IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 June 2022, accepted 11 July 2022, date of publication 19 July 2022, date of current version 17 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3192039

RESEARCH ARTICLE

A Large-Scale Wireless Cell Long-Term Daily-Granularity Forecasting Method

WEI FANG[1], YUN CHEN[2], NING PAN[1], AND BIN RAN[3]
[1]School of Automation, Wuhan University of Technology, Wuhan 430070, China
[2]School of Management, Wuhan University of Technology, Wuhan 430070, China
[3]Department of Civil & Environmental Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA

Corresponding author: Yun Chen (chenyun135@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 41971342 and Grant 17BGL230.

ABSTRACT Optimizing and managing wireless communication network, including improving the utilization of network resources, energy efficiency, automatically carrying out wireless network planning and network construction, is very important to the communication service providers (CSPs). Key performance indicators (KPIs) forecasting for wireless cells, especially the long-term forecasting task, plays a key role in wireless network planning and construction. A new adaptive combination forecasting method is proposed in this paper. The adaptive combination forecasting method has been verified by a real large-scale wireless network dataset which contains thousands of wireless cells and corresponding daily KPIs. After a series steps such as dataset analysis, and Auto-encoder algorithm, K-means algorithm and time series forecasting algorithms, we can obtain the prediction model, then compare its symmetric mean absolute percentage error (SAMPE) value with Holt exponential smoothing, Comb method and Theta method.Experimental results have demonstrated that the proposed method has a better performance, especially in the medium and long term forecasting scenario in terms of symmetric mean absolute percentage error (SMAPE) when compared with some existing methods. It proved that our method can be more suitable for complex wireless communication network environment.

INDEX TERMS Time series forecasting, wireless cell, wireless networks, combination forecasting method.

I. INTRODUCTION

With the rapid development of wireless communication technology, emergence of mobile applications, growing mobile users and the explosive growth of data traffic, the requirements for wireless service quality are continuously improved. All of these put forward new requirements and challenges for CSPs. It is very important to actively optimize and manage the wireless network, effectively improve the utilization of network resources and energy efficiency, and automatically carry out wireless network planning and network construction [1].

The key performance indicators (KPIs) of wireless network, especially KPIs of wireless cells, provide important information and understanding for wireless network insight. Therefore, KPIs forecasting for wireless cells is an important task for CSPs. A precisely forecasting KPIs of wireless cells can monitor the state of network, also can help CSPs reserve network resources, increase network bandwidth, adjust existing network resources in advance. All of these will improve the network efficiency and the stability of the network.

KPIs forecasting for wireless cells is an important branch of time series prediction. Due to the importance, complexity and large-scale characteristics of time series data, many challenges have been brought to the field of time series prediction, especially in the field of wireless communication.

At present, different time series prediction algorithms have been needed to solve the problems in the field of large-scale wireless communication, including the classical time series prediction algorithm, Holt-Winter's exponential smoothing algorithm [2],cloud-edge collaboration [32],joint offloading scheme [33], [34],burst traffic scheduling [35] and seasonal ARIMA model [3] as well as the neural network

The associate editor coordinating the review of this manuscript and approving it for publication was Ding Xu.

model [4], [5], [25]–[28], [30], [31], such as Convolutional Neural

Network [5], [25], [37], Bayesian Neural Network [4], [36], Recurrent neural Network [26], [27], Attention mechanism [27], [28], Convolutional Graph Autoencoder [24], auto-encoder [24], [30].

KPIs forecasting for wireless cells, especially the long-term forecasting task, plays a key role in wireless network planning and construction. To the best of our knowledge, most of the current works focus on short-term forecasting or on a small dataset [2]–[4]. The time span of these datasets is often not long enough to support long-term forecasting [1]–[7]. Some current works use individual time series forecasting method to solve the problem of KPIs forecasting for wireless cells. However, individual methods cannot produce a better performance than the combination methods [8], [9], [29].

Due to the complexity of wireless communication environment and user behavior, individual time series forecasting methods can only be applied in a single scene, which is difficult to adapt to the complex and changeable wireless communication environment. Therefore, it is particularly important to choose different time series forecasting methods for different wireless cell scenarios.

In this paper, in order to overcome the shortcoming of individual time series forecasting methods, which is difficult to adapt to the complex and changeable wireless communication environment, we use combination forecasting methods to improve the robustness adapting to the complex and changeable wireless environment. We proposed an adaptive combination forecasting method based on a real large-scale wireless network dataset which contains thousands of wireless cells and corresponding daily KPIs. The prediction method includes the following parts: Feature extraction of high-dimensional time series automatically by unsupervised learning; automatically classify time series by clustering algorithm; the forward selection algorithm is used to generate the best combination prediction mode for different types of time series. It is very important in the big data scenario, as it cannot manually select forecasting models for each time series dataset.

The rest of this paper is structured as follows. In Section II, we introduce the real large-scale wireless network dataset, together with the main machine learning algorithms including auto-encoder algorithm, K-Means algorithm and traditional time series forecasting algorithms. In Section III, we introduce the scheme design and the details of our method. In Section IV, the results of different forecasting methods are given and compared. Finally, we conclude this paper in Section V.

II. DATASET ANALYSIS METHODS

A. DATASET ANALYSIS

In this section, we provide details about the utilized dataset and present some observations of wireless cells characteristics. The dataset is collected from one of large cities in China, by one of largest CSP. It contains five thousand wireless cells spanning over one year during August 2017 and July 2018 in long-term-evolution (LTE) cellular network. The dataset contains three important KPIs to monitor the the status of cells, which include downlink physical resource block (PRB) utilization, average connections number of radio resource control (RRC) and downlink traffic.

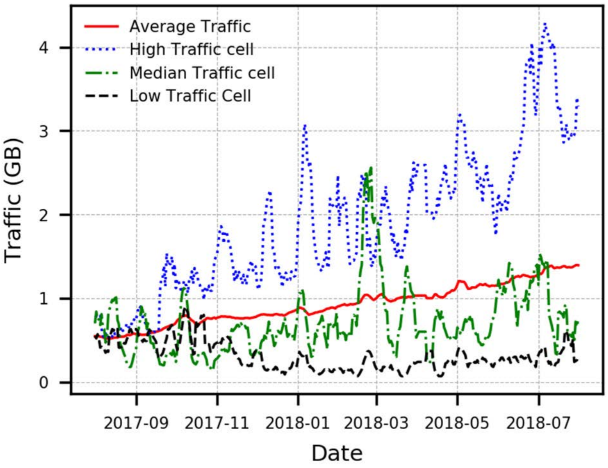


FIGURE 1. Time series for KPI of downlink traffic about average traffic of all wireless cells, high traffic wireless cell, median traffic wireless cell and low traffic wireless cell.

