International Journal of Energy
ISSN: 2957-9473 | Vol. 3, No. 3, 2023

Short-term Wind Power Prediction based on CEEMDAN-TCN

Ao Wang, Chenglin Feng

North China Electric Power University (Baoding), Baoding, China

Abstract: In order to improve the accuracy of wind power prediction, a wind power prediction method based on time series decomposition and error correction is proposed in this paper. Firstly, the maximum information coefficient (MIC) method is used to select the features with strong correlation with wind power, so as to reduce the complexity of the original data; Then, according to the non-stationary characteristics of wind power, the wind power is decomposed into several stationary subsequences by using adaptive noise complete set empirical mode decomposition (CEEMDAN); Finally, the time convolution network (TCN) is used to dynamically model the multivariable time series of wind power; In order to further improve the prediction accuracy, a light quantization hoist (LightGBM) is introduced to correct the error of the prediction value. The simulation results show that the proposed method has higher short-term wind power prediction accuracy than other prediction models.

Keywords: Wind Power Prediction; CEEMDAN; Temporal Convolution Network; Error Correction; MIC; LightGBM.

1. Introduction

With the increasing penetration of wind power in the power system, wind power has become one of the main power sources in the power grid [1-2], but the volatility and intermittency of wind power easily bring adverse effects on the power quality, economic efficiency and safe operation of the power system [3]. Accurate prediction of wind power can optimize power system scheduling, promote wind power consumption and enhance power system stability, which is of great practical significance for the development of wind power [4].

According to the prediction time scale and the corresponding control strategy, wind power forecasting can be divided into ultra-short-term, short-term, medium-term, and long-term forecasting [5]. The results of wind power short-term forecasting are important for optimizing unit combination schemes and developing power unit generation plans for wind farms [6].

In the study of wind power short-term prediction, the time-series decomposition algorithm is used to improve the prediction accuracy for the non-smooth characteristics of wind power data. Among them, the literature [7] proposed empirical modal decomposition and set-pair analysis model, and wind power was divided into three components of high frequency, medium frequency and low frequency after EMD decomposition, and the prediction model was established for the characteristics of the three components, which effectively improved the prediction accuracy, but the modal mixing problem may occur in the EMD decomposition process. In the literature [8], complementary empirical modal decomposition (CEEMD) is used for wind power, and the LSSVR prediction model optimized by the satin blue gardener algorithm is established for wind power prediction, which effectively improves the prediction accuracy. The literature [9] used the ensemble empirical modal decomposition (EEMD) to decompose the wind power series into subseries with different feature scales, and then established the SVM of whale optimization algorithm for the subseries, and finally superimposed the prediction values of the subseries to obtain the prediction results.

The research work on wind power prediction mainly focuses on improving the prediction accuracy of the model, ignoring the role of error in wind power prediction, and there is less research on model prediction error. Among them, the literature [10] used XGBoost to classify the wind speed and power errors into three categories: low wind speed error, medium wind speed error and high wind speed error, and finally the XGBoost model prediction results were added with the power error prediction to obtain the final prediction value. The literature [11] proposed the idea of stratifying the wind power prediction error for the statistical characteristics of the prediction error spikes and light tails, and gave the corresponding compensation scheme for the prediction errors in different cases. In the literature [12], after phase space reconstruction of wind power series, a VAR model is established to extract linear features, and an Elman model is established to extract nonlinear features for the residual series, and the predicted values of the two models are summed to obtain the final predicted values.

Based on the above problems, this paper proposes a short-term wind power short-term prediction method based on fully adaptive noisy ensemble modal empirical decomposition (CEEMDAN), temporal convolutional network (TCN) prediction and light gradient boosting machine (LightGBM) error correction. The wind power is decomposed into several subseries using CEEMDAN, a wind power prediction model with TCN of the subseries is established, and after superimposing the predicted values of the subseries, the error correction is applied to the predicted values using LightGBM to obtain the final predicted values. The proposed method is compared with TCN, LSTM and EMD-TCN, and it is verified that the proposed model in this paper has higher accuracy.

2. Data Preprocessing

2.1. MIC Feature Selection

Reshef et al. proposed the maximum information coefficient (MIC) method basedon mutual information [13]. Mutual information is a measure that describes the mutual relationship between two random variables x and y. The greater the mutual information value, the stronger the correlation between the two variables. Mutual information is defined as:

$$I(x,y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

In the formula: p(x,y) is the joint probability density function of x, y; p(x) and p (y) are edge density functions of x and y, respectively.

