An Improved Deep Neural Netowrk for Operation and Maintenance Prediction of Ultra High Voltage UHV Substation

Zhenyu Guo[1], Jinrui Gan[2], Xin Liu[1], Haibin Zhang[1,a,*], Long Chen[3], Jinjin Ding[1], Bin Zhang[4]

[1]State Grid Anhui Electric Power Co., Ltd., Hefei, China

[2]State Grid Smart Grid Research Institute Co., Ltd., Beijing, China

[3]Nanjing Nanrui Information Communication Technology Co., Ltd., Nanjing, China

[4]State Grid Intelligence Technology Co., Ltd., Jinan, China

[a]haibinzhang_sgcc@126.com

*Corresponding author

Keywords: UHV substation maintenance cost prediction; deep neural network; genetic algorithm; K-fold cross-validation

Abstract: The maintenance and operation costs of UHV substations are affected by many complex factors, and the maintenance cost data records are ambiguous and volatile. To solve the problem of unclear maintenance cost record is unknown, firstly, UHV substation maintenance entries are divided and processed using horizontal and vertical data analysis methods, and then deep neural network is used to predict maintenance costs. The deep neural network is then used to predict the overhaul costs. To improve the prediction accuracy of deep neural network, K-fold cross-validation is used to precisely adjust the original data training model, and genetic algorithm is applied to predict the deep neural network. Genetic algorithm is used to adjust and improve the initial values and threshold values of the deep neural network, to establish an improved deep neural network based on genetic algorithm for maintenance and operation cost prediction method based on genetic algorithm. The proposed method can effectively improve the accuracy of the model prediction. The comparison analysis shows that the proposed method can effectively improve the accuracy of the model prediction, and thus provide reference value for the grid to allocate maintenance costs to UHV substations.

1. Introduction

The UHV substation project is a costly and complex system, and the study of UHV substation project cost is generally analysed by life cycle cost (LCC). The UHV substation maintenance and operation cost is a part of LCC, which is the sum of all costs incurred during the operation of the UHV substation, including primary and secondary equipment, all line maintenance and operation costs and other costs in the station. And in the total assets of the UHV substation project accounted for a relatively large proportion, the accurate prediction of the maintenance costs has research significance.

Literature [1] made fuzzy prediction of reliability parameters of electrical equipment during the life cycle and established an optimization decision model of maintenance strategy. Literature [2] analysed the reliability of different categories of maintenance methods based on intelligent UHV substations, and provided reference for subsequent maintenance of intelligent UHV substations; literature [3] proposed a new optimization strategy of UHV substation equipment maintenance, which can effectively reduce UHV substation The literature [3] proposes a new optimization strategy for UHV substation equipment maintenance, which can effectively reduce the risk of power supply, improve the efficiency of equipment maintenance, and reduce its maintenance cost; the literature [4] minimizes UHV substation LCC by controlling the digital cost while improving the integration of the project and using the whole life cycle cost theory; the literature [5] analyses from the perspective of the lowest whole life cycle cost of power equipment and establishes a fault maintenance optimization cost model for 220 kV transformer equipment, but only for transformer fault However, the cost

  DOI: 10.25236/icceme.2022.032

optimization is only done for transformer faults, and does not involve the whole process of equipment operation.

Most of the existing predictions of maintenance and operation costs are reflected in one part of the whole life cycle cost of the equipment, and the maintenance and operation costs are obtained when the LCC is optimal under the condition of ensuring power supply reliability. Since LCC involves initial investment costs, this approach is only applicable to unbuilt or new UHV substation projects, not to built UHV substation projects. Moreover, in actual projects, the maintenance costs are allocated at UHV substation level and do not reach the equipment level directly. Therefore, the paper aims at prediction and analysis of the overhaul cost of the built UHV substation, by using the data of financial entries to classify and group the overhaul cost, and get the specific cost categories of overhaul cost according to the data collation and analysis, and use K-fold cross-validation to pre-process the data, and then based on the improved deep neural network of genetic algorithm to The data are pre-processed using K-fold cross-validation, and then improved deep neural network based on genetic algorithm for maintenance cost prediction to obtain accurate maintenance cost, which provides practical reference significance for power grid allocation of UHV substation maintenance cost.