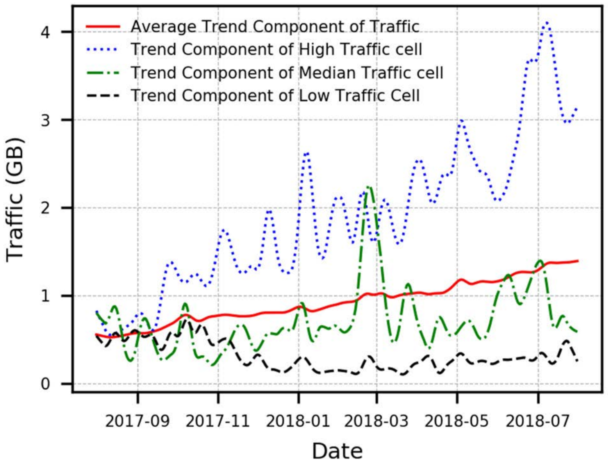


FIGURE 2. Trend component of downlink traffic about average trend component of all wireless cells, trend component of high traffic wireless cell, trend component of median traffic wireless cell and trend component of low traffic wireless cell.

Fig. 1 shows the time series for KPI of downlink traffic about average traffic of all wireless cells, high traffic wireless cell, median traffic wireless cell and low traffic wireless cell.

In order to better understand the characteristics of wireless cells. We used classic time series decomposition method to decompose the downlink traffic time series into trendcomponent, seasonal component and stochastic component [10].

Fig. 2 shows the trend component of downlink traffic about average trend component of all wireless cells, high traffic wireless cell, median traffic wireless cell and traffic wireless cell, respectively.

Fig. 3 shows the seasonal component of downlink traffic about average seasonal component of all wireless cells, high

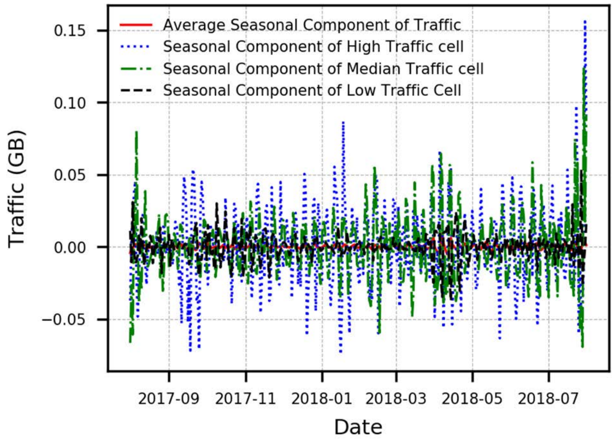


FIGURE 3. Seasonal component of downlink traffic about average seasonal component of all wireless cells, seasonal component of high traffic wireless cell, seasonal component of median traffic wireless cell and seasonal component of low traffic wireless cell.

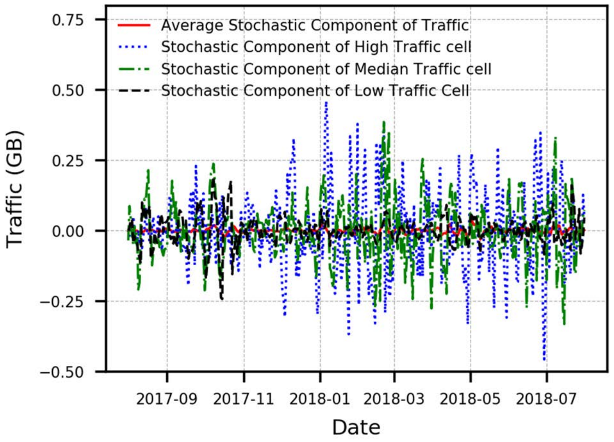


FIGURE 4. Stochastic component of downlink traffic about average stochastic component of all wireless cells, stochastic component of high traffic wireless cell, stochastic component of median traffic wireless cell and stochastic component of low traffic wireless cell.

traffic wireless cell, median traffic wireless cell and low traffic wireless cell, respectively.

Fig. 4 shows the stochastic component of downlink traffic about average stochastic component of all wireless cells, high traffic wireless cell, median traffic wireless cell and low traffic wireless cell, respectively.

It can be seen from Fig. 1 and Fig. 2 that the average traffic of all cells has an obvious growth trend. With the growth of time, the overall traffic is growing upward, but there is a great inconsistency between the traffic change of a single cell and the average traffic change. The trend component of a single cell has great volatility, which brings a great challenge to time series forecasting.

In addition, we can observe that the average traffic and the average trend component of all wireless cells have an obvious growth trend. The local trend of an individual wireless cell fluctuates greatly whether the downlink traffic is high, medium or low, which brings a great challenge for forecasting.

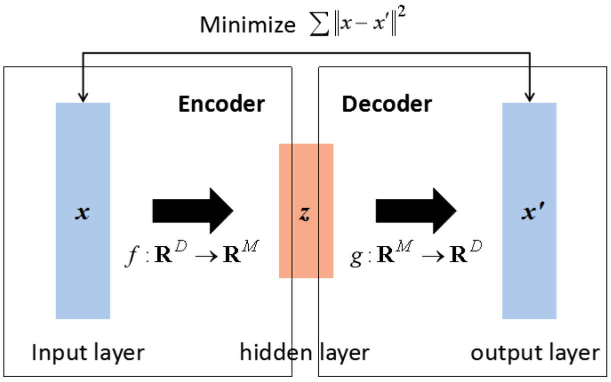

Minimize $\sum\|x-x'\|^2$
Encoder
Decoder
x
$f:\mathbf{R}^D\rightarrow\mathbf{R}^M$
z
$g:\mathbf{R}^M\rightarrow\mathbf{R}^D$
x'
Input layer
hidden layer
output layer

FIGURE 5. The structure of auto-encoder.

Compared with Fig. 1 to Fig. 3, we can find the amplitude of seasonal components is very smaller than the amplitude of time series or trend component. In other words, all of time series have no obvious seasonality.

