IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 March 2023, accepted 12 April 2023, date of publication 19 April 2023, date of current version 26 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3268437

METHODS

A Hybrid Deep Learning Method Based on CEEMDAN and Attention Mechanism for Network Traffic Prediction

DONG WANG, YU-YANG BAO, AND CHUAN-MEI WANG
School of Science, Wuhan University of Technology, Wuhan 430070, China
Corresponding author: Chuan-Mei Wang (wchuanmei@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFA0714202.

ABSTRACT Accurate prediction of network traffic trends is important for self-management, intelligent scheduling and network resource optimization of base stations. Network traffic prediction is a prerequisite for intelligent scheduling of base stations, and accurate prediction will be beneficial for improving network utilization and energy saving in scheduling. In this paper, a hybrid deep learning method for network traffic prediction, CEEMDAN-TGA which consists of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Temporal Convolutional Network (TCN), Gated Recurrent Unit (GRU), and Attention Mechanism is proposed. Firstly, CEEMDAN is introduced to decompose the original network traffic data into different modes, then reconstruct the modes into trend sequence and noise sequence. Secondly, TCN is used to extract the short-term local features in the network traffic, GRU is used to obtain the long-term data-dependent features, and the attention mechanism is used to improve the prediction accuracy and stability. Finally, through the comparison of experiments, the prediction effect and accuracy of the proposed method are verified to have significant advantages, and the network traffic scheduling strategy is proposed on the basis of prediction.

INDEX TERMS Network traffic prediction, deep learning, complete ensemble empirical mode decomposition with adaptive noise, temporal convolutional network, gated recurrent unit, attention mechanism.

I. INTRODUCTION

With the increasing digitization of society, network communication has become the foundation for building a smart society where everything is connected in the future. According to the October 2022 Global Digital Statistics Report, the number of Internet users reached 5.07 billion, accounting for 63.5% of the world's population, and 171 million new Internet users increased last year. The rapid growth of Internet users and the extended duration of Internet usage have led to a surge in Internet traffic. At the same time, the continuous rise of 5G communication, cloud services, Internet of things and other technologies have led to the rapid growth of network data services. Such explosive growth has increased the load pressure of network base stations continuously, and it led to a series of prominent problems such as aggravated network congestion, increased network delay, and unbalanced resource allocation [1], [2], [3].

Network traffic as one of the carriers of network communication data, a reasonable analysis will help to improve the quality of network services [4]. Accurately analyzing and predicting network traffic trends is a key technology for network traffic analysis, which will help to improve the self-management and intelligent scheduling of base stations. When the peak of network traffic is predicted, increasing the transmission power of base stations and deploying the scheduling strategy in advance are helpful to reduce the occurrence of network congestion. When the trough of network traffic is predicted, reducing the transmission power of base stations and letting some of them go into dormancy are helpful to reduce the energy consumption of base stations. Therefore, accurate network traffic prediction is

The associate editor coordinating the review of this manuscript and approving it for publication was Olutayo O. Oyerinde.

important to improve the quality of service in communication networks [5], [6], [7], [8], [9].

Current research shows that network traffic can be predicted [10] and is correlated in time-series [11]. In practice, the network traffic is very unstable, and the distribution is very complex. At the same time, the network traffic has the characteristics of nonlinearity, timeliness and randomness. In recent years, more and more scholars have carried out research about network traffic prediction, among which how to fully consider the instability and time-series correlation of network traffic has been a research hotspot [12].

Deep learning has a strong ability to capture nonlinear features and has become the mainstream method gradually. Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU) are widely used in time-series tasks and show significant advantages due to the ability to capture long-term dependencies of sequences better. Temporal Convolutional Network (TCN) uses causal convolution to extract the correlation of network traffic time-series, which also achieves significant advantages in some cases [26]. In order to fully extract the features of network traffic sequence, we try to combine the advantages of a single model and introduce the attention mechanism to improve the stability. In addition, we find that denoising is able to improve the prediction accuracy and solve the problem of network traffic sequence instability. The collaboration of the above methods can solve the problems of instability and time-series correlation of network traffic sequence effectively.

In this paper, we propose a hybrid deep learning method CEEMDAN-TGA based on CEEMDAN and attention mechanism to achieve accurate prediction of network traffic. About CEEMDAN-TGA, firstly, CEEMDAN is introduced to decompose network traffic data into several different modes, and then they are reconstructed into trend sequence and noise sequence according to their characteristics, in order to achieve effective denoising. Secondly, the hybrid method of TCN and GRU is used to extract the short-term local features and long-term dependent features of network traffic trend data respectively, for the sake of fully extract and learn the data features. Meanwhile, the attention mechanism is used to adjust the weight of the hybrid method to improve the prediction accuracy and stability. Finally, the performance of CEEMDAN-TGA is compared with common baseline methods, which shows the effectiveness of the prediction method in this paper. The main contributions of this paper can be summarized as follows.

- The adaptive decomposition method CEEMDAN is introduced to decompose and reconstruct the network traffic, which effectively eliminates the noise of the sequence. CEEMDAN suppresses the mode mixing and reduces the error caused by white noise.
- A multi-feature extraction method is proposed, and a hybrid deep learning prediction model based on TCN-GRU-Attention is constructed. TCN extracts short-term local features from network traffic, GRU extracts long-term dependent features, Attention mechanism is used to improve prediction accuracy, and the multi-feature extraction method achieves adequate extraction of hidden features.
- Bayesian Optimization Algorithm (BOA) is used to find the best combination of hyperparameters to improve training efficiency and robustness. Experiments on real base station dataset verify the performance advantages of the CEEMDAN-TGA, and the scheduling strategy based on the prediction method is proposed, which provides support for the next work.

The remainder of the work in this paper is as follows. In Section II, related work on network traffic prediction is presented. In Section III, we come up with the method CEEMDAN-TGA. In Section IV, the experimental results of the method on real datasets are described. In Section V, the base station scheduling and energy-saving strategies are put forward. Finally, we conclude the study and discuss future work.

II. RELATED WORK

Network traffic is a kind of data with time-series, and network traffic data has self-similarity and long-term dependence [13], [14]. The network traffic prediction problem is one of the tasks of time-series problems and has important applications in finance, natural science and other fields [15], [16]. Accurate prediction of network traffic can help network operators to schedule network resources in time, improve network efficiency and utilization, and provide more reliable network quality of service guarantee. In recent years, more and more scholars have carried out research on network traffic prediction, and a large number of prediction methods have been proposed. According to the characteristics of the methods, they can be divided into classical linear prediction methods and nonlinear methods based on neural networks.

A. CLASSICAL PREDICTION METHODS

Classical linear methods can be divided into two types: naive methods and parametric methods. The naive method only relies on historical data for statistical analysis, and one of the most commonly used methods is to take the average of historical data and use the average value as the prediction value at a certain moment in the future, but this method cannot deal with the burstiness of traffic data. Widely used parametric methods include Autoregressive (AR) [17], Moving Average (MA) [18], their combinations and improved methods, including Autoregressive Moving Average (ARMA) [19], Autoregressive Integrated Moving Average (ARIMA) [20], Fractal Autoregressive Integrated Moving Average (FARIMA) [21]. The advantage of these methods is that the theory is mature, but the disadvantage is that it is difficult to describe the long-term correlation and self-similarity of network traffic, on the other hand, it is hard to capture the nonlinear characteristics of network traffic.

