J. Commun. Inf. Netw.*, vol. 2, no. 1, pp. 52–65, Mar. 2017.

[17] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.

[18] H. Wang, J. Ding, Y. Li, P. Hui, J. Yuan, and D. Jin, "Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks," in *Proc. 7th Int. Workshop Hot Topics Planet-Scale Mobile Comput. Online Social Netw. (HOTPOST)*, New York, NY, USA, Jun. 2015, pp. 19–24.

[19] R. Li, Z. Zhao, J. Zheng, C. Mei, Y. Cai, and H. Zhang, "The learning and prediction of application-level traffic data in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3899–3912, Jun. 2017.

[20] Y. Shu, M. Yu, J. Liu, and O. W. W. Yang, "Wireless traffic modeling and prediction using seasonal ARIMA models," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, May 2003, pp. 1675–1679.

[21] B. Zhou, D. He, and Z. Sun, "Traffic predictability based on ARIMA/GARCH model," in *Proc. 2nd Conf. Next Gener. Internet Design Eng.*, Apr. 2006, pp. 200–207.

[22] F. Xu *et al.*, "Big data driven mobile traffic understanding and forecasting: A time series approach," *Services Comput.*, vol. 9, no. 5, pp. 796–805, Sep. 2016.

[23] 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.*, Jun. 2015, pp. 3585–3591.

[24] R. Li, Z. Zhao, X. Zhou, J. Palicot, and H. Zhang, "The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 234–240, Jun. 2014.

[25] J. G. D. Gooijer and R. J. Hyndman, "25 years of time series forecasting," *Int. J. Forecasting*, vol. 22, no. 3, pp. 443–473, 2006.

[26] J. Wang, B. Gong, H. Liu, and S. Li, "Multidisciplinary approaches to artificial swarm intelligence for heterogeneous computing and cloud scheduling," *Appl. Intell.*, vol. 43, no. 3, pp. 662–675, Oct. 2015.

[27] H. Zhang, L. Cao, and S. Gao, "A locality correlation preserving support vector machine," *Pattern Recognit.*, vol. 47, no. 9, pp. 3168–3178, Sep. 2014.

[28] J. Bai and H. Liu, "Multi-objective artificial bee algorithm based on decomposition by PBI method," *Appl. Intell.*, vol. 45, no. 4, pp. 976–991, Dec. 2016.

[29] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Malaysia, Asia: Pearson, 2016.

[30] 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)*, San Francisco, CA, USA, Mar. 2017, pp. 1–5.

[31] J. Wang *et al.*, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Atlanta, GA, USA, May 2017, pp. 1–9.

[32] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1656–1659, Aug. 2018.

[33] 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. (Mobihoc)*, New York, NY, USA, Jun. 2018, pp. 231–240.

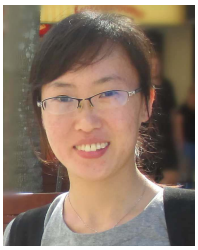
[34] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Short term traffic forecasting using the local linear regression model," in *Proc. 82nd Annu. Meet. Transp. Res. Board*, Washington, DC, USA, Jan. 2003, pp. 1–30.

[35] N. I. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 24–38, May 2009.

[36] 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.
[37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[38] G. Barlacchi *et al.*, "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," *Sci. Data*, vol. 2, Oct. 2015, Art. no. 150055.
[39] OpenCellID. *The World's Largest Open Database of Cell Towers*. Accessed: Jun. 29, 2018. [Online]. Available: https://opencellid.org/
[40] Google Inc. *Google Places API*. Accessed: Jun. 29, 2018. [Online]. Available: https://developers.google.com/places/
[41] Dandelion. *Dandelion API*. Accessed: Jun. 29, 2018. [Online]. Available: https://dandelion.eu
[42] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
[43] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

Chuanting Zhang (S'15) received the B.S. and M.S. degrees from the School of Information Engineering, Inner Mongolia University of Science and Technology, China, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree with the School of Information Science and Engineering, Shandong University, China. His research interests include wireless big data analysis, deep learning, computational social science, and image processing.

Haixia Zhang (M'08–SM'11) received the B.E. degree from the Department of Communication and Information Engineering, Guilin University of Electronic Technology, China, in 2001, and the M.Eng. and Ph.D. degrees in communication and information systems from the School of Information Science and Engineering, Shandong University, China, in 2004 and 2008, respectively. From 2006 to 2008, she was with the Institute for Circuit and Signal Processing, Munich University of Technology, as an Academic Assistant. From 2016 to 2017, she was a Visiting Professor with the University of Florida, USA. She is currently a Full Professor with Shandong University. Her current research interests include cognitive radio systems, cooperative (relay) communications, resource management, space–time process techniques, mobile edge computing, and smart communication technologies. She has been actively participating in many academic events. She has been serving as a TPC member and the session chair. She is giving invited talks in conferences. She is serving as a reviewer for numerous journals. She is an Associate Editor of the *International Journal of Communication Systems*.



Jingping Qiao (S'15) received the Ph.D. degree from Shandong University, China, in 2018, and the B.E. degree from the School of Information Engineering, Inner Mongolia University of Science and Technology, China, in 2012. She is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University, China. Her research interests include physical layer security, cooperative (relay) communications, signal processing for communications, and simultaneous wireless information and power transfer.

Dongfeng Yuan (SM'01) received the M.S. degree from the Department of Electrical Engineering, Shandong University, China, in 1988, and the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University, China, in 2000. From 1993 to 1994, he was with the Electrical and Computer Department, University of Calgary, AB, Canada. He was with the Department of Electrical Engineering, University of Erlangen, Germany, from 1998 to 1999, with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA, from 2001 to 2002, with the Department of Electrical Engineering, Munich University of Technology, Germany, in 2005, and with the Department of Electrical Engineering Heriot-Watt University, U.K., in 2006. He is currently a Full Professor with the School of Information Science and Engineering, Shandong University. His current research interests include cognitive radio systems, cooperative (relay) communications, and data processing for communications.

Minggao Zhang received the B.Sc. degree in mathematics from Wuhan University, China, in 1962. He was a Group Leader of the Radio Transmission Research Group, ITU-R. He is currently a Distinguished Professor with the School of Information Science and Engineering, Shandong University, and a Senior Engineer with the 22nd Research Institute, China Electronics Technology Corporation. He has been an Academician with the Chinese Academy of Engineering since 1999 and is currently a fellow of the China Institute of Communications. He has been involved in the research of radio propagation for decades. Many of his proposals have been adopted by international standardization organizations, including CCIR P.617-1, ITU-R P.680-3, ITU-R P.531-5, ITUR P.529-2, and ITU-R P.676-3. He has received seven national and ministerial Science and Technology Awards in China.