

Traffic Prediction on Communication Network based on Spatial-Temporal Information

Yue MA*, Bo PENG*, Mingjun MA**, Yifei WANG*, Ding XIAO**

*Information and Telecommunication Company, State Grid Jibei Electric Power Company Limited, Beijing China

**School of Computer Science, Beijing University of Posts and Telecommunications, Beijing China

mayue@sina.com, 25811247@qq.com, mamingjun@bupt.edu.cn, wangyf341@163.com, dxiao@bupt.edu.cn

Abstract— With the development of the communication and computer science technology, the traffic prediction of the communication network has attracted more and more interests from the scholars, meanwhile, it is also a significant problem in the real world. A good prediction result can monitor the diversification of the traffic volume and give an early alarm of the outlier. A key challenge of the traffic prediction in the communication network is that how to combine the spatial-temporal information together to make full use of the data. In this paper, we get two observations: (1) At the same timestamp, different square has different traffic volume, while at the same square, different timestamp also has different traffic volume. (2) There exists some periodicity in the traffic volume data along time. To address the challenges we mentioned before, we propose a novel Multi-Channel Spatial-Temporal framework (MCST) to model the spatial-temporal information. The three-channel CNN can mine the spatial information and enrich the temporal information, while the LSTM can model the temporal information. MCST can fuse the spatial-temporal information together to achieve the goal of giving a better prediction. Experiments on the public dataset of the communication network in Milan verify the effectiveness of the proposed model.

Keywords— Traffic prediction, Communication network, Deep learning, Spatial information, Temporal information

I. INTRODUCTION

A communication network is a system that interconnects geographically dispersed user terminal devices together using a switching device and a transmission device. The communication network is nearly everywhere. Forecasting the traffic volume in it can help the managers master the condition of the network and will give an instruction for further decision. With the accumulation of the time, there will exist a sequence data representing the history traffic volume. Different devices or stations in different locations will have various traffic volume. Using the history time series and the location information to predict the future volume in a certain place called spatial-temporal prediction.

The spatial-temporal prediction is a popular problem and has attracted a lot of interests in recent years. This problem is defined as predicting the future value based on the given series data of history and the spatial information. There are many spatial-temporal prediction cases in our daily life, like the road traffic volume prediction, the network traffic prediction and so

on. In this paper, we focus on the traffic prediction of the communication network. It is vital for managing the network and controlling it while congestion happens. However, in the spatial-temporal prediction problem, most of the existing work haven't considered to model the interactions between the spatial information and temporal information together.

To address the mentioned challenges, we propose a novel Multi-Channel Spatial-Temporal framework (MCST) for traffic prediction of communication network. MCST is a deep learning model based on the spatial-temporal information via CNN and LSTM. The CNN structure in MCST is three-channel, which aims to handle the spatial information by the CNN and add the previous important information through the three channels. The LSTM is used to handle the temporal dependency. We evaluate the proposed framework on the public dataset of Milan about the traffic of the communication network in 2013. The comparisons with other baselines show the effectiveness of our method. Our contributions are summarized below:

- We propose a three-channel CNN to solve the spatial dependency and enrich the information of the current region.
- We use the output of the CNN as the input of the LSTM to solve the interactions between the spatial information and the temporal information. Also, the LSTM can model the sequential dependency.
- We conduct experiments on the real datasets with varying proportion of training set. The results show that our method have the best performance among the baselines.

II. RELATED WORK

In a typical traffic prediction problem of the communication network, a series of history traffic data with its corresponding region will be given and one needs to predict the future traffic of a certain region at a certain time slice. A number of studies have investigated the traffic prediction of the communication network for decades. HA and ARIMA [1] are two traditional methods. In time series area, [2] analyses 9000 stations in the communication network and divides the traffic into predictable traffic and unpredictable traffic, which verifies the temporal relevance on the traffic data. [3] uses ARIMA to predict the traffic of the communication network. These early methods [4]