2. Operation and maintenance cost analysis

2.1. Operation and maintenance cost components

The total cost paid during the whole life cycle of the UHV substation project can be divided into initial investment cost, maintenance and operation cost and decommissioning disposal cost [6]. Among them, the initial investment cost can be estimated by the size of the new UHV substation project size of the corresponding information price document policy in different cities, the approximate value; scrapping costs are generally 3% to 5% of the initial investment cost; operation and maintenance costs in the whole life cycle of the cost accounted for a relatively large, generally 60% to 70% of the total cost, and affect the operation and maintenance costs of more factors, the cost fluctuations. The inspection and maintenance cost can be divided into operation cost, maintenance cost and fault outage cost [7-8].

Literature [7-8] are divided and calculated from the perspective of UHV substation equipment, but in the actual grid operation planning, maintenance costs are allocated at the local and municipal UHV substation level, and the specific maintenance cost data cannot be refined to the various equipment levels, so the analysis of UHV substation-level maintenance costs will be more reliable and practical significance.

2.2. Operation and maintenance cost data analysis method

As the UHV substation equipment of the power grid has a wide variety, each parameter of the equipment is not the same, and the quality level is not uniform. In the process of long-term operation, the chances of various types of faults will also increase, and the factors affecting the condition of UHV substation equipment are the operating status, maintenance, the working environment in which the equipment is located and the family quality history [9], the information of these conditions of the equipment is fuzzy and volatile, and the relationship between the variables is redundant and affect each other, therefore, the data of power equipment maintenance operation and maintenance cost records are also fuzzy and volatile of. Nowadays, the workload of grass-roots operation and maintenance is large, and the cost of maintenance and operation is increasing continuously every year. Considering the impact of various factors on the cost of maintenance, it is necessary to forecast the cost of maintenance in UHV substations. To solve the problem of ambiguous recorded data, the paper collates and analyses the UHV substation equipment overhaul cost entries from the financial system as shown in Figure 1, selects a reasonable data sample according to the actual situation, and processes the data using data analysis methods in both horizontal and vertical directions.

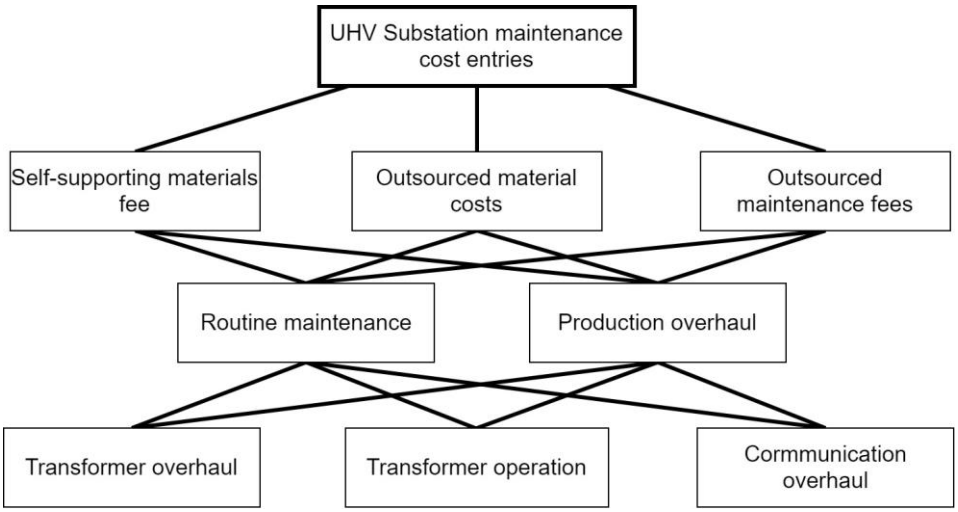

Figure 1: UHV substation overhaul operation and maintenance cost item

The method of dealing with the horizontal direction of the cost of the equipment is usually to take the average of the sum of the costs of the previous and following 2 years. This is due to the fact that in general the costs are smooth and continuous in the time series, i.e. the data of a certain year is close to the data of the year before and after it. If the difference is large, it means that the data of that year may have a large deviation due to some human factors. In this case, the data of that year should be discarded, or level treatment should be taken to correct.

$$\begin{cases} |Y(d,t) - Y(d,t-1)| > \alpha(t) \\ |Y(d,t) - Y(d,t-1)| > \beta(t) \\ Y(d,t) = \frac{Y(d,t)+Y(d,t-1)}{2} \end{cases} \tag{1}$$

