Research on Substation Project Prediction Method in Power Transmission and Transformation by Improved Neural Network Intelligent Model

Li Ma*, Lanfei He, Xuefei Zhang, Wei Wang, Liping Sun, Li Zhou, Qiupeng Zhou, Fangchao Ke
State Grid Hubei Electric Power Company Limited Economic Research Institute, Wuhan, China

*Corresponding author: mal12@hb.sgcc.com.cn

Abstract. On the basis of collecting a large number of practical engineering cases, this paper uses the improved neural network prediction model based on multiple linear regression to build an improved substation project cost prediction model based on the division of substation projects. Using the actual project sample data for empirical research, the results show that the prediction model has high accuracy for the prediction of substation engineering. This research work provides a reliable technical scheme for the cost prediction of power transmission project.

Keywords: substation engineering, neural network, multiple linear regression, cost prediction

1. Introduction
After the reform of transmission and distribution price, the profit model of power grid will change from earning price difference to "permitted cost + reasonable income", which will effectively release the dividend of power reform. The power transmission and distribution price is approved separately, and the incremental part of distribution side and sales side is released, which has a profound impact on the profit-making mode of power grid. At present, when the government verifies the transmission and distribution price of the provincial power company, the annual capital transfer rate is determined by a certain percentage according to the actual situation of the company's asset formation over the years. From the investment analysis, affected by the external environment and design depth of the project construction, the investment balance rate of some power grid infrastructure projects is on the high side (the deviation between the final investment and the planned investment is large), which affects the investment plan of the power grid company. The accuracy and rationality of planning and the actual transfer rate have a certain negative impact on the formation of effective assets and the verification of transmission and distribution price of power grid companies.

2. Research status of power transmission and transformation project cost prediction
Traditional prediction methods mainly include qualitative and quantitative prediction methods. Traditional prediction methods mainly include expert meeting method and subjective probability
prediction [1-3]. Quantitative prediction methods mainly include grey prediction method, fuzzy mathematical prediction method, moving average method and exponential smoothing method [4-6]. Due to the large number of engineering variables and the small number of samples, it is difficult to reasonably predict the project cost by using traditional methods [7-8]. This paper will analyze the latest collection of project historical cost data, improve data cleaning and conversion methods, improve the utilization of historical data information. In depth research on AI technology based on machine learning and deep learning, research on the latest intelligent algorithms such as regression prediction, on the basis of further improving the accuracy of data, combined with more historical project data, research and analyze the relationship between various factors in power grid infrastructure projects, train and optimize the power grid investment cost prediction model, and further improve the accuracy and application of the prediction model Availability of scenarios. At the same time, in order to improve the accuracy of data collection, we also developed the data automatic capture software to automatically extract the key data in the original data table, so as to further improve the accuracy of the research.

3. Research on cost prediction method of power transmission and transformation project

3.1. Construction of cost analysis database
At present, the project cost analysis work is to collect data offline, the sample data quality is not high, the error rate is large, the effective sample is small, the analysis work is manual sorting, analysis, calculation, which takes a long time, the analysis conclusion is single, and the utilization value of the project data resources is not high. The construction of cost analysis database can solve the lag of data collection afterwards, avoid the disadvantages of long project implementation cycle and lack of data, realize the collection and input of data in the process of project implementation, and dynamically optimize according to the evaluation opinions in the process of project implementation in real time.

At the same time, the collected data are imported into the database and found that there are some problems, such as different table standards, difficult to identify; different types of projects in different stages, different cost classification categories; different formats of sample tables exported by different software. By standardizing the data collection template and adopting standardized format sample table, the construction process of cost analysis database can be optimized. If we try to use the cost name in the project type division in the power grid construction budget preparation and calculation regulations, and try to keep the project name in the power industry quota or list specification unchanged, we can effectively improve the efficiency of cost analysis.