and [5] can study the future traffic value for a certain time series but fail to model the complex interactions between spatial information and temporal information. In recent years, deep learning has begun to rise and achieved enormous success in many tasks, like computer vision and natural language processing [6]. The success of the deep learning inspires many scholars to use this new technology to solve the traffic prediction problem. As for the typical deep neural network, many studies used several fully connected layers stacking together to fuse the multi-source data [7], including the traffic volume data, the weather condition data and so on. Some studies used convolutional structure to capture the relationship among the regions where the stations locate [8][9]. While others considered the prediction problem from the time aspect using the recurrent neural network to model the temporal dependency [10][11][12]. Although the above studies have taken the spatial or temporal information into account, the existing methods have a few limitations. First, the traffic prediction relies on both spatial aspect and temporal aspect. Considering both aspects and their interactions are significant. Second, different timestamps have different influence on the current timestamps. We observed that the recent time slots and the same slot of the current slot in yesterday have more affects. In summary, our proposed framework can explicitly model the spatial dependency and temporal dependency together and the three-channel structure can enrich the information to make the prediction have a better performance.

III. NOTATIONS AND PROBLEM DESCRIPTION

Traffic prediction of the communication network aims to predict the traffic volume of a certain region given the history traffic volume data in corresponding region. We split the whole area into an $m \times n$ grid map with N regions in total ($N = m \times n$) and use $\{1, 2, \dots, N\}$ to denote each of them. The division of the area can be seen in Figure 1 and each small grid represents a region. We split the whole time period into t equal-length continuous time periods, and each small period is 10 minutes. As for the format shown in Figure 1, we can see that the traffic of a random time can be represented as a matrix. For example, if X_i is a matrix of the traffic volume at time i in the whole $m \times n$ area, then $X_i^{(a,b)}$ represents the traffic volume at time i in the region (a,b) . Given a traffic volume sequence $\mathbf{X} = (X_{t-1}, \dots, X_{i-1}, X_i)$, we need to predict the volume at time $i + 1$.

IV. MULTI-CHANNEL SPATIAL-TEMPORAL FRAMEWORK

In this section, we will present the proposed MCST framework in detail. We will first introduce the foundation of the design and the overall framework of MCST. After that we will discuss the design of multi-time convolutional neural networks (MTCNN), and the LSTM part, respectively.

A. The foundation of the framework

The traffic prediction is a typical problem of spatial-temporal data forecasting. The traffic data is recorded at fixed points in time and fixed location in space and related to both the spatial aspect and temporal aspect. Figure 2 shows the spatial-temporal correlations of traffic volume data. We can see that the traffic at the same time is different for different square and the traffic is different at the same square for different time. This observation can be clearly seen from the bar in Figure 3. Figure 3(a) shows the traffic volume in the same square with different time and Figure 3(b) shows the traffic volume at the same time with different squares.

$(m,1)$	$(m,2)$	$\dots$	(m,n)
$\dots$	$\dots$	$\dots$	$\dots$
$(2,1)$	$\dots$	$\dots$	$(2,n)$
$(1,1)$	$(1,2)$	$\dots$	$(1,n)$

Figure 1. Division of the area.

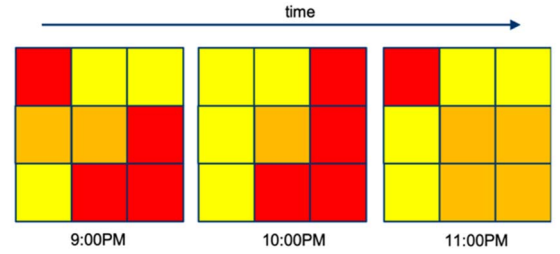


Figure 2. The spatial-temporal correlations of traffic volume data.

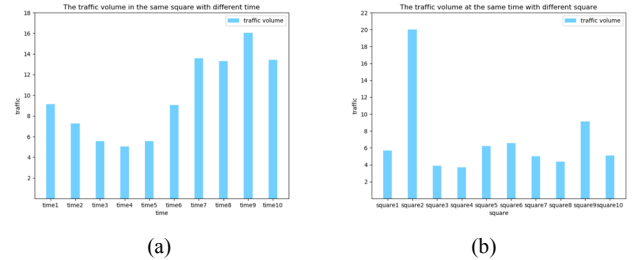


Figure 3. An example of the spatial-temporal changes. (a) The traffic volume of the same square with different square. (b) The traffic volume of the same time with different square.