As the relevant data of the same model of equipment of the same manufacturer such as family quality history problems and failure rate are clustered, the gap between different individuals should be maintained within a certain range and the error should be small, and if it exceeds its fixed range it is regarded as perturbed data. For the vertical processing method, a reasonable threshold of the data is set first, and if it exceeds this threshold, the data is judged to be wrong and a more reasonable value should be taken for replacement.

$$\begin{cases} |Y(d) - m(t)| > r(t) \\ Y(d,t) = \begin{cases} m(t) + r(t), Y(d,t) > m(t) \\ m(t) - r(t), Y(d,t) < m(t) \end{cases} \end{cases} \tag{2}$$

3. Improved deep neural network based on genetic algorithm

3.1. Deep neural network

deep neural network is essentially a reverse transfer process of the error between the actual value and the predicted value, with the direction of the negative gradient of the error on the derivative of the weights, the weights or thresholds are continuously corrected until the output error is within the set range. Compared with other prediction methods, deep neural network training to obtain the law performance by the modified network parameters, no specific mathematical model, and the relationship between the input parameters can be used without clarity, so the combination of the characteristics of deep neural network can be used for the prediction of maintenance and operation costs.

The deep neural network is shown in Figure 2, X and Y are the input vector of the neural network and the output vector of the hidden layer respectively; X0 and Y0 are the threshold values of the input layer and the hidden layer respectively, which are generally taken as -1; V and W are the weight vectors between the input layer, the hidden layer and the output layer respectively; if the output vector

O in the training result does not meet the convergence condition, V and W are adjusted by the error at this time, so that the output vector meets the condition.

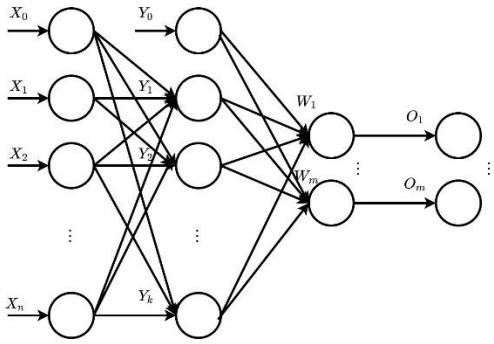

Figure 2: The structure of deep neural network.

3.2. Improved K-fold-GA-Deep neural network

The neural network training data is divided into a training set and a validation set, in order to objectively to objectively evaluate the application of the model to data outside its training set, the The data in the K-fold cross-validation is derived from the training data but not from the neural network. The data in the validation are derived from the training data, but are not involved in the training of the neural network. The data in the K-fold cross-validation are derived from the training data but are not involved in the training of the neural network. First, the initial data are divided into K groups, and each group of data is validated and trained in turn. If one group is set as the validation set, the remaining K-1 groups are validated and trained in turn. If one set is set as the validation set, the remaining K-1 data sets are set as the training set, and the general K value The network is trained for K times, and the evaluation error (mean squared error (MSE) of K times is summed and averaged as the final cross-validation The MSE of K times is the final cross-validation error. The cross-validation method utilizes limited data, which is effective for the accuracy of the model measurement set. The cross-validation method can be used as a model optimization cross-validation.

And the genetic algorithm (GA) is used to preprocess the deep neural network to obtain better initial values of the network. Firstly, the initial weights and thresholds of the deep neural network are preprocessed and represented by individuals in the GA; secondly, the prediction error value of the deep neural network is calculated and set as the individual fitness value; finally, the optimal value of the initial weights of the deep neural network is obtained by the operation of individual fitness selection, crossover and variation to meet the set accuracy range.

K-fold cross-validation is adopted to pre-process the data.

4. K fold- for UHV substation maintenance cost prediction GA-Deep neural network design

The structure of a deep neural network includes an input layer, an implicit layer and an output layer, where the input parameters enter the input layer, then pass through the implicit layer, and finally are output by the output layer [7]. Using K-fold-GA-deep neural network method he steps of UHV substation maintenance cost prediction are as follows.

1) Design of input and output layers. Based on the design characteristics of deep neural network the cost components of the UHV substation operation and maintenance cost can be selected as the input. The cost composition factors can be selected as input the output layer is the predicted cost value of maintenance and operation.

2) Selection of target values. According to the built UHV substation maintenance cost, the method of processing data in horizontal and vertical directions is used to estimate the target maintenance cost method to estimate the expected cost of target maintenance cost as the output target value. The target value is the output target value.

3) Model building.

Genetic algorithm part: firstly, the population is initialized using GA The initial weights and thresholds of the deep neural network are obtained. After the initialization, individuals are substituted

into the deep neural network for iteration to find the optimal initial Weights and thresholds.