B. DEEPING LEARNING METHODS

With the rapid development of artificial intelligence and deep learning, nonlinear methods based on neural networks have gradually replaced most linear methods. The characteristic of these methods is that they can capture the nonlinear characteristics of network traffic sequences better. In the fields of Computer Vision (CV) and Natural Language Processing (NLP), Convolutional Neural Network (CNN) [22] and Recurrent Neural Network (RNN) [23] have achieved remarkable results. In order to capture the complex features of network traffic sequences, researchers use Multilayer perceptron (MLP), Stacked Autoencoder (SAE), Support Vector Machine (SVR) and RNN to predict network traffic. RNNs have gradually become more popular as they have been shown to better capture the long-term dependence of time sequence [24], [25]. With the continuous research, numerous variants of RNN have emerged, among which LSTM and GRU have gradually become mainstream methods due to their ability to better overcome the problems of gradient disappearance and explosion during training. GRU has a simpler structure, fewer training parameters, and better adaptability than LSTM, and it is better at capturing long-term dependent features of sequence. TCN has been widely used for feature extraction problems on time sequences in recent years, and it has shown excellent performance in a number of tasks. TCN uses dilated causal convolution to obtain the correlation of sequences and employ residual connections to solve the problems such as the network degradation [26], [27].

C. SIGNAL DECOMPOSITION METHODS

Decomposing the signal by a decomposition method and extracting the features of the components are beneficial to improve the prediction performance.

Common decomposition methods include Empirical Mode Decomposition (EMD) [40], Ensemble Empirical Mode Decomposition (EEMD) [28], Variational Mode Decomposition (VMD) [29], and CEEMDAN [37]. EMD and EEMD can transform non-stationary time sequences into stationary time sequences by decomposition, and are widely used in time sequences prediction tasks, showing better prediction performance, but the problem of mode mixing occurs [30]. Compared with EMD and EEMD, VMD has better advantages in feature extraction, and solve the problem of mode mixing and boundary effect better. However, its influence parameters need to be set in advance, and the selection of parameters depends on experience [31], [32]. Although VMD obtains excellent results in signal denoising, these effects are determined by two critical parameters, namely, the number of modes K and the value of the penalty factor α, which are usually selected within a certain range. Selecting these parameters by trial-and-error would require a large number of operations and would waste a lot of time. In addition, the results of VMD decomposition may not be accurate enough, too many components of the decomposition will make it difficult to present regular features of the signal, too few components of the decomposition will not be comprehensive enough to express the features represented by the sequence [33], [34]. Hence, appropriate methods are needed to obtain the optimal values of these parameters. CEEMDAN is an adaptive decomposition method, which overcomes the problems of white noise error question of EEMD and the limitation of the parameter selection of VMD.CEEMDAN is able to achieve adequate decomposition of sequences adaptively, and the features of decomposed components are clearer, which shows obvious advantages in a part of the study [35]. Therefore, it is important to adopt an appropriate decomposition method for feature extraction.

In view of the different advantages of different deep learning methods, some scholars integrate the advantages of single neural network methods to achieve accurate prediction of network traffic through hybrid methods. Li combined LSTM with attention mechanism, LSTM can extract the temporal characteristics of network traffic, and attention mechanism can adjust the weight of hidden states to reduce the information loss rate and achieve higher accuracy prediction [36]. Xiong et al. combined CNN, LSTM, historical data component, and linear component, then the effect was significantly improved [37]. Branco et al. used wavelet decomposition and LSTM for network traffic prediction, and the results showed that the combined method outperformed the single LSTM [38]. Bi et al. used the SG filter to denoise the traffic data, and then features were extracted by the hybrid method of TCN and LSTM, and the prediction of network traffic was significantly improved [39].

In summary, there are two main problems with the existing mainstream neural network-based traffic methods. First, due to the interference of the base station by the external environment such as temperature and humidity, the network traffic data usually has a large noise influence, and many methods neglect the effective removal of noise. Second, the existing prediction methods often only considered a single method, neglecting the combination advantages of different methods to achieve adequate feature extraction.

Therefore, a hybrid deep learning method CEEMDAN-TGA is proposed in this paper, which is composed of CEENDAN, TCN, GRU and attention mechanism. There are clear motivations in our model selection. The classical LSTM can capture the correlation of time series well, but the LSTM uses three gate structures. The complex structure and parameters bring difficulties to the training, and it is difficult to meet the requirement of timeliness of network traffic prediction. Therefore, we use a three-layer GRU to improve it. The GRU has the characteristics of simple structure, small number of parameters, fast convergence speed and so on, which can capture and learn the long-term correlation characteristics of network traffic series to achieve high accuracy prediction. In addition to the long-term correlation features, the network traffic has obvious local short-term dependence features, which are specifically expressed as local burst and local periodicity features of traffic. Therefore, we introduce the TCN, which can capture and learn local features through one-dimensional convolution operation, and can extract high and low frequency information well.

To further optimize the weight assignment, the attention mechanism is introduced to adjust the hidden state weights and reduce the information loss rate. By combining TCN, GRU and attention mechanism, we achieve sufficient extraction of long-term global correlation features and short-term local dependency features to achieve more accurate prediction. In addition, the existing studies have found that the noise in network traffic sequence brings some interference and error to the prediction, and adaptive decomposition through CEEMDAN can effectively identify and remove noise.

To CEEMDAN-TGA, firstly, CEEMDAN is introduced to decompose the network traffic sequences and reconstruct the trend data, which eliminates the influence of noise effectively. Secondly, this paper combines and collaborates the advantages of TCN, GRU and attention mechanism innovatively. In the first stage, TCN is used to extract the short-term local features of the trend data. In the second stage, GRU is used to extract the long-term dependent features, and then the data features are fully extracted.

III. METHODOLOGY

In this section, the CEEMDAN-TGA is described in detail, as well as the basic principles and advantages of each module.

A. CEEMDAN

Network traffic sequences have the characteristics of nonlinearity, timeliness and randomness, and they are typical non-stationary data. These characteristics have a great impact on the accuracy of network traffic prediction. In order to reduce the influence of noise and improve the accuracy of prediction, this paper introduces the CEEMDAN to decompose the nonlinear and non-stationary sequences of network traffic into stationary components. Then identify the noise according to the characteristics of the components to reconstruct it into noise sequence and trend sequence. At last, we use the trend sequence to solve prediction tasks.

CEEMDAN [40] is proposed on the basis of EMD [41] and EEMD [28]. EMD is used to extract the fluctuation trend of non-stationary data sequences on different time scales. It can decompose the non-stationary data sequences into Intrinsic Mode Function (IMF) components, but the mode mixing phenomenon may occur in EMD. EEMD reduces the modal mixing phenomenon by adding pairs of positive and negative Gaussian white noise to the data to be decomposed, but there is residual white noise that affects the data. CEEMDAN adds a group of adaptive white noise with the same size and opposite sign to calculate the overall average after obtaining each order IMF, which not only overcomes the problem of incomplete decomposition, but also the reconstruction error is close to 0, and the calculation scale is reduced effectively [42].