3.2. Extraction of main factors
After collecting a large number of key data of historical transmission and transformation projects, BP neural network can be used to build artificial intelligence project cost prediction model, which provides theoretical reference for the current project investment plan. The main process is shown in the following figure:
3.2.1. **Construction of substation project cost model based on itemized cost prediction**

(1) Construction method

The prediction model of power transmission and transformation project is generally used to predict the static investment of the project. This modeling method is relatively simple and convenient for parameter adjustment, but it ignores the difference of the impact of different variables on the project investment. For the substation project with less project samples and more influence variables, the separate cost forecast can effectively improve the accuracy of the forecast. In this prediction model, the substation project is divided into construction cost, installation cost, equipment cost and other costs, which are respectively predicted and finally summarized to obtain the static investment of the project. The structure of the optimization model for substation project cost prediction and cost prediction is shown in the following structure diagram.
In this paper, multiple linear regression model is used to screen the main factors one by one. First, the evaluation is based on the adjusted R-square and kmo values. By constantly deleting the factor with the largest "significance" value, we can improve the "adjusted R square" and "kmo" values until the two values no longer change significantly. Second: find out the missing key factors, add the factors that are not added one by one, and keep the factors that increase or do not affect the value of "adjusted R" and "kmo". Because the more input attributes of neural network, the more accurate the trained model, so try to retain more factors with good correlation.

2) Factor screening
This model will take the factor screening of construction cost as an example to elaborate the screening process based on the itemized cost prediction model.
Firstly, the variables selected by experts' experience will be input into the model, and the linear regression relationship between each variable and construction cost will be initially established. Then the input variables are simplified according to the analysis results screened out by the software, and the input value of the dependent variable with higher importance is finally determined.

According to the experience of experts, the preliminary screening factors are: Survey and design cost (yuan), building area of the whole station (M2), unit price of main transformer equipment, land acquisition area (MU), site leveling cost, control cable quantity, retaining wall and slope protection cost, number of outgoing lines at medium voltage side, building area of main control building (M2), power cable quantity, number of circuit breakers at high voltage side, number of main transformer in current period, The number of outgoing lines at high voltage side, the cost of foundation treatment, the number of circuit breakers at medium voltage side, the number of circuit breakers at low voltage side, and the number of outgoing lines at low voltage side.

Table 1. Preliminary screening results of factor analysis of construction cost

| R-squared after adjustment | KMO     | Bartlett sphericity test   |
|---------------------------|---------|---------------------------|
|                           |         | The approximate chi-square | Freedom | Significance |
| 0.766                     | 0.532   | 398.239                   | 55      | 0.000        |

According to the preliminary screening results of the model, the R-square result adjusted by factor analysis of construction cost is 0.766, and the model fitting result is general. The kmo value of 0.532 indicates that the linear correlation of variables is general. The result of Bartley sphericity test is less
than 0.05, which means that the data samples are spherical distribution, excluding the linear correlation of variables. The preliminary screening results show that the input variables can be further screened.

Table 2. Preliminary screening results of significance coefficient of factor analysis of construction cost

| Factor                                      | Standardization coefficient | t     | Significance |
|---------------------------------------------|-----------------------------|-------|--------------|
| Number of main transformer in current period | -0.556                      | -2.193| 0.047        |
| Number of outgoing lines at high voltage side| 0.058                       | 0.510 | 0.618        |
| Number of outgoing lines at medium voltage side | -0.154                   | -0.727| 0.480        |
| Number of outgoing lines at low voltage side | 0.528                      | 1.262 | 0.229        |
| Land acquisition area (MU)                  | 0.063                       | 0.511 | 0.618        |
| Total station building area (M2)            | 0.340                       | 3.561 | 0.003        |
| Building area of main control building (M2) | 0.518                       | 4.643 | 0.000        |
| Number of high voltage side circuit breakers | 0.156                      | 1.365 | 0.195        |
| Number of medium voltage side circuit breakers | 0.522                    | 2.734 | 0.017        |
| Number of low voltage side circuit breakers | 0.127                      | 0.512 | 0.617        |
| Foundation treatment cost                   | 0.394                       | 3.634 | 0.003        |
| Site formation cost                         | 0.048                       | 0.566 | 0.581        |
| Cost of retaining wall and slope protection | 0.165                      | 1.866 | 0.085        |
| Quantity of power cable                     | 0.002                       | 0.019 | 0.985        |
| Control cable quantities                    | -0.149                      | -1.590| 0.136        |