(i) The weights between the input layer, the implicit layer and the output layer, the implicit layer and the threshold value between the output layer are encoded for the real string population.

(ii) The fitness value F of an individual is composed of its expected output value and the predicted the absolute value of the error of the output value constitutes.

$$F = k(\sum_{i=1}^{m}|p_i - o_i|) \tag{3}$$

Based on the results of individual fitness calculated in equation (3), the better individuals to form a new population, which is calculated as:

$$f_i = \frac{k}{F_i} \tag{4}$$

$$P_i = \frac{f_i}{\sum_{i=1}^{k} f_i} \tag{5}$$

Where Fi is the fitness value of the ith population, calculated from equation (3) to, k is the population size, and fi is the probability that population i is selected.

4) Parental generations P_k and P_l are randomly selected in the population, and the real number crossover at the jth point with certain probability is:

$$\begin{cases} P_{kj} = \alpha P_{lj} + P_{kj}(1-\alpha) \\ P_{lj} = \alpha P_{kj} + P_{lj}(1-\alpha) \end{cases} \tag{6}$$

5) By probability calculation, a random sample is drawn from the new population formed individual Pi to be mutated, and the jth point mutation is

$$P_{ij} = \begin{cases} P_{ij} + (P_{ij} - P_{max})\mu\left(1 - \frac{v}{V_{max}}\right), \mu > 0.5 \\ P_{ij} + (P_{min} - P_{ij})\mu\left(1 - \frac{v}{V_{max}}\right), \mu \leq 0.5 \end{cases} \tag{7}$$

6) Check whether the resulting individuals meet the optimal criteria, and for those who cannot meet the criteria The individuals that cannot meet the criteria are checked again for fitness; if they meet the criteria, the optimal individuals are used as the weights and thresholds of the deep neural network, and repeat If the criteria are met, the optimal individuals are used as the weights and thresholds of the deep neural network, and the number of training sessions is repeated.

7) Calculation of the model. MATLAB software is used for model calculation, and in the calculation process, the existing data of UHV substation equipment maintenance and operation costs are used as samples, and GA-deep neural network is used for training; to avoid overfitting situation and to improve the accuracy of GA-B neural network model at the same time, the training data are taken into K -fold cross-validation method is used to determine the accuracy of the model.

The paper selects energy consumption, labour cost, material cost, repair cost, electricity and transportation costs as the input layer, and the output layer is the UHV substation the initial weights are selected for learning, and the corresponding data from the input layer are used to predict the UHV substation operation and maintenance costs. The initial weights are selected for learning, and the UHV substation equipment maintenance and repair costs are predicted from the corresponding data in the input layer. The initial weights are selected for learning, and the UHV substation equipment maintenance and operation costs are predicted from the corresponding data in the input layer. The UHV substation equipment maintenance cost model based on K-fold-GA-deep neural network is shown in Figure 3.

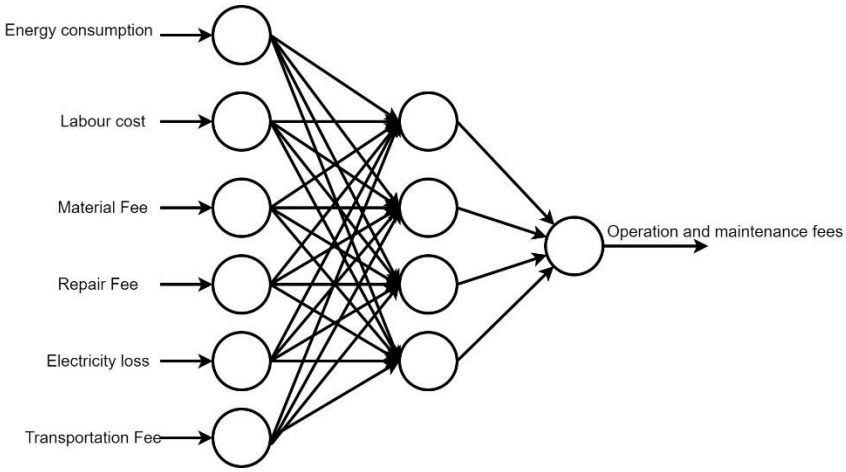

Figure 3: Overhaul operation and maintenance cost model of UHV substation equipment based on K-fold-GA-Deep neural network.