Firstly, define $E_i(\cdot)$ as the ith IMF after EMD decomposition, $\widetilde{C_i(\cdot)}$ as the ith IMF obtained after CEEMDAN decomposition, w^j is the Gaussian white noise sequence subject to the normal distribution, j is the times of adding white noise, and ε is the parameter value of white noise. The main steps of CEEMDAN include:

1) Add Gaussian white noise to signal $x(t)$ to get a new signal $x'(t) = x(t) + (-1)^m \varepsilon w^j$, where $m = 1, 2$. Then the new signal is decomposed by EMD to get the first IMF $C_1^j(t)$, the calculation formula is:

$$E(x'(t)) = C_1^j(t) + r^j \tag{1}$$

2) Calculate all IMFs components separately and calculate the overall average to obtain the first IMF $\widetilde{C_1(t)}$ of CEEMDAN, the calculation formula is:

$$\widetilde{C_1(t)} = \frac{1}{N}\sum_{j=1}^{N} C_1^j(t) \tag{2}$$

3) Calculate the residual $r_1(t)$ of the first stage.

$$r_1(t) = x(t) - \widetilde{C_1(t)} \tag{3}$$

4) The new signal is obtained by adding positive and negative paired Gaussian white noise to the residual $r_1(t)$ again, and the first IMF $D_1^j(t)$ is obtained by EMD, and then the overall average of the generated N IMFs is performed to obtain the second IMF $\widetilde{C_2(t)}$ of CEEMDAN, the calculation formula is as follows.

$$\widetilde{C_2(t)} = \frac{1}{N}\sum_{j=1}^{N} D_1^j(t) \tag{4}$$

5) Calculate the residual $r_2(t)$ of the second stage.

$$r_2(t) = x(t) - \widetilde{C_2(t)} \tag{5}$$

6) Repeat the above steps until the residual signal is a monotonic function, and then stop the decomposition. If the number of IMFs obtained is M, the original signal $x(t)$ can be decomposed as (6).

$$x(t) = \sum_{m=1}^{M} \widetilde{C_m(t)} + r_m(t) \tag{6}$$

By CEEMDAN decomposition, the non-stationary signal can be decomposed into multiple stationary components, and then according to the characteristics of each component, it is divided into trend and noise, and the trend component is reconstructed. The reconstructed signal is used as the denoised network traffic sequence. Denoising by CEEMDAN is much beneficial to improve the prediction accuracy and reduce the impact of noise.

B. TCN

TCN is widely used in the prediction of time sequences, because the effect of TCN in time sequence modeling is better than other recursive structures, and TCN has stronger memory ability [27]. TCN is a special one-dimensional convolutional network, which is composed of causal convolution, dilated convolution and residual block. Such a structure can make full use of the historical information in network traffic sequences, and the TCN structure diagram is shown in Figure 1.

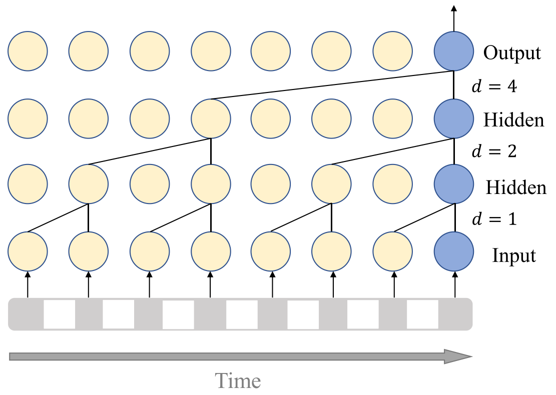

FIGURE 1. TCN structure diagram.

1) CAUSAL CONVOLUTION

In causal convolution, the value of the current network layer at time t only depends on the value of the previous layer at time t and before time t.The connection and information transfer of network layers strictly follow the causality relationship, which makes the network take full advantage of historical information. However, when enough historical information is considered, the number of convolutional layers must be increased, which leads to the problems of gradient disappearance, complex training, and poor fitting degree.

2) DILATED CONVOLUTION

The emergence of dilated convolution is to solve the problem of too many layers in causal convolution. Dilated convolution increases the receptive field by expanding the convolution field of view, so as to obtain a larger receptive field and more historical information with fewer network layers and lower calculation. The dilated convolution calculation formula is defined as follows:

$$F(s) = (x \odot f)(s) = \sum_{i=0}^{k-1} f(i) x_{s-d \cdot i} \tag{7}$$

where d is the dilation factor, k is the convolution kernel size, and $\odot$ is the convolution operation.

3) RESIDUAL BLOCK

In order to improve the depth of the model, prevent the problem of gradient disappearance and explosion due to high network depth, and prevent the data shape and dimension change caused by convolution, a one-dimensional convolution is introduced as the residual block. The input data of TCN needs to go through two rounds of dilated causal convolution, weight normalization, activation function and dropout layer, the one-dimensional convolution is used as the residual module for skip connection, and the final output result is sent to the next layer. The structure of the residual block is shown in Figure 2, and the formula of the residual connection is:

$$f_{output} = ReLU(x + f(x)) \tag{8}$$

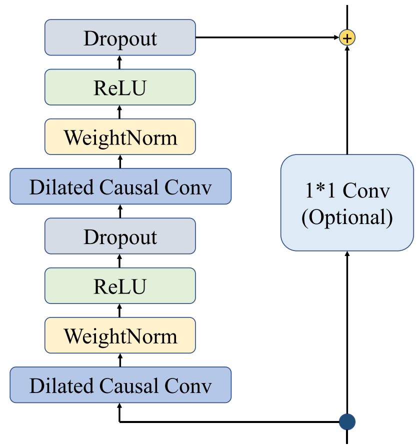

FIGURE 2. Residual block structure diagram.

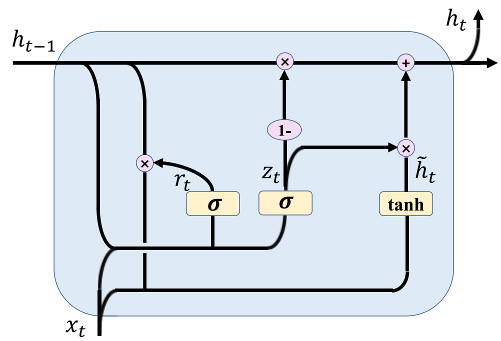

FIGURE 3. GRU structure diagram.

C. GRU

GRU is an improved method based on RNN and LSTM. Compared with LSTM, GRU improves the gate control structure, combines the input gate and the forgot gate in LSTM into an update gate (z_t), and changes the output gate into a reset gate (r_t) [43]. The function of the update gate is to control the amount of data stored in the memory information of the previous moment to the current moment, and the function of the reset gate is to control the amount of historical information that needs to be forgotten. The structure diagram of the GRU is shown in Figure 3.

Compared with LSTM, GRU has the advantages of simpler structure, less internal unit redundancy, fewer parameters, and faster calculation speed, which is more in line with the timeliness requirements in the field of network traffic prediction. The operation expression of GRU is:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{9}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{10}$$

$$\widetilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{11}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{12}$$

where r_t is the reset gate, σ is the activation function, W is the weight matrix, x_t is the input at time t, h_{t-1} is the output at time $t - 1$, and $\widetilde{h}_t$ is the candidate hidden state.

D. ATTENTION MECHANISM

Since network traffic data sequences are long sequences, they cannot fully consider and retain all necessary information when passing through GRU, which can lead to weakened effects. Using the feature of information filtering of attention mechanism, the collaboration of attention mechanism with GRU can effectively improve the prediction accuracy. Attention mechanism can evaluate the importance of information at different moments and allocate more attention to key parts to improve the prediction accuracy [46].

The attention mechanism has the effect of information filtering. The information with low correlation is actively lost by reducing the weight, and the information with high correlation is increased by increasing the weight. Network traffic data series often show strong periodicity of daily, weekly and monthly periods, so it is important to capture the periodicity and related information to improve the accuracy of prediction. The features with low or irrelevant correlation caused by bursty information and noise are effectively suppressed, so as to improve the overall performance and efficiency of the method.

E. CEEMDAN-TGA

In order to combine and collaborate the advantages of the above methods, reduce the impact of noise and maximize the extraction of hidden features, we propose the CEEMDAN-TGA, whose advantages include the denoising capability of CEEMDAN, the short-term feature extraction capability of TCN, the long-term dependency capturing capability of GRU, and the weight assignment capability of attention mechanism. CEEMDAN-TGA connects CEEMDAN, TCN, GRU, and attention mechanism end-to-end to make it a continuous whole. Firstly, CEEMDAN identified and removed noise through signal decomposition. Then, the advantages of the fitting ability of the nonlinear models of TCN and GRU are combined to achieve sufficient extraction of features, and then the weights are adjusted through the attention mechanism. Finally, the extracted long and short-term temporal features are mapped to the real traffic values through the fully connected layer to achieve accurate prediction of network traffic. The method of CEEMDAN-TGA is shown in Figure 4.