B. MCST framework

The framework of MCST is illustrated in Figure 4. We will introduce the architecture of MCST from bottom to up. As shown in Figure 4, MCST takes the history traffic volume as input. The output of the framework is the prediction value. For each piece of the traffic data, a specially designed MTCNN is used to process the history traffic volume and the spatial information of the station. MTCNN is an extension of

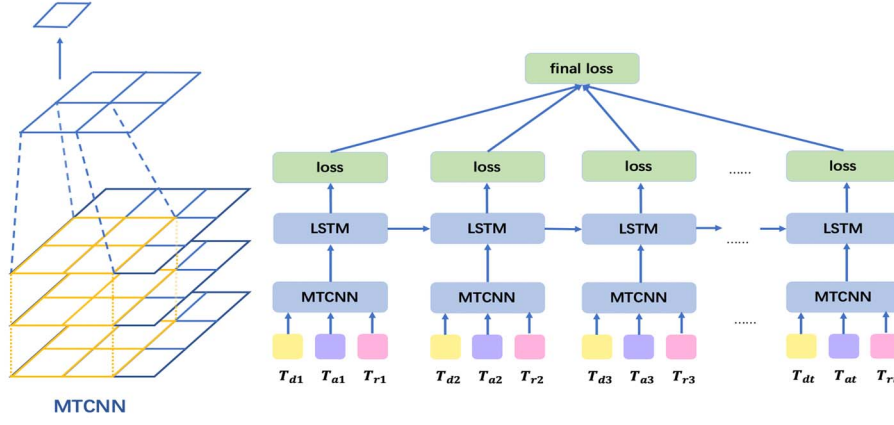


Figure 4. The overall of the framework

traditional CNN that allows more information can be added into the model. By MTCNN, we can get a value, which is a fusion of the traffic volume of the surrounding locations and the history volume. The output of the MTCNN is the input of LSTM. LSTM is used to model the temporal dependency along the time. With the length of the time sequence accumulation, the performance will be better and better. We will detail the process of prediction in the following section.

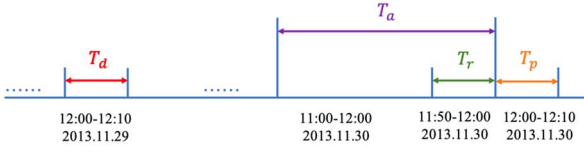


Figure 5. The segment of the time.

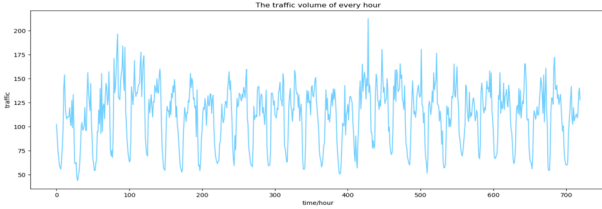


Figure 6. The periodicity of the traffic volume data.

C. Multi-Time CNN (MTCNN)

The input of the MTCNN is three matrices and each matrix represent a time point or a period. The elements of each matrix are the traffic volume of corresponding square. The details are shown in Figure 5. As we can see, the three layers are T_d , T_a and T_r . And T_p is the time period that we will predict. T_d is same period as T_p in yesterday. The reason why we choose T_d is that we can observe the traffic volume has periodicity of one day. This observation is shown in Figure 6. T_a is calculated as an average of the traffic volume sum of last one hour. The aim is to smooth the current volume and add more information. T_r is the value of the last 10 minutes, which is the closest time period of the current period. The architecture of MTCNN is

illustrated in left part of Figure 4. We align and stack the three matrices as:

$$Q = [[T_{d1} T_{a1} T_{r1}] [T_{d2} T_{a2} T_{r2}] \dots [T_{dn} T_{an} T_{rn}]] \in R^{d \times d \times 3}, \quad (1)$$

where $d \times d$ is the size of the matrix of each layer and n is the length of the sequence. After getting the multi-time input Q , we will apply multiple filters g to extract special local patterns. The local activation of sub-matrix Q_t with respect to g can be regarded as:

$$k_t = f(g * Q_t + b), \quad (2)$$

where $*$ is the convolution operator and b is a bias. And we use max-pooling operation on the output feature map to choose the largest feature to represent this corresponding step:

$$\tilde{k}_t = \max\{k_{t_1}, k_{t_2}, \dots, k_{t_m}\}. \quad (3)$$

The output of MTCNN is a scalar value which represents the current step in account of the periodicity and the spatial information. The output value will be served as the input of LSTM to capture the temporal dependency.