Compared with Fig. 1, Fig. 2 and Fig. 4, we can find the amplitude of average stochastic component of all wireless cells is smaller than the amplitude of average time series or average trend component. However, the amplitude of stochastic component of individual wireless cell is too high to be ignored. The high fluctuation of local trend and high amplitude of stochastic component bring great challenges for wireless cell forecasting task.

B. AUTO-ENCODER ALGORITHM

Effective characterization of time series plays a key role in subsequent time series prediction. In [30] rough auto-encoder (RAE) and rough denoising auto-encoder (RDAE) are used as hidden layers to extract the features of the time series of wind speed data. In [31] restricted Boltzmann machines and rough set theory are used to capture unsupervised temporal features from wind speed data. Our proposed method use auto-encoder extract the features of the time series of wireless cells.

Auto-encoder is an unsupervised neural network, which can extract the effective coding (or representation) of a dataset [11], [12]. The structure of auto-encoder can be divided into two parts: encoder and decoder, as shown in the Fig 5.

Assuming a set of D-dimensional samples $\mathbf{x}^{(n)} \in \mathbf{R}^D$, $1 \leq n \leq N$, the encoder $f : \mathbf{R}^D \rightarrow \mathbf{R}^M$ maps this set of samples to the feature space to get the code of each sample, $\mathbf{z}^{(n)} \in \mathbf{R}^M$, $1 \leq n \leq N$. Then the decoder $g : \mathbf{R}^M \rightarrow \mathbf{R}^D$ is expected to reconstruct the original samples $x'^{(n)} \in \mathbf{R}^D$, $1 \leq n \leq N$ as accurately as possible. Generally, **M**, the dimension of the feature space, is smaller than **D**, the dimension of the original space. The auto-encoder is equivalent to a dimension reduction or feature extraction method.

The learning objective of auto-encoder is to minimize the reconstruction errors L.

$$L = \sum_{n=1}^{N} \left\| \mathbf{x}^{(n)} - \mathbf{x}'^{(n)} \right\|^2 + \lambda \left\| \mathbf{W} \right\|_F^2 \tag{1}$$

TABLE 1. Time series algorithms used in this paper.

Algorithm Name	Formula
Random Walk (RW)	$y_{t+1} = y_t + \varepsilon_{t+1}$
Seasonal RW	$y_{t+h} = y_t + \varepsilon_{t+h}$
Simple exponential smoothing (SES)	$y_{t+h\|t} = l_t$ $l_t = \alpha y_t + (1-\alpha) l_{t-1}$
Seasonal SES	$y_{t+h\|t} = l_t + s_{t+h-m(k+1)}$ $l_t = \alpha y_t + (1-\alpha) l_{t-1}$ $s_t = \gamma(y_t - l_{t-1}) + (1-\gamma) s_{t-m}$
Holt exponential smoothing (HES)	$y_{t+h\|t} = l_t + h b_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}$
Seasonal LES	$y_{t+h\|t} = l_t + h b_t + s_{t+h-m(k+1)}$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}$ $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma) s_{t-m}$
Damped exponential smoothing (DES)	$y_{t+h\|t} = l_t + \phi_h b_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$
Seasonal DES	$y_{t+h\|t} = l_t + \phi_h b_t + s_{t+h-m(k+1)}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$ $s_t = \gamma(y_t - l_{t-1} - \phi b_{t-1}) + (1-\gamma) s_{t-m}$
Auto Regressive Integrated Moving Average (ARIMA) [14]	$ARIMA\,(p,d,q)$
Seasonal ARIMA [14]	$ARIMA(p,d,q)\,(P,D,Q)_m$
Theta	$y_{t+1\|t} = \frac{1}{2}\left[LR\left(f_t(0)\right) + LEM\left(f_t(2)\right)\right]$
Linear Regression (LR)	$y_{t+1\|t} = g(t+1 \| W^*)$ $W^* = \arg\min_W \sum_{t=1}^{T} \left(y_t - g(t\|W)\right)^2$

TABLE 2. The meaning of mathematical symbols.

Symbol	Explanation
y_t	the real value at time t
$y_{t+h\|t}$	t+h prediction after t
s_t	the periodic change of time t
l_t	level value predicted at time t
b_t	trend change value at time t
$\alpha, \beta, \gamma, \phi$	algorithm parameters, the value range is (0, 1)
m	periodic value
W	vector to be solved in LR
h	time interval
k	Integral part of (h-1)/m
ε_t	noise value at time t

where λ is the regularization coefficient, **W** is the shared weight parameter matrix of encoder and decoder (only transposed), which is called tied weight. The neural network is easier to learn because of the reduction of parameters. In addition, tied weight can also play a role in regularization.

C. K-MEANS ALGORITHM

K-Means is the most commonly used clustering algorithm. K-Means divides the set of n samples into k classes, and the distance from each sample to the center of its class is the smallest [13]. The essence of K-Means is the selection of function from samples to classes。

Assuming a set of n samples, $X = \{x_1, x_2, \ldots, x_n\}$, where each sample is represented by an m-dimensional eigenvector. The number of clusters is less than the number of samples. The samples are divided into k classes $G_1, G_2, \ldots, G_\mathrm{k}$, where

$$G_i \cap G_j = \emptyset, \quad \bigcup_{i=1}^{k} G_i = X \tag{2}$$

And C presents division, and one division corresponds to one clustering result.

The loss function $W(C)$ is defined as the sum of the distances between each sample and its class center.

$$W(C) = \sum_{l=1}^{k} \sum_{C(i)=l} \|x_i - \bar{x}_l\|^2 \tag{3}$$

where $\bar{x}_l = (\bar{x}_{1l}, \bar{x}_{2l}, \cdots, \bar{x}_{ml})^T$ is the center of the l-th class.

K means is to solve the optimization problem.

$$\begin{aligned} C^* &= \arg\min_C W(C) \\ &= \arg\min_C \sum_{l=1}^{k} \sum_{C(i)=l} \|x_i - \bar{x}_l\|^2 \end{aligned} \tag{4}$$

D. TIME SERIES FORECASTING ALGORITHMS

The list of time series prediction algorithms to be used in this paper is shown in Table 1. In the offline stage, according to different clustering results, the forward selection algorithm will select the most suitable combination from the algorithms list. The variables involved in these algorithms are further illustrated in Table 2.