The reason that we use the hybrid method to extract features is that the features of network traffic have short-term local features and long-term dependent features in the time dimension. Different methods have different features and advantages, and the hidden features can be explored by using appropriate methods for training and learning. TCN uses convolutional networks, which can capture short-term local information of the data well, while the receptive field of the TCN can be flexibly adjusted and is suitable for network traffic prediction in most cases [45]. GRU has the characteristics of small parameters and fast convergence, which can better deal with the sudden and timely characteristics of network traffic. Using a hybrid model of TCN and GRU to extract both features simultaneously will be more accurate and effective in the network traffic prediction task compared to a single model.

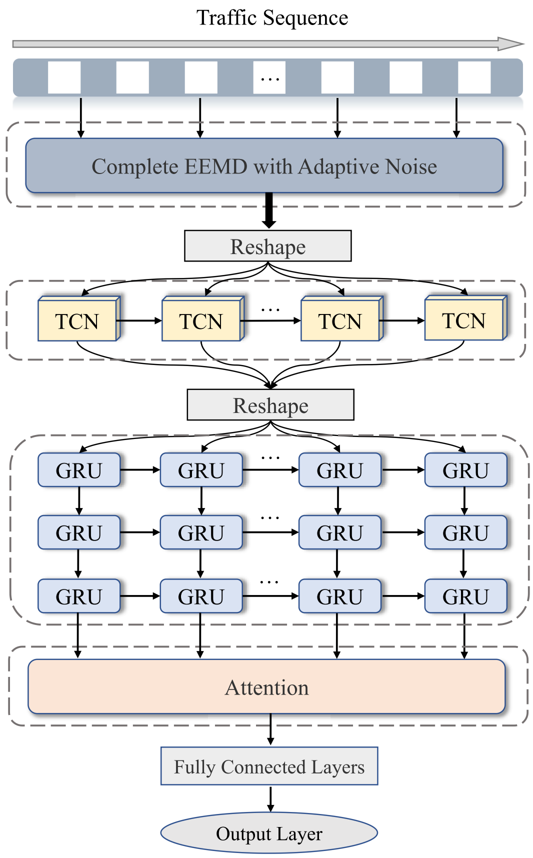

FIGURE 4. CEEMDAN-TGA structure diagram.

IV. EXPERIMENT AND RESULT ANALYSIS

In this section, CEEMDAN-TGA is experimented on a real base station dataset to verify the effectiveness and performance of the proposed method.

A. DATA SET DESCRIPTION

The dataset used in the experiment is from the open data science competition "AIIA Home Network Competition: Network Traffic Forecasting", organized by the China Mobile and China Artificial Intelligence Industry Development Alliance. The dataset contains network traffic data for three anonymized regional base stations from January 2017 to November 2018, which is typical of time series data. The dataset of this experiment is the base station traffic data of a sampling area of the dataset, from January 1, 2017 to March 31, 2017. The data is sampled every hour, with a total of 2160 data [47]. In order to extract the periodic features of the sequence, the input step is set to 24 to predict the network traffic value at a future moment. In the experiment, the data is divided according to the ratio of 8:2. The training set has 1728 data, which are used to input into the TGA for training and tuning, and the Bayesian optimization algorithm is used

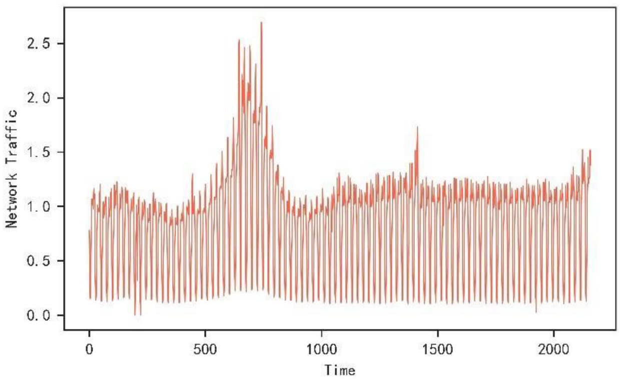

FIGURE 5. The distribution of original data.

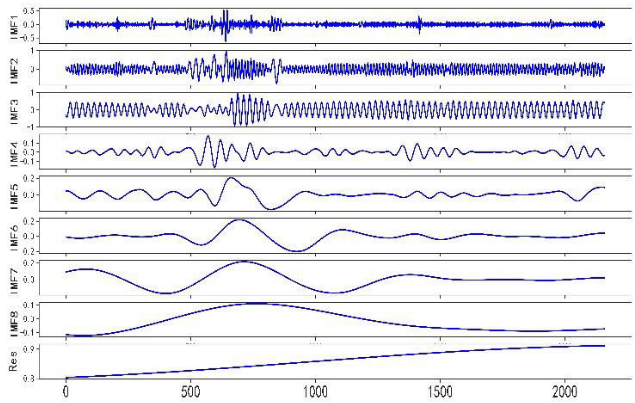

FIGURE 6. The IMFs obtained by CEEMDAN decomposition.

to tune the hyperparameters of the TGA. The remaining 432 data are used as the test set to calculate evaluation metrics to test the performance of the model. The original data distribution is shown in Figure 5, where the horizontal axis Time represents the time dimension in hours, and the vertical axis Network Traffic represents the traffic in GB/s.

The data set shows that the network traffic presents certain periodicity, burstiness and nonlinearity. The periodicity is reflected in the cycle of days, which is at a higher level at about 11:00-23:00 each day, showing a "peak"; it has a lower level at about 3:00-8:00 each day, showing a "trough". Burstiness usually occurs at a certain time or period. At a certain time, it is reflected in the high and low value of network traffic, which is due to the influence of a variety of sudden factors and base stations. A certain period is reflected in data 620th-790th of the dataset, which is due to the fact that these days are "The Chinese New Year", and in such a special time, the network usage and the number of people will surge. Nonlinearity is the most typical characteristic of network traffic data.

B. DATA PREPROCESSING

The data preprocessing stage includes normalization and denoising. The distribution of network traffic is very complex, and it is beneficial to use normalization to reduce gradient disappearance and accelerate convergence. In this experiment, Min-Max Normalization is used, which is a linear transformation of the original data and maps the whole data into the interval [0,1]. The normalization formula is:

$$X_i^* = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{13}$$

where X_i is the original network traffic data, X_{min} and X_{max} are the minimum and maximum values of all original network traffic data, and X_i^* are the normalized network traffic data.

In the second stage of data preprocessing, CEEMDAN is used to decompose the normalized sequences. The sequence is decomposed into multiple IMFs in the process of iteration, each IMF represents a different characteristic, and the decomposition ends when the residual sequence presents a monotone function. During the experiment, CEEMDAN decomposes the sequence into 8 IMFs and 1 residual sequence with monotonicity (Figure 6). The different IMFs present certain characteristics such as stationary periodicity, trend and non-stationary noise, respectively. The results of CEEMDAN decomposition show that IMF1 and IMF2 present certain noise characteristics, IMF2, IMF3 and IMF4 present certain periodic characteristics, and IMF4-IMF8 present certain trend characteristics. These characteristics represented by IMF cannot be completely determined by experience only.