Finally, the main factors screened out are: the number of main transformers in the current period, the building area of the whole station, the building area of the main control building, the quantity of control cables, the number of medium voltage side circuit breakers, the cost of site leveling, the number of high voltage side circuit breakers, the cost of foundation treatment, and the number of low voltage side circuit breakers.

Table 3. Final screening results of factor analysis of construction cost

| R-squared after adjustment | KMO  | Bartlett sphericity test |
|----------------------------|------|--------------------------|
|                            |      | The approximate chi-square | Freedom | Significance |
| 0.915                      | 0.752| 198.719                  | 55      | 0.000        |

After screening the input variables of the model, the adjusted R square of the final screening results is 0.915, and the model fitting results are very good. Kmo value and Bartlett sphericity test results also show that there is no strong linear correlation between the model variables, which is suitable for factor analysis. The variables after final screening can be used to predict the construction cost.

Table 4. Final screening results of significance coefficient of factor analysis of construction cost

| Factor                                      | Standardization coefficient | t     | Significance |
|---------------------------------------------|-----------------------------|-------|--------------|
| Number of main transformers in current period | -0.338                      | -2.657| 0.014        |
| Total station building area (M2)            | 0.358                       | 5.553 | 0.000        |
| Site formation cost                         | 0.114                       | 1.887 | 0.072        |
| Number of medium voltage side circuit breakers | 0.433                      | 2.792 | 0.011        |
| Number of high voltage side circuit breakers | 0.181                      | 2.285 | 0.032        |
| Building area of main control building (M2) | 0.448                       | 5.340 | 0.000        |
| Foundation treatment cost                   | 0.354                       | 4.424 | 0.000        |
| Number of low voltage side circuit breakers | 0.379                       | 2.554 | 0.018        |
| Control cable quantities                    | -0.106                      | -1.397| 0.176        |
Figure 4. Normal P-P diagram of regression standardized residuals

(3) Model training

According to the screening process of the above itemized cost prediction model, the key influencing factors of construction cost, installation cost, equipment cost and other costs are screened. Then select the training sample set and verification sample set of 110kV new substation data. A neural network with one input layer, one output layer and nine hidden layers is built. The number of neurons in each hidden layer is 16, 8, 4, 8, 4, 8, 8, 8 and 4 respectively. The number of training rounds was 10000 and the learning rate was 0.01.

By using the optimized neural network intelligent prediction model, the investment prediction model of power transformation project is divided into the prediction training of construction cost, installation cost, equipment cost and other costs from the static cost prediction.

3.3. Model validation results

The trained model is used to verify the data of 8 verification sets, and the final static investment results are as follows after summarizing the predicted installation cost, equipment purchase cost, construction cost and other costs. It can be seen that the average deviation rate of 8 projects is 2.07%, and the static investment deviation rate of all projects is predicted to be less than 5%, so the prediction result is good.