In algorithm Theta, f_t is a function of variable θ, which is expressed as

$$f_t(\theta) = x_1 + (i-1)(x_2 - x_1) + \theta \left(\sum_{i=2}^{t-1} (t-i) y''_{i+1} \right) \tag{5}$$

III. PROPOSED METHOD

In this section, the scheme of our proposed method is described firstly, then the detailed description of offline phase and online phase are presented.

A. SCHEME DESIGN

The scheme of our proposed method is showed in Fig. 6, which includes offline phase and online phase.

The offline phase consists of auto-encoder algorithm, K-Means algorithm and forward selection algorithm. The time series of all wireless cells are used as features and labels

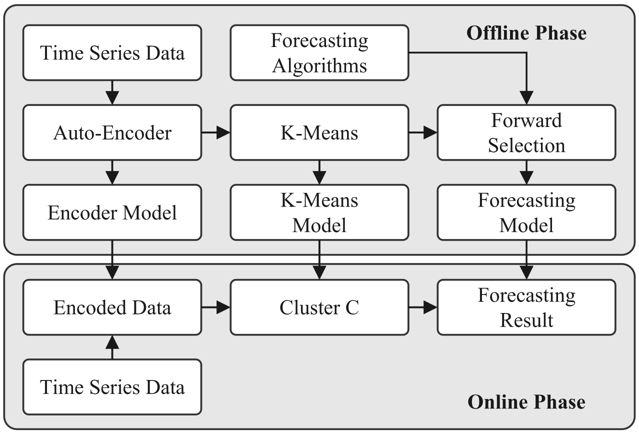


FIGURE 6. The overall flow chart of offline phase and online phase.

for the auto-encoder algorithm, then a encoder model are generated by training dataset. The feature vectors output by encoder model are used as input of K-Means algorithm. Then K-Means algorithm will output the K-Means model and the class corresponding to the time series of each wireless cell. Finally, the forward selection algorithm is used to generate the combination forecasting model for each class.

The offline phase shown in Fig. 6 includes auto-encoder algorithm, K-means algorithm and forward selection algorithm.

The time series data are input into the auto-encoder as a training set, and the auto-encoder parameters are trained to generate the Encoder model, and the encoding feature vector corresponding to the training set is generated to extract the time series features of the wireless cell.

The encoding feature vector of Auto-encoder is input into the K-means algorithm, which generates the K-means clustering model and outputs the clustering results of the training set and the clustering labels of each sample, so as to facilitate the subsequent search for the optimal solution.

The K-means training output clustering results and clustering labels are input into the forward selection algorithm to generate each type of corresponding combination prediction model.

In the offline phase, the Encoder model generated by auto-encoder algorithm, the K-means clustering model generated by K-means algorithm, and the combined forecasting model generated by forward selection algorithm are used as the trained model and the input of the online phase model.

In online phase, the time series to be predicted generates Encoded data by Encoder model. Encoded data is input into the K-means model to generate the corresponding clustering label Clutter C, and the corresponding combination prediction algorithm is found according to the clustering label and Forecasting model, which effectively solves the impact on our work caused by the large amplitude of the random component in a single wireless cell, the high volatility of the local trend and the high amplitude of the random component. Time series data is predicted according to the combination prediction algorithm to generate the final prediction results.

```
Algorithm 1 Forward Selection of Combined Forecasting Model
Input: compressed time series of a cluster class
Output: combined forecasting model of a cluster class
Symbol: $S_0$: All candidate time series forecasting algorithms
        $S$: Candidate time series forecasting algorithms in
        iteration, $S=\{s\}$.
        $A_0$: Selected combined forecasting algorithm set in
        last iteration
        $A$: Selected combined forecasting algorithm set

        $e(S^*)$: forecasting error of combined forecasting
        algorithm set $S^*$

1  Initial: Let $A = \varphi$, $S_0 = \{$12 candidate time series
   forecasting algorithms$\}$.
2  Let $A_0 = A$, $S = S_0$.

3  Select the time series forecasting algorithm $s$ with the
   smallest forecasting error from $S$

4  if $e(A + s) < e(A)$: let $A = A + s$.

5  $S = S - s$.

6  if $S \neq \varphi$: go to step 2
   elif $S = \varphi$ and $A = A_0$: return $A$.
   else: $A_0 = A$, go to step 1.
```

In online phase, the time series pass through the encoder model and K-Means model in sequence to obtain the corresponding forecasting model firstly. Then, this model is used to generate the forecasting results. The details of online and offline phases are presented as follows.

B. METHODS OF OFFLINE PHASE

In the offline phase, we train a encoder model, a K-Means clustering model, and a combined forecasting model for each cluster class.

Firstly, the time series are compressed by auto-encoder, and then the encoder model is generated. Secondly, K-means algorithm is utilized to cluster the compressed time series to generate a k-means clustering model include all the clustering results of the training set and the clustering labels of each sample. Finally, for each cluster tag and cluster result, we utilize the forward selection algorithm to generate a combined prediction model. The details of forward selection algorithm are as follows.

In each iteration, a candidate algorithm from table 1 is added to combined forecasting algorithm set A if the combined algorithm with s has lower forecasting error. The forward selection algorithm stops iterating until the forecasting error is no longer reduced.

C. METHODS OF ONLINE PHASE

In online phase, the vector $Y = (y_1, y_2, \ldots, y_T)^T$ is used to represent the known time series values, where the subscript T is the number of elements in the set. Then, y_{T+i} is defined as the value to be predicted, where i is the i-step ahead for forecasting and $i \geq 1$. The overall flow chart of the online phase is shown in Fig. 6.

Firstly, the time series set Y is input into the encoder trained in the offline stage to obtain the coding vector $\hat{Y} = (y'_1, y'_2, \ldots, y'_n)^T$, which is expressed as:

$$\hat{Y} = f(Y) \tag{6}$$

Secondly, the distance $d_i = \|Y - M_i\|^2$ between $\hat{Y}$ and K-Means cluster centers is calculated, and the clustering center corresponding to the minimum distance is taken as the category c of $\hat{Y}$, that is

$$c = \arg\min_i \{d_i \,|i = 1, \ldots, m\} \tag{7}$$

Finally, the corresponding combined forecasting algorithm set is selected based on category c, which is denoted as C_f, $C_f = \{f_k\}$, where f_k is individual time series forecasting method. For the input time series $Y = (y_1, y_2, \ldots, y_T)^T$, the i-step ahead forecasting result y_{T+i} is denoted as:

$$y_{T+i} = \frac{1}{|C_f|} \sum_{f_k \in C_f} f_{k,T+i}, \quad f_{k,T+i} = f_k(Y; i) \tag{8}$$

where $f_k\,(Y; i)$ is the predictive value of prediction algorithm in i-step ahead of f_k, $i \geq 1$.