The maximum information coefficient is defined as follows: assuming that D(x,y) is a finite two-dimensional data set, the current two-dimensional space was divided into m intervals and n intervals in the x and y directions, respectively, to form a grid of m * n. There are multiple ways to divide grids with the same interval [14]. Under grids of different scales, the maximum mutual information value was selected for calculation, and the maximum information coefficient is defined as:

$$MIC(x,y)=\max_{m\times n<B}\frac{I(x,y)}{\log(\min(m,n))} \tag{2}$$

In the formula, m*n<B represents the constraint condition of the total number of grid divisions. It can be seen from literature[14] that the optimal effect can be achieved when it is set to the 0.6th power of the total amount of data. Therefore, this setting method was also used in this paper.

2.2. Normalization

When using multivariate time series to predict wind power, the dimensions of different variables are different, and the numerical differences are also large [15]. In orderto make the prediction model consider the effect of variables on wind power equally, the wind speed, temperature, wind direction angle and active power in the wind power data are normalized to the [0,1] range.

$$X_{norm}=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{3}$$

In the formula: Xmin and Xmax are the minimum and maximum values of variables respectively.

3. Prediction Model Analysis

3.1. CEEMDAN Principle

CEEMDAN is based on EMD decomposition. In the process of decomposing wind power signals, it adaptively adds white noise to weaken the mode aliasing problem, and the decomposition process is characterized by integrity and almost no reconstruction error [16].

The CEEMDAN decomposition algorithm is implemented as follows:

1)CEEMDAN decomposition algorithm makes use of EMD for signal X(t)+εi0n(t)is decomposed repeatedly for N times, and the first modal component is obtained through mean value calculation:

$$\overline{IMF_1(\mathrm{t})}=\frac{1}{N}\sum_{i=1}^{N}IMF_1^i(t) \tag{4}$$

2)Calculate the first residual signal r1(t) decomposed by CEEMDAN as:

$$r_1(t)=X(t)-\overline{IMF_1(t)} \tag{5}$$

3)The signal is decomposed repeatedly for N times to obtain the second modal component:

$$\overline{IMF_2(t)}=\frac{1}{N}\sum_{i=1}^{N}E_1(r_1(t)+\varepsilon_1E_1(n^i(t))) \tag{6}$$

4)For k=2,...,K, calculate the kth residual signal:

$$r_k = r_{k-1} - \overline{IMF_K(t)} \tag{7}$$

5)Repeat the calculation process in step 3) to obtain the k+1 modal function:

$$\overline{IMF_{k+1}(t)}=\frac{1}{N}\sum_{i=1}^{N}E_1(r_k(t)+\varepsilon_kE_k(n^i(t))) \tag{8}$$

6)Repeat steps 4) and 5) until the residual signal meets the termination condition of decomposition, and finally obtain k modal components. The final residual signal decomposed is:

$$R(t)=X(t)-\sum_{k=1}^{k}\overline{IMF_k(t)} \tag{9}$$

The original signal can be decomposed into:

$$X(t)=R(t)+\sum_{k=1}^{k}\overline{IMF_K(t)} \tag{10}$$

3.2. TCN Wind Power Prediction Model

Temporal Convolutional Networks (TCN) is an algorithm that uses a convolutional structure to solve time series problems, mainly consisting of dilated causal convolution and residual connectivity modules. The advantage of TCNs over RNNs is that they can complete operations in parallel as well as avoid the problem of gradient disappearance or explosion [17].

3.2.1. Dilated Causal Convolutions

The TCN uses causal convolution to ensure that no future information is leaked and that the input Xt at the current moment t is only related to the inputs X0,X1,X2...Xt-1 at past moments [18].

Since causal convolution is limited by the linear size of the network depth in time series modeling, dilated convolution is introduced to achieve exponentially larger sensory fields that can capture long time scale dependencies in time series.

The expanded causal convolution operation for sequence elements is：

$$F(s)=\sum_{i=0}^{k-1}f(i)\cdot X_{s-d\cdot i} \tag{11}$$

Where: $F(s)$ represents the output of the network during the dilation convolution calculation; $f(i)$ is the filtering operation on the ith input; k is the size of the filtering window; d is the dilation factor; and is the past input.