D. LSTM part for temporal dependency

LSTM is famous for its ability to capturing the temporal dependency. The output of each MTCNN cell together forms a time sequence. LSTM is controlled by the gate mechanism. The first step is to use the forget gate to calculate the number, which indicates how much will the cell state will be abandoned. The forget gate is gained by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, \tilde{k}_t] + b_f), \quad (4)$$

where h_{t-1} is the hidden state of last time step and $\tilde{k}_t$ is the input of the current time. W_f and b_f are parameters.

The input gate of the LSTM is calculated as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, \tilde{k}_t] + b_i), \quad (5)$$

After that we should calculate the current memory as:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, \tilde{k}_t] + b_c). \quad (6)$$

The cell state of the current state is calculated as:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \quad (7)$$

The forget gate will control how many things does this model should forget from history. Finally, the output gate is written as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, \tilde{k}_t] + b_o). \quad (8)$$

And the output of each LSTM cell is:

$$h_t = o_t * \tanh(C_t). \quad (9)$$

The output of each LSTM cell is the prediction value of every time step. With the sequence becomes longer and longer, the performance of the LSTM will be better and better.

E. Loss function

For each time step i , there will exist two values. One is the actual value of the current time y_{a_i} and the other is the output of the prediction value of the same time y_{p_i} . The aim of the model is to let the prediction value closest to the actual value. In this work, the loss function is defined as:

$$L = \sum_{i=1}^N (y_{a_i} - y_{p_i})^2, \quad (10)$$

where i is the current step and N is the total length of the sequence. For training the model, we should adjust the parameters to minimize the loss function.

V. EXPERIMENTS

A. Datasets

We evaluate our proposed method on a public real-word dataset from Milan, Italy [13]. The dataset contains many aspects, such as communication traffic volume, the social network of the local residents and so on. We only extract the data related to our work. The datasets we extracted contain 4320 communication traffic records of each square, from 11/01/2013 to 11/30/2013.

B. Pre-processing

The whole Milan city is split into 100×100 regions. The length of each time interval is set as 10 minutes, so there are 4320 records in one day. We first select the column we need and each of the record is in the format of “square ID, timestamp, traffic”. In our experiments, we set the proportion of the training set varies from 50% to 80%.

C. Evaluation and baselines

In our experiments, RMSE is used for evaluation. RMSE means rooted mean square error and is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{a_i} - y_{p_i})^2}{N}}. \quad (11)$$

We compare MCST with traditional models and deep learning models, including:

- Historical average (HA): As for the prediction of the current time step, we calculate the average of the last 6 values as prediction result. This method only concerns the history nearby.
- Autoregressive integrated moving average (ARIMA): It is a traditional method for time sequence prediction. In our experiments, we use $p = 5$, $d = 1$ and $q = 0$.
- Decision Tree Regressor (DTR): It can be used as a regression method. We use the last 6 values as the feature and the current value as the target to train the model. The number of the maximum leaf nodes of the model is 5 and the number of the minimum samples leaf is 2.
- Long-short term memory network (LSTM): LSTM is a deep learning model for time sequence. It can

remember the important things and ignore the unimportant things as time goes on. The number of the units is 10 and the look back is set as 1. And a dense layer is added to convert the 10 units to a scalar value.

- Bi-directional LSTM (Bi-LSTM): Bi-LSTM can capture the information from the begin to the end and also the end to the begin. The parameters set is the same with LSTM.
- Convolutional neural network (CNN): We select 3×3 grids traffic volume data of the last time step as input and the current time center value of the grid as the target to train the model. We set 6 filters and use the stride as 1. The activation function is ReLU.