The purpose of CEEMDAN decomposition is to find the noise and eliminate it. In order to accurately identify which IMFs represent noise, experiments are needed to judge them. Firstly, the IMFs suspected of presenting noisy features are eliminated, and then the other IMFs are reconstructed. Then, the reconstructed sequence is inputted into the hybrid deep learning model for training and tuning until the best parameters are obtained, and the prediction results are output. Finally, the results of each experiment are compared, and the combination of IMFs corresponding to the optimal result is the best combination, and the IMFs eliminated by the combination are considered to be noisy sequences.

The experimental results show that the results are optimal if and only if the IMF1 is removed. So IMF1 is considered as the noise sequence, IMF2-IMF8 and the residual sequence are considered as the trend sequences. Therefore, IMF2-IMF8 and the residual sequence are reconstructed, and the reconstructed data is the denoised data. The denoised sequence is inputted into the TGA for training and prediction. Figure 7 shows the trend sequences distribution diagram, and Figure 8 shows the noise sequence distribution diagram.

C. EVALUATION METRICS

The performance of network traffic methods is usually measured by several error evaluation metrics. The smaller the error of the method, the better the prediction; the higher the accuracy, the better the prediction effect. Commonly used evaluation metrics are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 [4], and

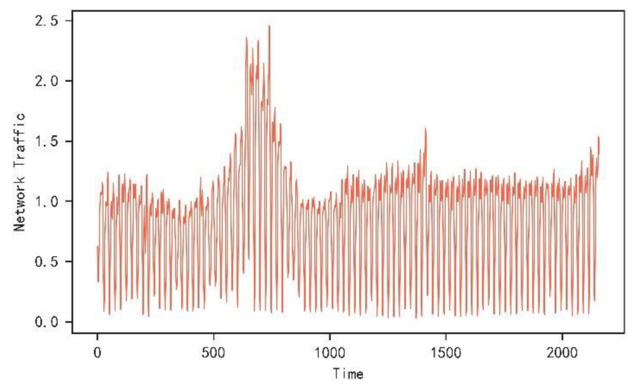

FIGURE 7. The distribution diagram of trend sequences.

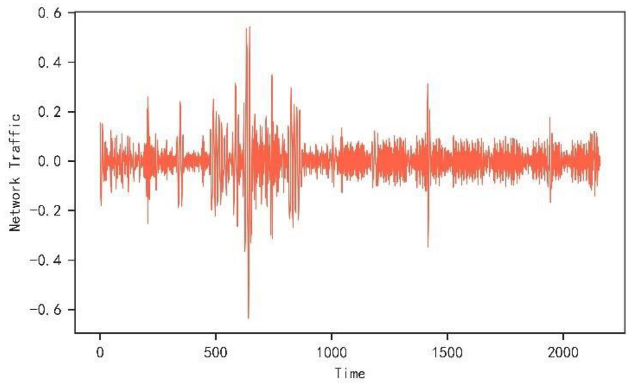

FIGURE 8. The distribution diagram of noise sequence.

the formulas are:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|x_i - x_i'\right|^2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - x_i'\right| \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - x_i')^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{16}$$

D. BASELINES

In order to verify the advantages of the CEEMDAN-TGA, the current widely used methods were selected for comparison: ① TGA (TCN+GRU+Attention); ②SG+TCN+LSTM [39]; ③SSA+LSTM [52]; ④Wavelet+LSTM [38]; ⑤XGBOOST [47]; ⑥SVR [48]; ⑦ARIMA [49]. The baseline method ①TGA is added to compare and verify the reasonableness and effectiveness of CEEMDAN for denoising.

E. PARAMETER SETTING

The components of CEEMDAN-TGA contains a large number of hyperparameters, and the setting of hyperparameters directly affects the experimental results. In order to find the optimal parameter combination efficiently, we adopted

TABLE 1. Comparison of TCN parameters in CEEMDAN-TGA.

Parameter	Value	RMSE	MAE	R^2
Kernel _ size	2	0.052751	0.042345	0.9819
	3	**0.051496**	**0.042075**	**0.9828**
	4	0.053554	0.042803	0.9814
	5	0.052479	0.042063	0.9821
Dilations	1	0.058045	0.046061	0.9781
	2	**0.051496**	**0.042075**	**0.9828**
	3	0.055855	0.044511	0.9797
	4	0.056445	0.046451	0.9793
Activation Function	*ReLU*	**0.051496**	**0.042075**	**0.9828**
	Sigmoid	0.063007	0.048872	0.9742
	Tanh	0.051664	0.042393	0.9827
	LeakyReLU	0.051986	0.042896	0.9827

the Bayesian optimization algorithm. The specific idea of Bayesian optimization algorithm is to use the prior probability distribution of the objective function and the known observation points to update the posterior probability distribution, and then find the next minimum point according to the posterior probability distribution. In this process, the minimum decreases continuously until the optimal parameter is obtained. Bayesian optimization employs a Gaussian process as an agent model to continuously update the prior, which is characterized by a low number of iterations, high speed, and high robustness. The main parameters include the size of the convolution kernel *Kernel_size*, the dilations coefficient *Dilations* and the activation function of the TCN; the number of hidden layer neurons *Hidden_size* and the number of layers *Number_layers* of the GRU.

To Bayesian optimization, the best parameters after optimization can be obtained by sampling from the region where the global optimal solution is most likely to occur and the region that is not sampled through the sampling function, and minimizing the loss function through continuous iteration. The preselection ranges of the parameters are given firstly and Bayesian optimization searches through these preselection ranges to find the best combination of parameters for the current model. In fact, in order to compare the evaluation metrics of different parameters more obviously, we also conducted experiments on some pre-selected values and obtained a complete comparison of parameter performance (Table 1 and Table 2).The results show that when the $Kernel\_size = 3$, $Dilations = 2$, residual block using*ReLU*, $Hidden\_size = 64$, $Number\_layers = 3$, the RMSE and MAE reach the minimum, R^2 reaches the maximum, and the performance is optimal at this time.

In addition to the main parameters mentioned above, the optimizer *Optimizer*, the number of iterations *Epochs*, the

TABLE 2. Comparison of GRU parameters in CEEMDAN-TGA.

Parameter	Value	RMSE	MAE	R^2
Hidden_size	16	0.052936	0.043330	0.9818
	32	0.051527	0.042099	0.9828
	64	**0.051496**	**0.042075**	**0.9828**
	96	0.053297	0.043577	0.9816
Number_layers	1	0.053062	0.043705	0.9817
	2	0.053601	0.043792	0.9813
	3	**0.051496**	**0.042075**	**0.9828**
	4	0.051595	0.042284	0.9827

TABLE 3. Parameter settings of CEEMDAN-TGA.

Parameter	Value
Kernel_size	3
Dilations	2
Activation Function	*ReLU*
Hidden_size	64
Number_layers	3
Optimizer	*Adam*
Epochs	500
Input	24
Output	1

network input length *Input* and the network output length *Output* are also included. The experimental results show that when the *Adam* optimizer is used, the convergence rate and loss value of the method reach the best, so *Adam* is used as the optimizer. The final parameter Settings of CEEMDAN-TGA are shown in Table 3.

F. PREDICTION RESULT ANALYSIS

The experiment compares and analyzes the prediction results from two aspects of qualitative analysis and quantitative calculation, using the fitting diagram and accuracy table.

The qualitative analysis is to compare the predicted values of the traffic data on the test set with the fitted images of the true values (Figure 9). The experimental results show that CEEMDAN-TGA is almost consistent with the trend of the real traffic curve, and it has a high degree of fit and good prediction effect. However, the predicted results of other methods have a large deviation from the true results. In order to further verify the effectiveness and robustness of CEEMDAN-TGA, comparative analysis through quantitative calculation is still needed.

The quantitative calculation is compared with the baseline method through RMSE, MAE, and R^2 (Table 4). The evaluation metrics can effectively reflect the accuracy of the prediction. The results in Table 4 show that CEEMDAN-TGA achieves the lowest error in RMSE and MAE, and the highest accuracy in R^2. Therefore, the predicted value of CEEMDAN-TGA is the best fit to the true value.