Table 5. Summary of static investment forecast of Substation

| Order | Name of individual project | Voltage level | Approved final investment (10000 yuan) | Static investment forecast (10000 yuan) | Deviation rate (%) |
|-------|-----------------------------|---------------|---------------------------------------|----------------------------------------|-------------------|
|       |                             |               | Construction cost | Equipment cost | Installation cost | Other expenses | total | Construction cost | Equipment cost | Installation cost | Other expenses | total |               |
| 1     | A substation project        | 110           | 1803.72           | 1726.63        | 614.89           | 821.48         | 4966.72 | 1929.54           | 1732.43        | 550.38           | 785.67         | 4998.02 | 0.63%        |
| 2     | B substation project        | 110           | 1021.99           | 1195.5         | 402.44           | 702.80         | 3322.73 | 1010.1            | 1249.59        | 454.57           | 710.69         | 3424.95 | 3.08%        |
| 3     | C substation project        | 110           | 948.41            | 1748.11        | 507.91           | 734.86         | 3939.29 | 883.72            | 1877.29        | 485.38           | 746.90         | 3993.29 | 1.37%        |
| 4     | D substation project        | 110           | 906.44            | 1071.07        | 455.03           | 543.07         | 2975.61 | 840               | 1061.15        | 451.46           | 581.25         | 2933.86 | 1.40%        |
| 5     | E substation project        | 110           | 1306.03           | 1709.66        | 555.37           | 708.76         | 4729.82 | 1283.6            | 1865.47        | 539.03           | 703.72         | 4391.82 | 2.62%        |
| 6     | F substation project        | 110           | 838.67            | 1072.63        | 434.96           | 506.73         | 2852.99 | 822.89            | 1188.93        | 402.65           | 529.72         | 2944.19 | 3.20%        |
| 7     | G substation project        | 110           | 893.96            | 958.74         | 385.36           | 668.46         | 2906.52 | 816.6             | 947.9          | 389.09           | 670.89         | 2824.48 | 2.82%        |
| 8     | H substation project        | 110           | 801.25            | 1092.33        | 370.31           | 602.37         | 2866.26 | 776.37            | 1167.5         | 391.75           | 572.41         | 2908.03 | 1.46%        |
4. Conclusion
In this paper, the linear regression model and factor analysis theory are used to screen the factors of the itemized cost prediction model, and the factors with small influence factors and insignificant weight coefficient are screened out. For the new substation, the key factors affecting the construction cost, installation cost, equipment purchase cost and other costs are constructed.

By using the optimized neural network intelligent prediction model, the investment prediction model of power transformation project is divided into the prediction model of construction engineering cost, installation engineering cost, equipment cost and other costs. Finally, the static investment prediction model of substation project is established to meet the demand of cost prediction dimension refinement.

Through the verification of the optimized neural network intelligent prediction model, the prediction model has certain prediction ability for each item cost of new substation, and the prediction result is more accurate. It can assist the investment forecast in the preparation of the front-end investment plan of the project.

Acknowledgments
This work was financially supported by Science and Technology Project of State Grid Hubei Electric Power Company Limited Economic Research Institute (521538200007)

References
[1] Teng Lingyun. Research on construction cost prediction model based on BP neural network [J]. Housing industry, 2020 (12): 110-113
[2] Zhang Yuchen, Zhang Yulin, Feng Chunfei, Wang Chen. Research on transmission line cost prediction model based on BP neural network model [J]. Power big data, 2020, 23 (06): 35-42
[3] Li Xiao. Prediction of EPC cost index based on pso-aco-bp neural network [D]. Xihua University, 2020
[4] Wang Yao. Prediction of housing project budget based on BP neural network [D]. Lanzhou Jiaotong University, 2020
[5] Miao Xujuan, Liu Jinning, Gao Liang, Li Changling, Zhang Xinyan, Xu Qidan. Prediction model of technological transformation project cost based on principal component analysis and neural network [J]. Science and technology and economy of Inner Mongolia, 2019 (19): 37-40 + 42
[6] Yang Kai, Yu Bo, Xiao Yanli, He Yongping, Wang Fengxiao. Distribution network project cost prediction based on GA-BP neural network [J]. Automation instrument, 2019, 40 (07): 91-93 + 99
[7] Lu Yan, Lv Ke, Xie Jinghai, sun MI, Zhao Jia. Application of particle swarm optimization BP neural network algorithm in cost prediction of transmission line project [a]. Proceedings of 2018 Symposium on development and application of new smart grid technologies [C]. China Electric Power Research Institute: Editorial Department of computer engineering and application, 2018:5
[8] Mo Junwen, Wu Hailong. Research on housing construction cost prediction based on RBF artificial neural network [J]. Project cost management, 2018 (06): 15-23