IV. EXPERIMENTAL RESULTS

Use In this section, we use the dataset introduced in Section II to verify the effectiveness of our method. We divide the dataset into two parts. One is used in the offline phase and learn the models including auto-encoder model, K-Means model and combining forecasting model, which ranging from 2017-08-01 to 2018-07-03. The other is used to conduct the experiments to verify and validate the accuracy of our proposed method, which ranging from 2017-08-01 to 2018-07-31. The 28 days of data ranging from 2018-07-04 to 2018-07-31 is used as evaluation data to evaluate the forecasting performance metrics.

To evaluate the different time series forecasting methods, we should introduce the performance metrics which can measure their forecasting accuracy. There are many different performance metrics to evaluate the time series forecasting methods. Among them, the most common and widely performance metrics for time series forecasting methods are root mean square error (RMSE) and mean absolute error (MAE). They are defined as follows.

$$RMSE_k = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\hat{y}_{k,t} - y_{k,t}\right)^2} \tag{9}$$

$$RMSE = \frac{1}{K} \sum_{k=1}^{K} RMSE_k \tag{10}$$

$$MAE_k = \frac{1}{T} \sum_{t=1}^{T} \left|\hat{y}_{k,t} - y_{k,t}\right| \tag{11}$$

$$MAE = \frac{1}{K} \sum_{k=1}^{K} MAE_k \tag{12}$$

where T is the total number of samples in a forecasting period, K is the total wireless cell numbers, $y_{k,t}$ is the true value and $\hat{y}_{k,t}$ is the forecasting value.

However, both RMSE and MAE performance metrics are not normalized. They are varied according to the true values. Since the symmetric mean absolute percentage error (SMAPE) performance metric is a normalized value, we use SMAPE to evaluate the different time series forecasting methods in this paper. The SMAPE is defined as follows.

$$SMAPE_k = \frac{100\%}{T} \sum_{t=1}^{T} \frac{\left|\hat{y}_{k,t} - y_{k,t}\right|}{\left(\left|\hat{y}_{k,t}\right| + \left|y_{k,t}\right|\right)/2} \tag{13}$$

$$SMAPE = \frac{1}{K} \sum_{k=1}^{K} SMAPE_k \tag{14}$$

In order to verify the effectiveness of our method, we compare the proposed method with Holt exponential smoothing method [15], [16], Theta method [17], [18] and Comb method which a combination based on the simple arithmetic average of the Simple, Holt and Damped exponential smoothing models [8].

In the training of auto-encoder method, the time series are normalized by subtracting the mean value and dividing by the variance. The mean-absolute-error (MAE) is used as the loss function for auto-encoder method and the dimension of the feature space is 14 dimensions. Auto-encoder input dimension 154 days of day granularity data. The Adam method [19] is used to train the parameters of the auto-encoder. The auto-encoder is realized using Tensorflow package [20]. The 14-dimension vectors are input into the K-Means method to train the K-Means model. The parameter K of K-Means method are determined by silhouette coefficient [21]. The K-Means method is realized using scikit-learn package [22]. The classical time series forecasting methods introduced in Section II are realized using statsmodels package [23]. As the forecasting steps are very long, the forecasting steps are divided into four non-overlapping segments. Each segment has seven forecasting points, and uses Forward selection method to find the best combining forecasting methods.

We highlight to the reader that any results reported in this section should be considered as lower bounds of system performance. The performance of adaptive capacity prediction method can be improved by super parameter optimization of auto-encoder and k-means algorithm, or by introducing better feature extraction algorithm, or by introducing the latest neural network time series prediction algorithm.

The SMAPE performance metrics among four different methods on the three KPIs are presented in Fig. 7.

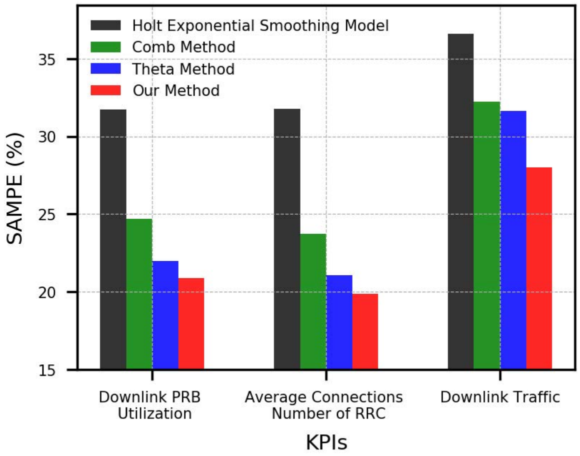


FIGURE 7. The SMAPE performance metrics comparison among four different methods.

For the KPI of downlink PRB utilization, SMAPE of Holt exponential smoothing method is 31.71% whereas it is 24.69% for Comb method and 21.98% for Theta method. Our proposed method has a lowest SMAPE among four different forecasting methods, which is 20.87%. Our method's SMAPE is close to 5.1% more accurate than the Theta method.

For the KPI of average connections number of RRC, SMAPE of Holt exponential smoothing method is 31.71% whereas it is 23.71% for Comb method and 21.06% for Theta method. Our proposed method has a lowest SMAPE among four different forecasting methods, which is 19.86%. Our method's SMAPE is close to 5.7% more accurate than the Theta method.

For the KPI of downlink traffic, SMAPE of Holt exponential smoothing method is 36.60% whereas it is 32.24% for Comb method and 31.62% for Theta method. Our proposed method has a lowest SMAPE among four different forecasting methods, which is 28.03%. Our method's SMAPE is close to 11.4% more accurate than the Theta method.

From Fig. 7, we can obtain observations. The combination methods have a better performance metrics compared with the individual method on different KPIs. Our proposed method has a lowest SMAPE among the four forecasting methods.

The SMAPE performance metrics for KPI of downlink PRB utilization on the different forecasting steps among four different methods are showed in Fig. 8.

For the first 7 forecasting steps, SMAPE of Holt exponential smoothing method is 16.48% whereas it is 14.87% for Comb method, 14.18% for Theta method and 14.32% for our proposed method. Except for Holt exponential smoothing method, the forecasting errors of the other three algorithms are similar.