TABLE 4. Evaluation metrics values of CEEMDAN-TGA and baseline method.

Method	RMSE	MAE	R^2
CEEMDAN-TGA	0.05149	0.04207	0.9828
TGA	0.05856	0.04306	0.9777
SG+TCN+LSTM	0.05580	0.04298	0.9798
SSA+LSTM	0.05758	0.04421	0.9785
Wavelet+LSTM	0.05976	0.04617	0.9769
XGBOOST	0.06320	0.04318	0.9741
SVR	0.07758	0.05949	0.9610
ARIMA	0.11749	0.08010	0.9239

The results in Table 4 show that when CEEMDAN is used to denoise, the RMSE, MAE and R^2 of CEEMDAN-TGA are 0.05149, 0.04207 and 0.9828, respectively. When the baseline method ①TGA is not denoising with CEEMDAN, the RMSE, MAE and R^2 are 0.05856, 0.04306 and 0.9777, respectively. The use of CEEMDAN denoising reduces RMSE by 12.07%, MAE by 2.3%, and increasesR^2 by 0.52%. Therefore, the noise can be effectively removed by CEEMDAN, and denoising is beneficial to improve the prediction accuracy.

In order to show the performance advantages of CEEMDAN-TGA, we choose the current popular methods with strong performance advantages for comparison, including ③SG+TCN+LSTM, ④SSA+LSTM and ⑤ Wavelet+LSTM, which have been widely used in time sequences prediction tasks. The experimental results show that the RMSE of CEEMDAN-TGA is reduced by 7.72%, the MAE is reduced by 2.12%, and the R^2 is increased by 0.31%, compared to the best current method ⑤. Compared with the traditional methods of ⑤XGBOOST, ⑥SVR and ⑦ARIMA, CEEMDAN-TGA shows a more significant performance advantage.

In order to further show the feasibility of CEEMDAN-TGA, the time complexity of CEEMDAN is analyzed. CEEMDAN adopts a double-layer cyclic nesting and its time complexity is quadratic order $O(n^2)$, where n is the total length of the sequence. TCN adopts the structure of CNN, and its time complexity is linearly and positively correlated with the dimension d, and its time complexity is $O(n)$. The time complexity of GRU-Attention is $O(3(2Nm + m^2))$, where N is the number of samples and m is the number of neurons, and its complexity is $O(n^2)$. To CEEMDAN-TGA, only when the highest order term is retained, its time complexity is $O(n^2)$. The experimental procedure shows that the efficiency of the model is improved by using Bayesian optimization, and the training time of CEEMDAN-TGA is 280.11 seconds and

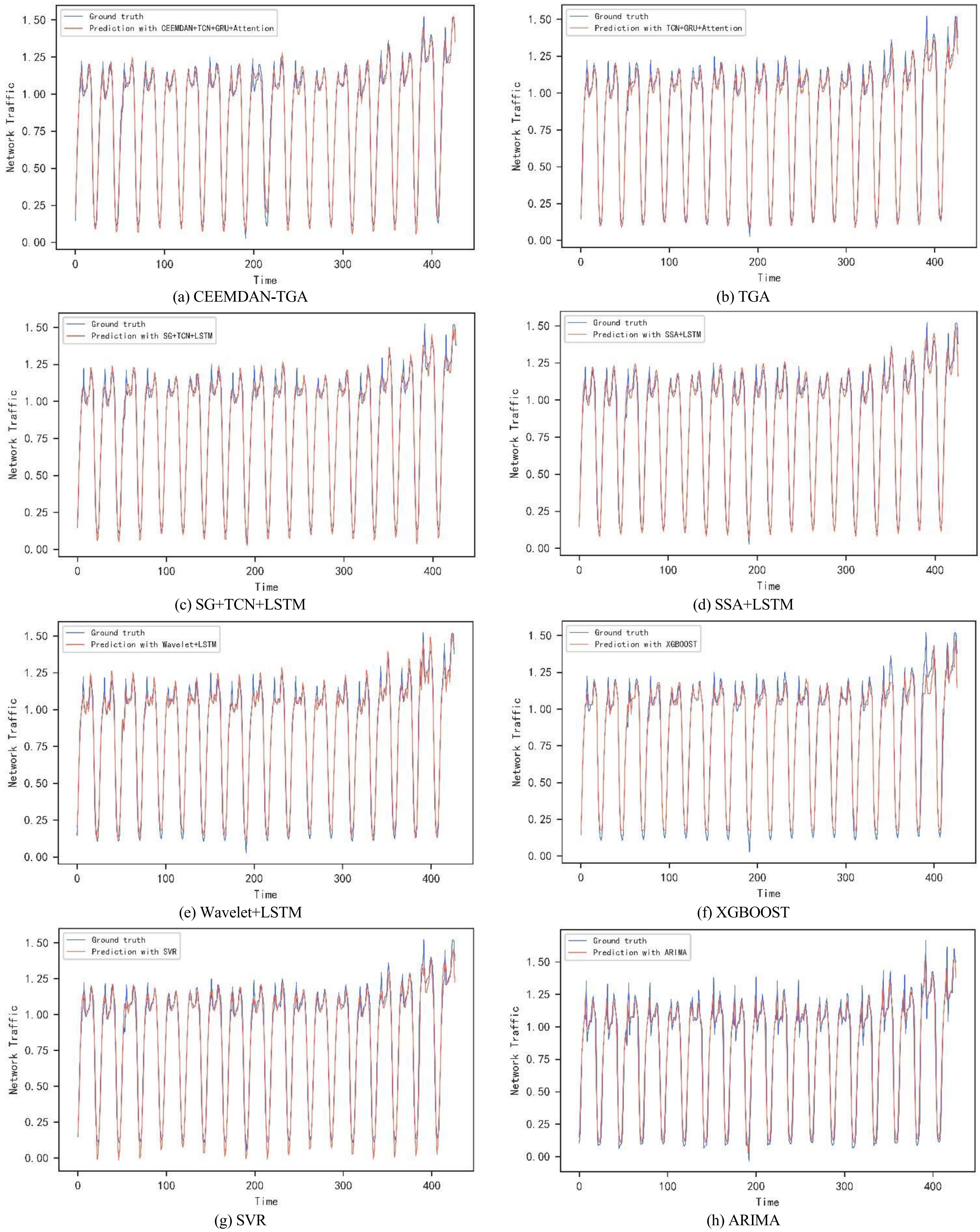

FIGURE 9. **Predicted values of CEEMDAN-TGA with the baseline method.**

the prediction time is 0.1086 seconds, which is of practical application.

Further analysis shows that the reason why CEEMDAN-TGA can show its advantages is that the convolution operation of TCN can effectively extract the local burst changes of network traffic sequence, which is more sensitive to short-term local features, and thus has significant advantages in the extraction of burst features. Secondly, GRU can effectively capture the periodic characteristics of network traffic sequence, which is more sensitive to long-term dependence features, and thus has significant advantages for the extraction of periodic features. Finally, the introduced CEEMDAN can remove the noise effectively, and the introduced attention mechanism can help important information to be filtered, which can effectively improve the overall accuracy and robustness. The combined advantages of these methods make CEEMDAN-TGA perform well in the prediction experiments.

V. SCHEDULING AND ENERGY-SAVING STRATEGIES

Network traffic prediction is the basis of self-management and intelligent scheduling of base stations. Accurate prediction helps base station to deploy scheduling strategy in advance, which plays an important role in reducing network congestion and green energy saving. The main work of the scheduling strategy is as follows: Firstly, the traffic value of the base station at one or several moments in the future is predicted according to the network traffic prediction method. Then, the network utilization rate was calculated according to the predicted value of network traffic, and the network status was classified. Finally, the corresponding network link scheduling strategy was proposed for each level according to its characteristics.