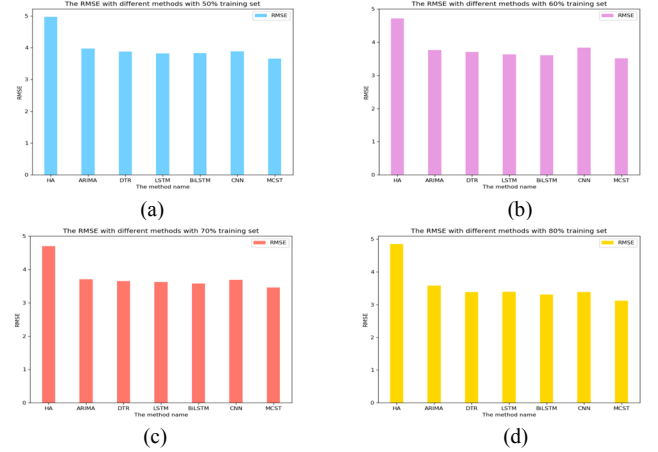


Figure 7. The results of the experiments with different size of the training set. (a) The RMSE with different methods with 50% training set. (b) The RMSE with different methods with 60% training set. (c) The RMSE with different methods with 70% training set. (d) The RMSE with different methods with 80% training set.

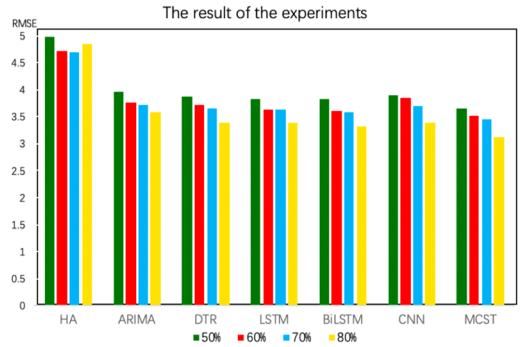


Figure 8. The overall results of the experiments.

D. Results and Analysis

Table 1 shows the prediction performance of our proposed method as compared to all other competing methods. We run each baseline 5 times and report the average results of each baseline. From the table we can get the conclusion that our proposed MCST significantly outperforms all competing baselines by achieving the lowest RMSE at the same training set proportion on the datasets. Specially, the traditional time-series prediction methods (HA and ARIMA) do not perform well, because they only rely on historical records and ignore the spatial information. For DTR, this method is better than HA and

ARIMA, but sometimes can't beat the deep learning method. The deep learning methods are almost better than the traditional method. CNN concerns the spatial information, while LSTM and Bi-LSTM concern the temporal information. Among the deep learning model, CNN performs the worst. The reason may be the traffic data relies on temporal information more than spatial information. As for our method, it concerns both the spatial and temporal information and the interaction between them. Periodicity is also added into the model, so the performance is the best among all the baselines at the same training set proportion. From the perspective of the proportion of the training set, Figure 7(a)(b)(c)(d) show us the RMSE with different size of training set. No matter what the size is, MCST always performs the best. From the perspective of the methods, in Figure 8, we can find that with the proportion of the training set increase, the RMSE is getting smaller and smaller, except HA. To emphasize, HA is only a method to calculate the average of the history data and the proportion means the data we use to get the final result. So, the performance is not stable.

TABLE 1. COMPARISON OF DIFFERENT METHODS

Method	RMSE(50%)	RMSE(60%)	RMSE(70%)	RMSE(80%)
HA	4.97	4.72	4.70	4.85
ARIMA	3.97	3.76	3.71	3.58
DTR	3.88	3.71	3.65	3.38
LSTM	3.82	3.63	3.62	3.39
Bi-LSTM	3.83	3.61	3.58	3.31
CNN	3.89	3.84	3.69	3.38
MCST	3.66	3.51	3.46	3.12

E. Case Study

For a better understanding of the performance, we will select 400 time steps and use MCST framework to predict the traffic volume. The visualization result can be seen in Figure 9. In the figure, we can see the red line is the prediction value and the green line is the actual value. Our model can predict the trend of the traffic volume and give a good performance.