For the forecasting steps ranging from 8 to 14, SMAPE of Holt exponential smoothing method is 30.65% whereas it is 25.03% for Comb method, 22.44% for Theta method and 20.96% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods.

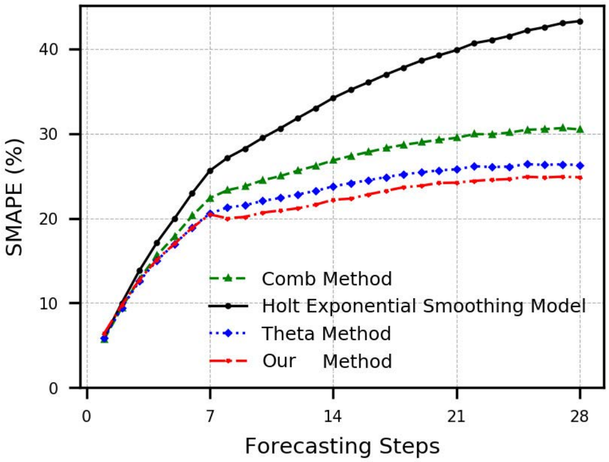


FIGURE 8. The SMAPE performance metrics for KPI of downlink PRB utilization on the different forecasting steps among four different methods.

Our method's SMAPE is close to 6.6% more accurate than the Theta method and 16.3% more accurate than the Comb method.

For the forecasting steps ranging from 15 to 21, SMAPE of Holt exponential smoothing method is 37.68% whereas it is 28.55% for Comb method, 25.08% for Theta method and 23.47% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 6.4% more accurate than the Theta method and 17.8% more accurate than the Comb method.

For the forecasting steps ranging from 22 to 28, SMAPE of Holt exponential smoothing method is 42.05% whereas it is 30.30% for Comb method, 26.24% for Theta method and 24.72% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 5.8% more accurate than the Theta method and 18.4% more accurate than the Comb method.

From Fig. 8, we can find that the farther the forecasting step is, the higher the forecasting error is. Our proposed method has a better performance metrics in the medium and long term forecast compared with the other three forecasting methods for the KPI of downlink PRB utilization.

The SMAPE performance metrics for KPI of average connections number of RRC on the different forecasting steps among four different methods are showed in Fig. 9.

For the first 7 forecasting steps, SMAPE of Holt exponential smoothing method is 15.41% whereas it is 13.38% for Comb method, 12.55% for Theta method and 11.94% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 4.9% more accurate than the Theta method and 10.8% more accurate than the Comb method.

For the forecasting steps ranging from 8 to 14, SMAPE of Holt exponential smoothing method is 29.70% whereas it is 23.09% for Comb method, 20.40% for Theta method and 19.32% for our proposed method. Our proposed method has

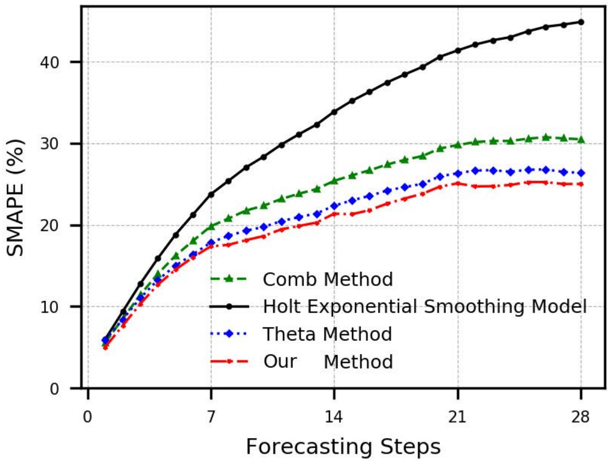


FIGURE 9. The SMAPE performance metrics for KPI of average connections number of RRC on the different forecasting steps among four different methods.

a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 5.3% more accurate than the Theta method and 16.3% more accurate than the Comb method.

For the forecasting steps ranging from 15 to 21, SMAPE of Holt exponential smoothing method is 38.38% whereas it is 27.95% for Comb method, 24.66% for Theta method and 23.21% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 5.9% more accurate than the Theta method and 17.0% more accurate than the Comb method.

For the forecasting steps ranging from 22 to 28, SMAPE of Holt exponential smoothing method is 43.58% whereas it is 30.44% for Comb method, 26.61% for Theta method and 24.98% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 6.1% more accurate than the Theta method and 17.9% more accurate than the Comb method.

From Fig. 9, we can find that the farther the forecasting step is, the higher the forecasting error is. Our proposed method has a better performance metrics, especially in the medium and long term forecast, compared with the other three forecasting methods for the KPI of average connections number of RRC.

The SMAPE performance metrics for KPI of downlink traffic on the different forecasting steps among four different methods are showed in Fig. 10.

For the first 7 forecasting steps, SMAPE of Holt exponential smoothing method is 22.28% whereas it is 20.93% for Comb method, 22.04% for Theta method and 19.73% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 10.5% more accurate than the Theta method and 5.7% more accurate than the Comb method.

For the forecasting steps ranging from 8 to 14, SMAPE of Holt exponential smoothing method is 37.17% whereas it is 33.40% for Comb method, 32.72% for Theta method and 27.64% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 15.5% more accurate than the Theta method and 17.2% more accurate than the Comb method.

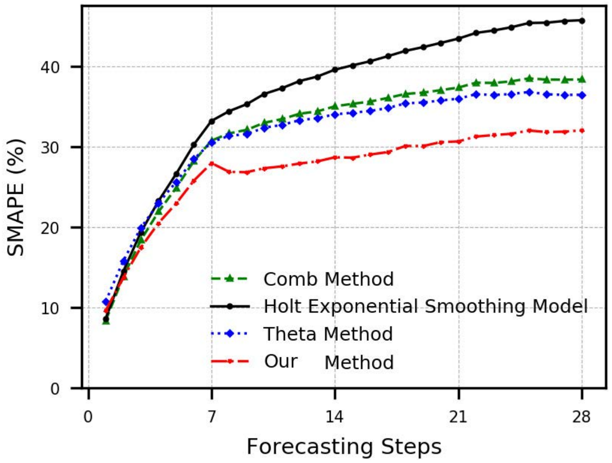


FIGURE 10. The SMAPE performance metrics for KPI of downlink traffic on the different forecasting steps among four different methods.