The utilization of the network represents the operating state of the network, so grading according to the utilization can indicate different network states. Define the network utilization and preference curve (Figure 10), and the curve adopts the idea of *Softmax*. The horizontal axis represents the utilization of the network, and the vertical axis represents the selection preference of the network links. The link with higher preference is given a higher probability of being selected during routing scheduling, which has certain randomness and dynamics [50]. Define α as the utilization of the network, which is the ratio of the flow value to the total capacity value at the current moment. The network status is graded according to utilization:

1) When $0 \leq \alpha \leq 40\%$, it is defined as Mice Flow (MF), which is characterized by a small amount of short-time data transmission through network links. A low utilization rate indicates that certain network links are not fully utilized and there is a problem of resource waste.

The scheduling strategy of the mouse flow is Preference Algorithm (PA): the link with low utilization rate is given a lower selection preference, and the lower preference makes the link less likely to be selected, so as to inhibit the new data flow to the link. After a period of time, part of the link utilization drops to 0, and then the link starts to sleep. Some of the multiple original links are no longer used and go to sleep, while others are used more fully, so as to save energy and improve efficiency.

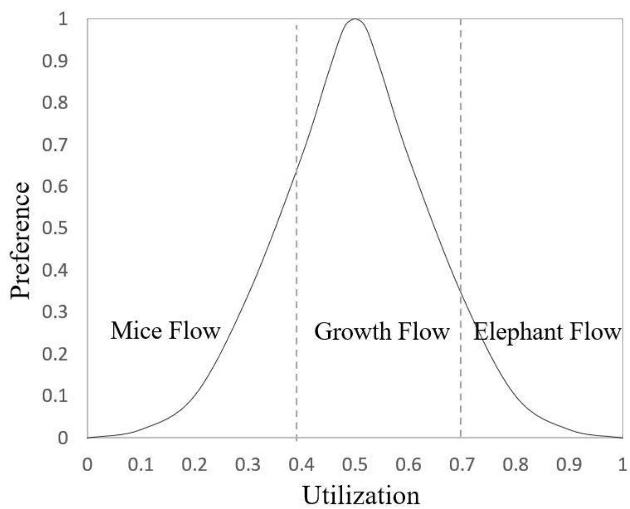

FIGURE 10. Network utilization and preference curve.

2) When $40\% < \alpha \leq 70\%$, it is defined as Growth Flow (GF), which is characterized by the data transmission of the network link in a continuously higher state. Stable and high utilization is a performance of the resources. As the utilization continues to grow steadily, the preference increases, which promotes the network, but the transmission delay may increase and packet loss may occur.

The scheduling strategy of growth flows is Preference Algorithm based on K-Shortest Paths (PA-KSP). (i) Calculate K shortest paths based on topology information of network nodes. (ii) Preference was given to the utilization of K shortest paths respectively. (iii) The link selection is made according to the link selection preference. The purpose of selecting K shortest paths is to ensure that the network transmission process as short as possible, and the final selection is made according to the link preference, which can promote the network more effectively, so as to realize full utilization.

3) When $70\% < \alpha \leq 100\%$, it is defined as Elephant Flow (EF), which is characterized by large and continuous data transmission over network links. Higher network utilization indicates better bandwidth balance. With the rapid increase of utilization, it is easy to cause network congestion. Solving the problem of network congestion and achieving better load balancing is the key to this situation [51].

The scheduling strategy of elephant flow is the Maximum Bottleneck Bandwidth Algorithm (MBBA). The idea of this method is to choose the path with the maximum bottleneck bandwidth as the next traffic path. (i) Calculate K shortest paths based on topology information of network nodes. (ii) Calculate the bottleneck bandwidth of the K paths at the next moment. The calculation formula is shown in (17), where $X^k_{bottleneck}$ is the bottleneck bandwidth of the kth path, $C^k_{capacity}$ is the capacity of the kth path, $x^1_{predicted}$ is the predicted value at the next moment. (iii) The link with the largest residual bandwidth is selected as the routing choice to solve the congestion.

$$X^{k}_{bottleneck} = C^{k}_{capacity} - x^{1}_{predicted} \tag{17}$$

The development of 5G technology has brought great convenience to communication, and it also has huge energy consumption problems. The power of 5G base station is about three times that of 4G base station, and the high-frequency characteristics of 5G base stations make their coverage range only a quarter of that of 4G, which leads to huge energy consumption of 5G base stations. In China, for example, there are more than 2 million 5G base stations, and operators need to pay about RMB 30 billion in electricity costs each year. Up to now, the number of 5G base stations in the world has reached more than 3 million. The existing studies show that reasonable scheduling and energy-saving strategies can save energy by about 5-25%. Rough calculations suggest that the global savings could amount to hundreds of millions of dollars at least. Therefore, it is of great significance to use network traffic prediction as a service for scheduling and energy saving.

VI. CONCLUSION AND FUTURE WORK

Accurate prediction of network traffic plays an important role in intelligent scheduling, congestion control and anomaly detection of base stations. In the face of the complex characteristics of network traffic sequence, denoising and fully extracting features can effectively improve the prediction accuracy. In this paper, a hybrid deep learning method CEEMDAN-TGA is proposed for the first time, which combines CEEMDAN, TCN, GRU and attention mechanism. CEEMDAN is used to decompose and denoise the data, TCN is used to obtain short-term local features of the data, and GRU is used to obtain long-term dependent features of the data, attention mechanism is used to adjust the weight of information to improve the accuracy and robustness of prediction. The experimental results verify the effectiveness of CEEMDAN denoising, and the significant performance advantages of CEEMDAN-TGA are verified from two aspects of qualitative analysis and quantitative calculation. Based on the accurate prediction, the scheduling and energy-saving strategies for base stations are proposed, which provide ideas for the subsequent work.

In future work, we plan to investigate two questions. First, find a metric that can distinguish the noise sequence and the trend sequence to remove the noise completely. Second, carry out more in-depth research on network traffic scheduling, so that network traffic prediction and scheduling can be fully combined to solve the problem from a systematic perspective.

REFERENCES

[1] K. Sundaresan, "5G: An evolution towards a revolution," in *Proc. 24th Annu. Int. Conf.*, Oct. 2018, p. 659.

[2] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, Nov. 2014.

[3] W. Xu, G. Zhang, and B. Bo, "Ten key ICT challenges in the post-Shannon era," *Scientia Sinica Mathematica*, vol. 51, no. 7, pp. 1095–1138, May 2021.

[4] M. Kang, J. Song, and P. Fan, "Survey of network traffic forecast based on deep learning," *Comput. Eng. Appl.*, vol. 57, no. 10, pp. 1–9, Oct. 2021.

[5] W. Jiang, "Cellular traffic prediction with machine learning: A survey," *Exp. Syst. Appl.*, vol. 201, Sep. 2022, Art. no. 117163.

[6] C. Pan, Y. Wang, H. Shi, J. Shi, and R. Cai, "Network traffic prediction incorporating prior knowledge for an intelligent network," *Sensors*, vol. 22, no. 7, p. 2674, Mar. 2022.

[7] N. Jiang, Y. Deng, O. Simeone, and A. Nallanathan, "Online supervised learning for traffic load prediction in framed-ALOHA networks," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1778–1782, Oct. 2019.

[8] Y. Ji, Y. Wu, D. Zhang, Y. Yuan, S. Liu, R. Zarei, and J. He, "A novel flash P2P network traffic prediction algorithm based on ELMD and garch," *Int. J. Inf. Technol. Decis. Making*, vol. 19, no. 1, pp. 127–141, Jan. 2020.