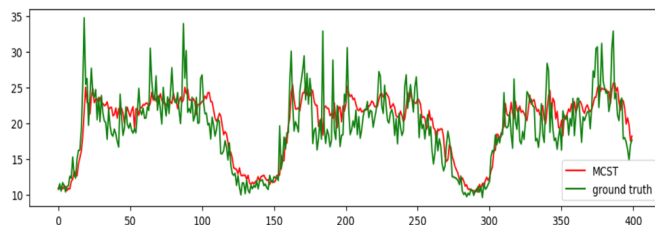


Figure 9. The prediction results of 400 time steps.

VI. CONCLUSIONS AND DISCUSSION

In this paper, we propose a novel Multi-Channel Spatial-Temporal (MCST) model for traffic prediction in communication network. Our method combines the spatial information and the temporal information together. The method can enrich the input and also considers the periodicity of the time sequence. The evaluations on the public dataset show that the proposed model has the best performance among the baselines. In the future, we decide to introduce the attention mechanism to the model and make the model more explainable.

ACKNOWLEDGMENT

This paper is funded by the State Grid Jibei Power Co. Ltd. Science and Technology Project "Research on Data Fusion and Visualization Key Technologies Based on Full-service Unified Data Center" (52018E18006N).

REFERENCES

- [1] Williams, B. M., and Hoel, L. A. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering* 129(6):664–672.
- [2] Xu F , Lin Y , Huang J , et al. Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach[J]. *IEEE Transactions on Services Computing*, 2017, 9(5):796-805.
- [3] SHU, Yantai, YU, et al. Wireless traffic modeling and prediction using seasonal ARIMA models[C]// *IEEE International Conference on Communications*. IEEE, 2003.
- [4] Xiaolong L I , Pan G , Zhaohui W U , et al. Prediction of urban human mobility using large-scale taxi traces and its applications[J]. *Frontiers of Computer Science*, 2012, 6(1):111-121.
- [5] Moreira-Matias L , Gama J , Ferreira M , et al. Predicting Taxi-Passenger Demand Using Streaming Data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14(3):1393-1402.
- [6] Lecun Y, Bengio Y, Hinton G. Deep learning.[J]. 2015, 521(7553):436.
- [7] Wang, D.; Cao, W.; Li, J.; and Ye, J. 2017. DeepSD: Supply-demand prediction for online car-hailing services using deep neural networks. In *ICDE*, 243–254. IEEE.
- [8] Zhang J , Zheng Y , Qi D . Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction[J]. 2016.
- [9] Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *International Joint Conference on Artificial Intelligence*, 3634–3640.
- [10] Yu, R.; Li, Y.; Demiryurek, U.; Shahabi, C.; and Liu, Y. 2017. Deep learning: A generic approach for extreme condition traffic forecasting. In *Proceedings of SIAM International Conference on Data Mining*.
- [11] Cui, Z.; Ke, R.; and Wang, Y. 2016. Deep stacked bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. In *ACM SIGKDD Workshop on Urban Computing*.
- [12] OLIVEIRA T P, BARBAR J S, SOARES A S. Computer Network Traffic Prediction: A Comparison between Traditional and Deep Learning Neural Networks [J]. *International Journal of Big Data Intelligence*, 2016, 3(1): 28-37
- [13] Barlacchi G , De Nadai M , Larcher R , et al. A multi-source dataset of urban life in the city of Milan and the Province of Trentino[J]. *Scientific Data*, 2015, 2:150055.



Yue Ma (1977.12--). Yue is the deputy general manager and senior engineer of Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. Yue received his bachelor degree at North China Electric Power University. Yue's research interest is communication network data analysis.



Bo Peng (1982.7--). She is a senior engineer at the Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. Her research interest is communication network data analysis.



Mingjun Ma (1996.1--). She is a Master Candidate at Beijing University of Posts and Telecommunications. Her research interest are deep learning and traffic prediction.



Yifei Wang (1988.4--). She is an engineer at the Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. Her research interest is communication network data analysis.



Ding Xiao (1966.4--). He is a lecture of Beijing University of Posts and Telecommunications. His research interests are software engineering and artificial intelligence.