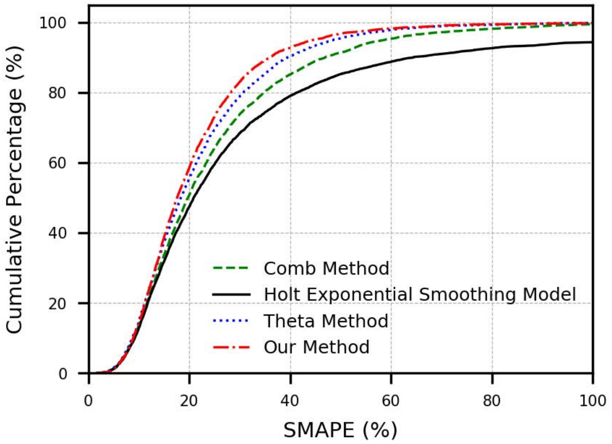


FIGURE 11. The cumulative percentage of SMAPE of wireless cell for KPI of downlink PRB utilization about four different methods.

For the forecasting steps ranging from 15 to 21, SMAPE of Holt exponential smoothing method is 41.83% whereas it is 36.40% for Comb method, 35.18% for Theta method and 29.79% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 15.3% more accurate than the Theta method and 18.2% more accurate than the Comb method.

For the forecasting steps ranging from 22 to 28, SMAPE of Holt exponential smoothing method is 45.11% whereas it is 38.24% for Comb method, 36.56% for Theta method and 31.74% for our proposed method. Our proposed method has a lowest SMAPE among four different forecasting methods. Our method's SMAPE is close to 13.2% more accurate than the Theta method and 17.0% more accurate than the Comb method.

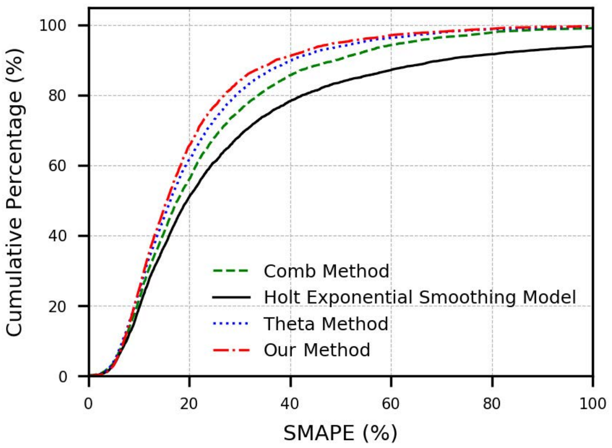


FIGURE 12. The cumulative percentage of SMAPE of wireless cell for KPI of average connections number of RRC about four different methods.

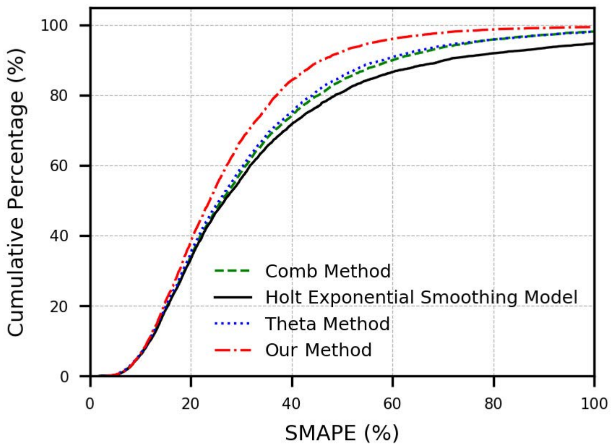


FIGURE 13. The cumulative percentage of SMAPE of wireless cell for KPI of downlink traffic about four different methods.

From Fig. 11, we can find that the farther the forecasting step is, the higher the forecasting error is. Our proposed method has a better performance metrics, especially in the medium and long term forecast, compared with the other three forecasting methods for the KPI of downlink traffic.

The cumulative percentage of SMAPE of wireless cell for KPI of downlink PRB utilization are presented in Fig. 11.

From the Fig. 11, we can find that our proposed method has more wireless cells with low forecasting error compared with the other three methods. Taking SMAPE less than 30% as an example, the cumulative percentage is 68.32% for Holt exponential smoothing method, 75.62% for Comb method, 80.96% for Theta method whereas it is 84.02% for our proposed method.

The cumulative percentage of SMAPE of wireless cell for KPI of average connections number of RRC are presented in Fig. 12.

From the Fig. 12, we can find that our proposed method has more wireless cells with low forecasting error compared with the other three methods. Taking SMAPE less than 30% as an example, the cumulative percentage is 68.36% for Holt exponential smoothing method, 73.70% for Comb method, 78.84% for Theta method whereas it is 82.94% for our proposed method.

The cumulative percentage of SMAPE of wireless cell for KPI of downlink traffic are presented in Fig. 13.

From the Fig. 13, we can find that our proposed method has more wireless cells with low forecasting error compared with the other three methods. Taking SMAPE less than 30% as an example, the cumulative percentage is 56.22% for Holt exponential smoothing method, 58.08% for Comb method, 59.38% for Theta method whereas it is 67.04% for our proposed method.

V. CONCLUSION

In this paper, we carried out experiments on a real large-scale wireless network dataset which contains thousands of wireless cells and corresponding daily KPIs. Experimental results have demonstrated that the proposed method has a better performance, especially in the medium and long term forecasting scenario in terms of SMAPE when compared with some existing methods. It proved that our method can be more suitable for complex wireless communication network environment.

The future direction of our work includes: (1) Use effective and efficient time series extractor to further improve feature extraction and feature expression, such as denosing auto-encoder, sparse auto-encoder, variational auto-encoder; (2) The proposed method uses the classical time series prediction algorithm as the basic predictor, which will limit the final performance of our adaptive combination prediction method. In the future work, the recent neural network model, and machine learning model will be introduced as the basic prediction algorithm to further improve the prediction performance of the adaptive combination prediction method. (3) At present, our proposal method is mainly used in the field of wireless communication network, and this predicition method will be applied to more time series prediction fields later.

REFERENCES

[1] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 796–805, Sep./Oct. 2016.

[2] D. Tikunov and T. Nishimura, "Traffic prediction for mobile network using Holt–Winter's exponential smoothing," in *Proc. 15th Int. Conf. Softw., Telecommun. Comput. Netw.*, 2007, pp. 1–5.

[3] J. Guo, Y. Peng, X. Peng, Q. Chen, J. Yu, and Y. Dai, "Traffic forecasting for mobile networks with multiplicative seasonal ARIMA models," in *Proc. 9th Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Aug. 2009, pp. 377–380.