8) Prediction effect evaluation index. When the results of the model are examined When testing the results of the model, the reliability and validity of the model need to be considered. Confidence refers to the reliability of the model. reliability of the model, which is expressed as the reproducibility and stability of the model results. The validity refers to the degree of validity of the model, which is the extent to which the results of the model match with the content to be examined [8]. The validity refers to the degree of validity of the model, which is the extent to which the model results match the content to be examined.

In this paper, eMSE was used to evaluate the effectiveness of K-fold cross-validation on the model. to assess the effect of K-fold cross-validation on model accuracy. The e_{MSE} was used to evaluate the effect of K-fold cross-validation on the accuracy of the model, which was calculated as:

$$e_{MSE} = \frac{1}{n}\sum_{t=1}^{n}(D_t - Y_t)^2 \tag{8}$$

The mean absolute percentage error (e_{MAPE}) and root mean square error (e_{RMSE}) were used to measure the validity of the network, i.e., the predictive validity The calculation formula is:

$$e_{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{D_t - Y_t}{Y_t}\right| \times 100\% \tag{9}$$

$$e_{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(D_t - Y_t)^2} \tag{10}$$

Where D_t is the actual output of the model and Y_t is the expected output. To test the reliability of the model, the paper takes the e_{MAPE}, e_{RMSE}after multiple runs of standard deviation as the final test index.

5. Simulation

5.1. Parameter setting and description

The maintenance cost of a typical local UHV substation in a Chinese province is selected data as an example for analysis. The 2014-2019 local and municipal overhaul allocation data as the data source, and use Matlab simulation to first apply the legacy Genetic algorithm population size is set to 50, the number of iterations is 500, the crossover probability is 0.2, the variation probability is 0.01, and then deep training, the deep neural network hidden layer is set to single The number of nodes in the hidden layer is generally based on the number of nodes in the hidden layer. The number of nodes in the hidden layer is generally selected based on experience. In principle, the value should be as small as possible, and the calculation formula is:

$$L = \sqrt{n + m} + c \tag{11}$$

Where L is the number of implied layer nodes to be set; n, m are the number of nodes in the input and output layers, respectively; c is a constant value between 1 and 10. is a constant value between 1 and 10.

According to equation (10) the trade-off optimum can determine the nodes of the hidden layer number. The network learning rate is set to 0.05, the training error is chosen to be $1\times$ The hidden layer activation function tensing, the output activation function purelin, and weights and thresholds are obtained by genetic algorithm The optimization is done by genetic algorithm.

5.2. Inspection cost forecast analysis

To test the prediction accuracy of maintenance costs, three methods for prediction analysis, and verify each model through prediction effect evaluation effectiveness.

(1) The traditional deep neural network algorithm is used for prediction.

(2) Using GA-deep neural network algorithm based on genetic algorithm to improve algorithm for prediction.

(3) K-fold cross-validation was used to pre-process the training data, and then the GA-deep neural network was used for prediction. (3) K-fold cross-validation is used to pre-process the training data, and then the GA-deep neural network algorithm is used to make predictions.

The results and relative errors of the three different methods for the training of maintenance and operation cost data are shown in the table below. The results and relative errors of the training are shown in Tables 1 and 2. Taking the 2017-year data, for example, the maintenance cost in 2017 is 55.83.7 million the traditional deep prediction is 58.1976 million yuan, and the GA-deep the prediction result is 56.88.921 million yuan, and the prediction result of K-fold-GA-deep is 5.527.372 million yuan.

The relative errors of the three methods are 0.040 70, 0.018 84 and 0.010 09, respectively. The relative errors of the three methods are 0.040 70, 0.018 84 and 0.010 09, respectively. The K-fold-GA-deep method is the best, with a relative error of only 1.009%. 1.009%, while the relative error of the traditional deep prediction is larger, reaching 4.070%. The relative error of the traditional deep prediction value is larger, reaching 4.070%, which indicates that the prediction effect of this method is not good.