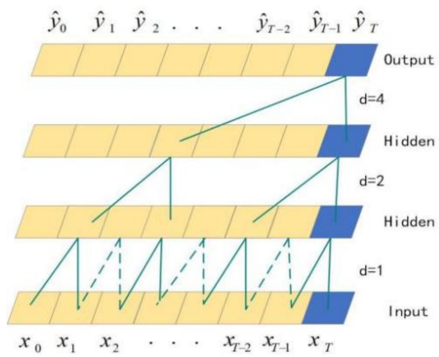

Figure 1. Extended convolution

3.2.2. Residual Connection

To a certain extent, the performance of neural networks is enhanced with the increase of network layers, but too many network layers can lead to gradient disappearance or gradient explosion, and the residual connection effectively solves the problem of accuracy degradation due to too many layers in neural networks [19].

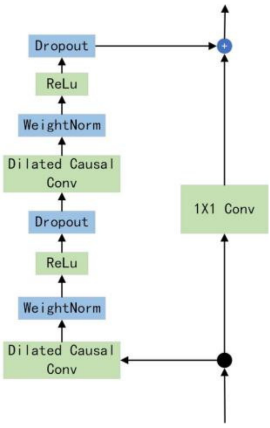

Figure 2. Residual connection

The residual connection of the TCN is shown in Figure 2. Within the residual block, the WeightNorm module normalizes the weight values; Relu is the activation function in the network; Dropout module is to prevent the network from overfitting; and Dilated Causal Conv is the expansion causal convolution.

3.3. LightGBM Wind Power Error Correction

Gradient Boosting Decision Tree (GBDT) does not work well in the case of high feature dimensionality and large amount of data.

Gradient-based One-Side Sampling (GOSS) is used in LightGBM to retain sample points with large gradients and randomly reject sample points with small gradients to obtain more accurate gain estimates compared to the uniform random sampling method, and the use of Exclusive Feature Bundling, EFB) after sorting the features by mutual exclusivity, merges the features by adding offsets to the feature values to reduce the data feature dimensionality, which effectively improves the training speed and accuracy [20].

The LightGBM algorithm is implemented as follows:

1)The purpose of each iteration is to obtain a weak learner $h_t(x)$ that minimizes the loss function $L(y,F_t(x))$ of the iterations [21].

$$L(y,F_t(x))=L(y,F_{t-1}(x)+h_t(x)) \tag{12}$$

where: $F_{t-1}(x)$ is the strong learner of the last iteration and $L(y,F_t(x))$ the loss function of this iteration.

2)The negative gradient obtained in the above equation is used to fit the loss approximation of this iteration.

$$r_{ti} = \frac{\partial L(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)} \tag{13}$$

3)The weak learner $h_t(x)$ is fitted using the squared difference approximation.

$$h_t(x) = \arg\min_{h\in H} \sum (r_{ti} - h(x))^2 \tag{14}$$

4)The final strong learner $F_t(x)$ is obtained as follows:

$$F_t(x) = h_t(x) + F_{t-1}(x) \tag{15}$$

4. Wind Power Prediction Model

4.1. Forecast Process

The flow of short-term wind power prediction method proposed in this paper is as follows:

(1) Wind power datasets A and B use the MIC method for feature selection of wind power data to select features with strong correlation with wind power, which is conducive to speeding up network training;

(2) Based on CEEMDAN decomposition algorithm, wind power time series data was decomposed to obtain several stable sub sequences;

(3) TCN network prediction model is established for each subsequence. After training, the prediction results of each subsequence were obtained, and the prediction values were obtained after superposition.

(4) Where dataset B is divided into training set and test set according to 7:3, the predicted value $\hat{y}$ is obtained by decomposing the prediction, and its true value $y-\hat{y}$ is obtained as the error value, and the error dataset is obtained by combining the wind power variable with the error value.

(5) The features of the initial prediction values of dataset A were input into the error correction model LGB to obtain the prediction error values;

(6) The initial prediction value and the prediction error value are added to obtain the final prediction value of the wind power dataset A.