[9] C. Katris and S. Daskalaki, "Dynamic bandwidth allocation for video traffic using FARIMA-based forecasting models," *J. Netw. Syst. Manage.*, vol. 27, no. 1, pp. 39–65, Jan. 2019.

[10] A. Sang and S.-Q. Li, "A predictability analysis of network traffic," *Comput. Netw.*, vol. 39, no. 4, pp. 329–345, Aug. 2002.

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

[12] L. Nie, X. Wang, S. Wang, Z. Ning, M. S. Obaidat, B. Sadoun, and S. Li, "Network traffic prediction in industrial Internet of Things backbone networks: A multitask learning mechanism," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7123–7132, Oct. 2021.

[13] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 1566–1579, Feb. 1995.

[14] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[15] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106181.

[16] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110227.

[17] Z. Zhang, X. Li, and S. Yang, "Short-term prediction of significant wave height based on AR-SVR model," *Acta Energiae Solaris Sinica*, vol. 42, no. 7, pp. 15–20, Aug. 2021.

[18] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London A, Containing Papers Math. Phys. Character*, vol. 226, no. 1927, pp. 267–298, Apr. 1927.

[19] Z. Duan, E. Sun, and Y. Zhang, "The prediction of network traffic based on ARMA," *J. China Academy Electoronics Inf. Technol.*, vol. 4, no. 4, pp. 352–356, Aug. 2009.

[20] R. Zhang and C. Zhao, "Application research on network traffic prediction base on ARIMA," *Comput. Simul.*, vol. 28, no. 2, pp. 171–174, Feb. 2011.

[21] Y. Shu, Z. Jin, and L. L. Zhang, "Traffic prediction using FARIMA models," in *Proc. IEEE Int. Conf. Commun.* Vancouver, BC, Canada, Jun. 1999, pp. 891–895.

[22] H. Shin, H. Roth, and M. Gao, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[23] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[24] X. Chen, Y. Liu, and J. Zhang, "Traffic prediction for Internet of Things through support vector regression model," *Internet Technol. Lett.*, vol. 5, no. 3, p. 336, May 2022.

[25] T. P. Oliveira, J. S. Barbar, and A. S. Soares, "Computer network traffic prediction: A comparison between traditional and deep learning neural networks," *Int. J. Big Data Intell.*, vol. 3, no. 1, pp. 28–37, 2016.

[26] W. Zhao, Y. Gao, T. Ji, X. Wan, F. Ye, and G. Bai, "Deep temporal convolutional networks for short-term traffic flow forecasting," *IEEE Access*, vol. 7, pp. 114496–114507, 2019.

[27] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[28] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[29] K. Dominique and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[30] Y. Zhao, Nurbol, and S. Whshour, "Prediction method of network traffic based on timing EEMD," *Comput. Simul.*, vol. 35, no. 11, pp. 466–469, Nov. 2018.

[31] Y. Li, B. Tang, and S. Jiao, "Optimized ship-radiated noise feature extraction approaches based on CEEMDAN and slope entropy," *Entropy*, vol. 24, no. 9, p. 1265, Sep. 2022.

[32] Y. Li, B. Tang, X. Jiang, and Y. Yi, "Bearing fault feature extraction method based on GA-VMD and center frequency," *Math. Problems Eng.*, vol. 2022, pp. 1–19, Jan. 2022.

[33] M. Civera and C. Surace, "A comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark," *Sensors*, vol. 21, no. 5, p. 1825, Mar. 2021.

[34] Z. Li, L. Gao, W. Lu, D. Wang, H. Cao, and G. Zhang, "A novel noise suppression and artifact removal method of mechanomyography based on RLS, IGWO-VMD, and CEEMDAN," *J. Sensors*, vol. 2022, Sep. 2022, Art. no. 4239211.

[35] Y. Li, B. Tang, and Y. Yi, "A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy," *Appl. Acoust.*, vol. 196, Jun. 2022, Art. no. 108899.

[36] M. Li, Y. Wang, Z. Wang, and H. Zheng, "A deep learning method based on an attention mechanism for wireless network traffic prediction," *Ad Hoc Netw.*, vol. 107, Oct. 2020, Art. no. 102258.

[37] H. Xiong, J. Liu, and J. Wang, "Network traffic prediction based on neural network and autoregressive model," *J. Comput. Appl.*, vol. 41, pp. 180–184, Jun. 2021.

[38] N. W. Branco, M. S. M. Cavalca, S. F. Stefenon, and V. R. Q. Leithardt, "Wavelet LSTM for fault forecasting in electrical power grids," *Sensors*, vol. 22, no. 21, p. 8323, Oct. 2022.

[39] J. Bi, X. Zhang, and H. Yuan, "A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM," *IEEE Trans. Autom. Sci. Eng.*, vol. 99, no. 3, pp. 1–11, May 2021.

[40] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[41] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[42] F. Zhou, Z. Huang, and C. Zhang, "Carbon price forecasting based on CEEMDAN and LSTM," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118601.

[43] L. Yin and Y. Wu, "Traffic flow combination prediction model based on improved VMD-GAT-GRU," *J. Electron. Meas. Instrum.*, vol. 36, no. 7, pp. 62–72, Jul. 2022.

[44] Y. Bin, Y. Yang, F. Shen, N. Xie, H. T. Shen, and X. Li, "Describing video with attention-based bidirectional LSTM," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2631–2641, Jul. 2019.

[45] S. Zhou, C. Song, T. Wang, X. Pan, W. Chang, and L. Yang, "A short-term hybrid TCN-GRU prediction model of bike-sharing demand based on travel characteristics mining," *Entropy*, vol. 24, no. 9, p. 1193, Aug. 2022.

[46] D. Guo, X. Xia, L. Zhu, and Y. Zhang, "Dynamic modification neural network model for short-term traffic prediction," *Proc. Comput. Sci.*, vol. 187, pp. 134–139, Jun. 2021.

[47] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[48] X. Zhao, H. Chen, and W. Lai, "Prediction of mobile network traffic by SVR with optimized parameter," *Comput. Appl. Softw.*, vol. 39, no. 9, pp. 278–284, Sep. 2022.

[49] Z. Tian, S. Li, Y. Wang, and X. Wang, "Network traffic prediction based on ARIMA with Gaussian process regression compensation," *J. Beijing Univ. Posts Telecommun.*, vol. 40, no. 6, pp. 65–73, Jun. 2017.

[50] Y. Lu, X. Wang, and F. Li, "Dynamic load balancing and energy saving mechanism in software defined networking," *Chin. J. Comput.*, vol. 43, no. 10, pp. 1969–1982, Oct. 2020.

[51] L. Wang, "Design and simulated implementation of a SDN-based traffic prediction and scheduling mechanism," M.S. thesis, School Comput. Sci. Eng., Northeastern Univ., Liaoning, China, 2018.

[52] D. Ren, Z. Zuo, and H. Chen, "Short-term passenger flow prediction of urban rail transit based on SSA-LSTM," *J. Wuhan Univ. Technol.*, vol. 44, no. 2, pp. 44–52, Feb. 2022.

DONG WANG received the B.S. degree from the Wuhan University of Science and Technology, in 2020. He is currently pursuing the M.S. degree in mathematics with the Wuhan University of Technology. His research interests include deep learning and network traffic prediction.

YU-YANG BAO received the B.S. degree from the Wuhan University of Technology, in 2021, where he is currently pursuing the B.S. degree in applied statistics. His research interests include statistical analysis and network traffic prediction.

CHUAN-MEI WANG received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2008. She is currently an Associate Professor with the School of Science, Wuhan University of Technology. Her research interests include machine learning and applied statistics and management.