Table 1: Trainging reusltls of the operation and mintenance cost data by three methods

Yeah	**Groundtruth**	**Preditictions**		
		deep	GA-deep	K-fold_GA-deep
2015	3877.0	3893.673	3943.525	3874.328
2016	4404.0	4444.600	4339.355	4397.532
2017	5604.4	5423.453	5556.078	5631.609
2018	5583.7	5810.976	5688.921	5527.372
2019	6529.2	6385.458	6473.004	6591.704
2020	7457.9	7511.336	7408.192	7429.191

Table 2: Relative errors of three methods of forecasting results

Yeah	**Relative errors**		
	deep	GA-deep	K-fold_GA-deep
2015	0.004300	0.01716	0.00067
2016	0.009219	0.01468	0.00147
2017	0.032290	0.00862	0.00486
2018	0.040700	0.01884	0.01009
2019	0.022150	0.00861	0.00957
2020	0.007165	0.00299	0.00385

The comprehensive evaluation index data of the training effects of the different models constructed As shown in Table 2, it can be seen that the K-fold-GA-deep prediction models of MAPE and RMSE are 38.063 800 and 0.005 087, respectively. It can be visualized that the comprehensive evaluation indexes with K-fold-GA-deep method the minimum error of the comprehensive evaluation indexes

of the K-fold-GA-deep method indicates that the model constructed by this method is suitable for this city. The model constructed by this method is suitable for predicting the maintenance and operation cost of this city, and the effect is good. The results are good.

6. Conclusion

For the problem of allocated cost of maintenance of constructed UHV substation equipment, firstly, the historical data are processed, and the true value of the historical maintenance cost data is imputed according to the analysis and collation of UHV substation maintenance financial entries; to avoid the traditional deep neural network will fall into local optimum, then the training data are K-fold processed, and the deep improved by genetic algorithm is used neural network to predict UHV substation maintenance costs. Conclusions are drawn from the study.

(1) The process of collecting historical data of UHV substation operation and maintenance overhaul costs is difficult, and the relevant overhaul costs are obtained by analyzing and organizing the financial entries of UHV substation engineering overhaul, and processing the data in horizontal and vertical directions to ensure the real reliability of the obtained data. And analyze the various constituent factors affecting UHV substation overhaul operation and maintenance costs and use them as prediction input of the model.

(2) When training the deep neural network on the original data, K-fold cross-validation can better ensure the accuracy of the prediction model and avoid the phenomenon of overfitting.

(3) The K-fold-GA-deep neural network model is used to predict the future maintenance allocation costs of constructed UHV substations. Through the training and prediction of typical municipal UHV substation maintenance cost data, it is known that the model can accurately predict the future maintenance costs. In the case of large fluctuation of maintenance cost every year, this method can be used to get the maintenance allocation value more accurately, which can provide reference for the grid to allocate maintenance cost to UHV substations.

Acknowledgement

This work is supported by Science and Technology Program of State Grid Corporation of China(5500-202224139A-1-1-ZN).

References

[1] Gu Jie, Du Jian, Qin Jie, et. Al. A fuzzy programming model based on life cycle cost for equipment maintenance in substations [J]. Automation of Electric Power Systems, 2014, 38(1): 44-50.

[2] He Xu, Jiang Xianguo, Zhang Peichao, et. Al. Availability analysis of smart substation protection system considering maintenance strategies [J]. Power System Technolgoy, 2015, 39(4): 1121-1128.

[3] Guo Liujuan, Tao Songmei, Zhang Wei, et. Al. Maintenance strategy optimization for main equipment in substations considering both power supply reliabilibty and life cycle cost [J]. Automation of Electric Power Systems, 2016, 40(15): 99-105.

[4] Liu Su, Yao Jiangang, Wei Yilong, et. Al. Life cycle cost study of main transformer based on Monte Carlo [J]. Chinese Journal of Power Sources, 2017, 41(2): 312-315.

[5] Cai Jia, Jiang Xiubo. Bayesian econometrics method based maintenance cost model for 220kV transformers [J]. Journal of Electric Power Science and Technology, 2013, 28(4): 15-18.

[6] Qian Hai, Jia Songjiang, Yang Fei, et. Al. Intelligent operation and maintenance technology for relay protection equipment based on mobile internet [J]. Smart Power, 2019, 47(11): 60-66.

[7] Han Jun, Xie Zhenjian, Huang He, et. Al. Unit mesh reliability planning for distribution network based on outage cost calculation and cost analyssi [J]. Smart Power, 2020, 48(1): 63-68.

[8] Fu Chengran, He Yongxiu. Analysis of influencing factors of physical assets benefit of substation equipment based on interpretative structural model [J]. Shandong Electric Power, 2019, 46(5):35-40.

[9] Li Zhengming, Liang Caixia, Wang Manshang. Short-term power generation output prediction based on a PSO-DBN neural network [J]. Power System Protection and Control, 2020, 48(8): 149-154.