4.2. Evaluation Index

In this paper, three evaluation indicators, mean absolute error (MAE), mean relative error (MRE) and root mean square error (RMSE), are used to analyze the wind power prediction results. The evaluation indicators are as follows:

$$MAE = \sum_{t=1}^{n} \frac{|p_t - \hat{p}_t|}{np_N} \tag{16}$$

$$MRE = \sum_{t=1}^{n} \frac{|p_t - \hat{p}_t|}{np_t} \tag{17}$$

$$RMSE = \frac{1}{p_N}\sqrt{\sum_{t=1}^{n} \frac{(p_t - \hat{p}_t)^2}{n}} \tag{18}$$

In the formula: n is the total number of forecasts; P_N is the capacity of wind turbine; P_t is the Actual power of wind turbine; $\hat{p}_t$ is the predictive value of wind turbine power.

5. Example Analysis

5.1. Data Preparation

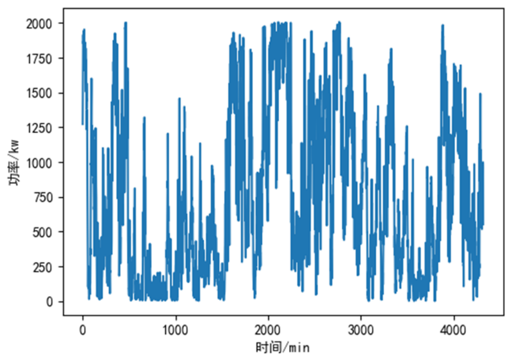

Figure 3. Raw data

Simulation experiments are carried out using measured data from a wind farm in China, where two adjacent wind turbines are selected and divided into Wind Turbine A and Wind Turbine B. The data collected from the wind farm include wind turbine ID, wind speed, wind direction angle and ambient temperature, with a sampling period of 10 min.

The SCADA data of wind generators A and B for 30 consecutive d in June 2016 are used to predict the wind power of wind generator A on 1 July. Where the data of wind turbine B is predicted by the model to form an error set to error correct the wind power of wind turbine A. The raw data of wind turbine A was shown in Figure 3.

The wind speed, temperature, air pressure and wind direction of neighboring wind turbines in a wind farm are affected by similar factors, and there is a correlation between the output power.

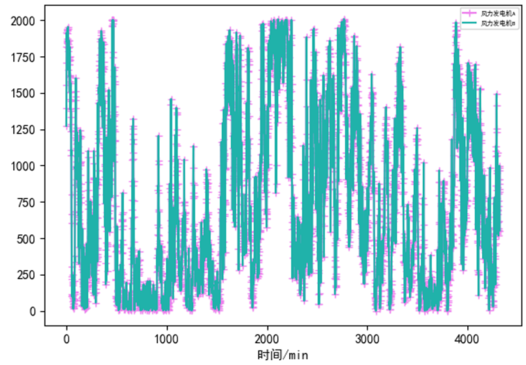

Figure 4. Change in power output of the two wind turbines in June

As shown in Figure 4, there is a strong similarity in the output power volatility of the two adjacent wind turbines in the wind farm, demonstrating that the prediction errors of the adjacent wind turbines can be used to error correct the predicted values of the wind turbines.

5.2. MIC Feature Selection

There are many types of parameters of wind power SCADA data set recording units. If all dimensions of the data set are input into the prediction model, a large number of parameters will be generated in the network model, and the training speed will be slow, affecting the prediction accuracy. The maximum information coefficient method was used for feature selection of wind turbine multivariable time series to reduce the data scale and complexity.

Table 1. MIC feature selection results

influence factor	*WS*	*GS*	*RS*	*WD*	*GT*	*T*
MIC coefficient	0.982	0.965	0.971	0.250	0.810	0.143

In Table 1, six variables with the strongest correlation with output power were selected by using the maximum information coefficient method for feature selection. *WS* represented wind speed; *GS* represented the generator speed; *RS* stands for impeller speed; *WD* represented wind direction angle; *GT* represented gearbox temperature; *T* were the ambient temperature.

5.3. CEEMDAN Decomposition Results

In this paper, the CEEMDAN algorithm is used to decompose the wind power time series data by adding the number of white noise groups NR=100, the noise standard deviation Nstd=0.05, the maximum number of iterations MaxIter=300, and the decomposition yields 10 subsequences, and the first 5 subsequences are shown in Fig5.