[4] O. Narmanlioglu, E. Zeydan, M. Kandemir, and T. Kranda, "Prediction of active UE number with Bayesian neural networks for self-organizing LTE networks," in *Proc. 8th Int. Conf. Netw. Future (NOF)*, Nov. 2017, pp. 73–78.

[5] 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.

[6] J. Riihijarvi and P. Mahonen, "Machine learning for performance prediction in mobile cellular networks," *IEEE Comput. Intell. Mag.*, vol. 13, no. 1, pp. 51–60, Feb. 2018.

[7] 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. 2016.

[8] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward," *Int. J. Forecasting*, vol. 34, no. 4, pp. 802–808, Oct. 2018.

[9] S. Makridakis and M. Hibon, "The M3-competition: Results, conclusions and implications," *Int. J. Forecasting*, vol. 16, no. 4, pp. 451–476, 2000.

[10] C. Robert, C. William, and T. Irma, "STL: A seasonal-trend decomposition procedure based on loess," *J. Off. Statist.*, vol. 6, no. 1, pp. 3–73, 1990.

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[12] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[13] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.

[14] G. E. P. Box *et al.*, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[15] E. S. Gardner, "Exponential smoothing: The state of the art—Part II," *Int. J. Forecasting*, vol. 22, no. 4, pp. 637–666, Oct. 2006.

[16] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. London, U.K.: Bowker-Saur, 2014.

[17] V. Assimakopoulos and K. Nikolopoulos, "The theta model: A decomposition approach to forecasting," *Int. J. Forecasting*, vol. 16, no. 4, pp. 521–530, 2000.

[18] R. J. Hyndman and B. Billah, "Unmasking the theta method," *Int. J. Forecasting*, vol. 19, no. 2, pp. 287–290, Apr. 2003.

[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Comput. Sci.*, 2014.

[20] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–284.

[21] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, 2013.

[23] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 1–5.

[24] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 571–583, Apr. 2020, doi: 10.1109/TSTE.2019.2897688.

[25] D. Liang, J. Zhang, S. Jiang, X. Zhang, J. Wu, and Q. Sun, "Mobile traffic prediction based on densely connected CNN for cellular networks in highway scenarios," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2019, pp. 1–5.

[26] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Int. J. Forecasting*, vol. 36, no. 3, pp. 1181–1191, Jul. 2020.

[27] F. Liu, Y. Lu, and M. Cai, "A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting," *IEEE Access*, vol. 8, pp. 62423–62438, 2020.

[28] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021.

[29] Z. Liu, Z. Zhu, J. Gao, and C. Xu, "Forecast methods for time series data: A survey," *IEEE Access*, vol. 9, pp. 91896–91912, 2021, doi: 10.1109/ACCESS.2021.3091162.

[30] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec. 2017, doi: 10.1109/TII.2017.2730846.

[31] M. Khodayar, J. Wang, and M. Manthouri, "Interval deep generative neural network for wind speed forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3974–3989, Jul. 2019, doi: 10.1109/TSG.2018.2847223.

[32] Y. Zerui, G. Jie, and J. Zhijian, "User side data application framework of power Internet of Things based on cloud edge end collaboration," *Power Construct.*, vol. 41, no. 7, pp. 1–8, 2020.

[33] Z. Luying, L. Xiaokai, L. Zhao, X. Fangmin, and Z. Chenglin, "Prediction based dynamic resource allocation method for edge computing first networking," *J. China Univ. Posts Telecommun.*, pp. 1–9, May 2022, doi: 10.19682/j.cnki.1005-8885.2022.1010.

[34] D. Mei, Z. Junhua, L. Dunqiao, C. Shizhao, and W. Yifei, "Joint intelligent optimization scheme of MEC computing unloading and resource allocation," *J. Beijing Univ. Posts Telecommun.*, vol. 45, no. 2, pp. 65–71, 2022, doi: 10.13190/j.jbupt.2021-145.

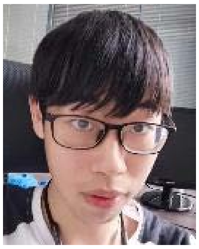
[35] L. Xin, "Scheduling non-stationary bursts of real-time and non-real-time traffic in ATM networks," *High Technol. Lett.*, no. 4, pp. 61–69, 2000.

[36] Z. Yuan, D. Tian, J. Li, and Z. Niu, "Magnetic moment predictions of odd—A nuclei with the Bayesian neural network approach," *Chin. Phys. C*, vol. 45, no. 12, Dec. 2021, Art. no. 124107.

[37] H. Qinwei, T. Qing, W. Nini, C. Qingzheng, W. Tenghui, and Z. Xiaodong, "Multiscale feature fusion convolution neural network method for steady-state visual evoked potential target recognition," *J. Xi'an Jiaotong Univ.*, vol. 56, no. 4, pp. 185–193 and 202, 2022.

WEI FANG received the Graduate degree from the Wuhan University of Technology. During his Ph.D. degree, he went to Madison, University of Wisconsin, USA, for two years of joint doctoral training. He is currently working as an Associate Professor with the School of Automation, Wuhan University of Technology.

YUN CHEN received the Ph.D. degree in management science and engineering from the Wuhan University of Technology, China. She was a Visiting Scholar with the Joseph M. Katz Graduate School of Business, University of Pittsburgh, USA, from 2016 to 2017, a Visiting Scholar with the Center for Research in Economics and Business, Tilburg University, The Netherlands, in 2010, and was employed as a part-time Associate Professor with the Graduate School of Innovation and Technology Management, Yamaguchi University, Japan, in 2012. She is currently an Associate Professor in the field of innovation management with the Wuhan University of Technology. Her research interests include technological innovation, project management, and risk management.



NING PAN is currently a Junior with the School of Automation, Wuhan University of Technology.

BIN RAN received the Ph.D. degree from the University of Illinois, Chicago, USA, in 1993. He is currently a Professor with the Department of Civil and Environmental Engineering, University of Wisconsin–Madison, WI, USA, and the Director of the Research Center for Internet of Mobility, Southeast University, Nanjing, China. He is one of the Co-Founder of the Chinese Overseas Transportation Association. He was the first Chairman. He has authored or coauthored over 90 articles in international journals, including *Transportation Science*, *Transportation Research—Part B: Methodological*, and IEEE ACCESS.