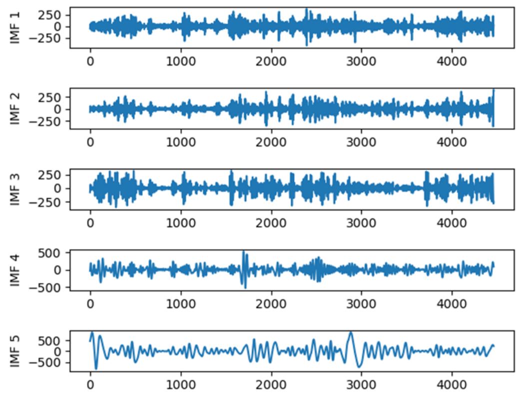

Figure 5. Raw data

5.4. Model Parameter Settings

The Hidden of TCN is set to 8 layers, the dilation factor d=[1,2,4,8,16,32,64,128], and the convolution kernel size is set to 2. The field of view of the prediction model can feel all the data points of the wind turbine in a day.

In order to accelerate the training speed of LightGBM, enhance the error correction ability, and prevent overfitting, the main parameters of LightGBM are tuned by grid searching using GridSearchCV integrated in the machine learning library sklear.

5.5. Results

Figure 6 shows the July 1 power change curve of wind turbine A predicted by each model, with a prediction resolution of 10 min and a total of 144 data points.

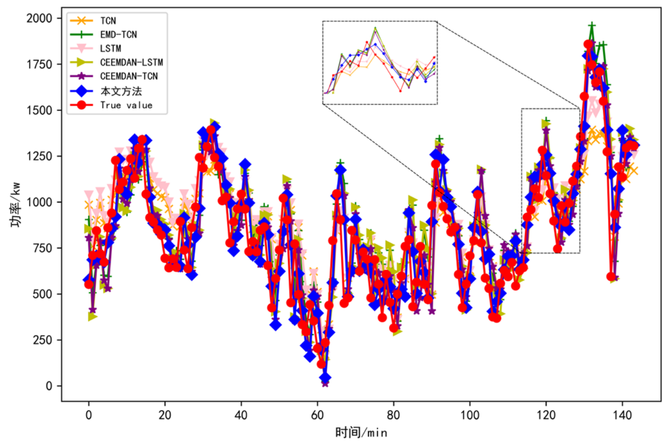

Figure 6. The effect of each prediction model

The MAE, MRE and RMSE indexes of each model were calculated for evaluation and analysis, as shown in Table 2.

Table 2. Evaluation index of each prediction model

Model name	MAE/%	MRE/%	RMSE/%
LSTM	8.96	25.64	10.74
TCN	8.57	26.41	10.41
EMD-TCN	8.48	25.44	10.21
CEEMDAN-LSTM	8.68	25.94	10.66
CEEMDAN-TCN	7.86	24.90	10.05
Proposed Method	5.83	17.37	7.58

After training the model by the same data, it can be seen from Table 2 that the time series decomposition algorithm can effectively reduce the effect of wind power uncertainty, and the prediction accuracy becomes higher after decomposition.

Among them, taking TCN as the benchmark model, CEEMDAN has better prediction effect than EMD, and the MAE, MRE and RMSE are reduced by 0.62%, 0.54% and 0.16% respectively, mainly because there is no modal aliasing problem during decomposition. The evaluation indexes show that the error correction of this paper's method can effectively reduce the prediction error when comparing the models such as LSTM, CEEMDAN-TCN and EMD-TCN. In summary, the prediction model algorithm proposed in this paper has higher prediction accuracy compared with other models.

6. Conclusion

A wind power prediction model based on CEEMDAN-TCN and error correction is proposed. The proposed model has a more accurate prediction effect as verified by the measured data of a wind farm in China. The main conclusions are as follows:

(1) The CEEMDAN algorithm is used to decompose the wind power time series data with respect to the fluctuation characteristics of wind power, and the CEEMDAN algorithm effectively solves the problem of modal aliasing in the EMD and improves the prediction accuracy of the prediction model.

(2) TCN can change the field of view of the network by changing the size of the Hidden layer and convolutional kernel, while for LSTM, if the time series is too long, the network will forget the previous information. For long time series prediction, TCN has higher prediction accuracy than LSTM.

(3) For the problem of prediction error of the prediction model itself, using LightGBM for error correction can further improve the prediction accuracy, and the method in this paper decreases the MAE, MRE and RMSE decomposition by 2.03%, 7.53% and 2.47% compared with CEEMDAN-TCN.

References

[1] Sun Rongfu, Zhang Tao, et al. Overview of key technologies and applications of wind power prediction[J]. High Voltage Technology, 2021,47(04):1129-1143.

[2] Zhang Yachao, Liu Kaipei,et al. Study on Multi time Scale Source Load Coordinated Dispatching Model of Power System with Large Scale Wind Power[J]. High Voltage Technology, 2019,45(02):600-608.

[3] Tao Yubo, Chen Hao, et al. Short term wind power prediction concept, model and method[J]. Power Engineering Technology, 2018, 37 (05):7-13.

[4] Zhao Dongmei, Du Gang, et al. Wind power combination prediction model based on time series decomposition and machine learning [J/OL]. Modern Power,1-11[2021-10-21].

[5] Liu Xing, Wang Yan, JI Zhicheng. A short-term prediction method for wind power based on random forest[J/OL]. Journal of System Simulation,1-10[2021-10-30].

[6] Ding Tingting,Yang Ming,Yu Yixia,Sz Zhiyuan,Zhang Qiang. Short-term wind power integration prediction method based on error correction[J/OL].High Voltage Engineering, 1-11[2021-10-30].

[7] Yang Mao,Chen Yulin. Real-Time Prediction for Wind Power Based on EMD and Set Pair Analysis[J]. Transactions of China Electrotechnical Society,2016,31(21):86-93.

[8] Zhou Xiaolin,Tong Xiaoyang. Ultra-short-term wind power combination prediction based on CEEMD-SBO-LSSVR[J]. Power System Technology,2021,45(03):855-864.

[9] Yue Xiaoyu,Peng Xianggang,Lin Li. Whale optimization support vector machine for short-term wind power prediction [J]. Proceedings of the CSU-EPSA,2020,32(02): 146-150.

[10] Ding Tingting,Yang Ming,Yu Yixia,et al. Short-term wind power integration prediction method based on error correction[J/OL]. High Voltage Engineering,1-11[2021-10-21].

[11] Liu Shuai,Zhu Yongli,Zhang Ke, et al. Short-term wind power prediction based on error-corrected ARMA-GARCH model[J]. Acta Energiae Solaris Sinica,2020,41(10):268-275.

[12] Wang Lan, Wang Hei, Li Huaqiang, Diao Fangyu. Chaotic time series prediction model for wind power based on phase space reconstruction and error compensation[J].Proceedings of the CSU-EPSA,2017,v.29;No.164(09):65-69.

[13] Reshef D N,Reshef Y A,Finucane H K,et al.Detecting Novel Associations in Large Data Sets[J].Sicence,2011, 334 (6062): 1518-1524.

[14] Zheng Ruicheng,Gu Jie,Jin Zhijian,Peng Hongqiao,Cai Long. Research on short-term load forecasting input variable selection method by fusion of data-driven and forecast error-driven[J].Proceedings of the CSEE,2020,40(02):487-500.

[15] Zhu Qiaomu,Li Hongyi,Zi Qi et al. Ultrashort-term prediction of wind farm generation power based on long and short-term memory networks[J].Power System Technology,2017, 41(12): 3797-3802.

[16] Shu Chang,Jin Xiao,Li Zipin,et al. Noise diagnosis method of distribution transformer discharge fault based on CEEMDAN [J]. High Voltage Engineering,2018,44(08):2603-2611.

[17] Bai Shaojie, J.Zico Kolter, and VIadlen Koltun.An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.arXiv:1803.01271(2018).

[18] Li Zikang, Liu Hao,Bi Tianshu,et al. Data-driven robust power system disturbance identification[J/OL]. Proceedings of the CSEE,1-15[2021-10-26].

[19] He K ,Zhang X ,Ren S,et al. Deep Residual Learning for Image Recognition[C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.

[20] Ke G,Meng Q,Finley T,et al.LightGBM: A Highly Efficient Gradient Boosting Decision Tree[G]. Guyou I,Luxburg U V,Bengio S,et al.Advances in Netural Information Processing System30(NIPS2017).LongBeach:NIPS,2017:3146-3154.

[21] Wu Qiong, Li Ronglin, Hong Haisheng,et al. Low-voltage trip prediction of distribution substation based on hybrid resampling and LightGBM algorithm[J].Power System Protection and Control ,2021,49(12):71-78.