Analysis of the Influence Factors of Transformer Substation Project Cost Based on the Random Forest Model

Ming ZHOU¹, Sheng-wei LU¹, Si-cong WANG², Zi-xia SANG², Jia-qi HUANG² and Shu WANG²

¹State Grid HuBei Electric Power Company Limited Economic
²State Grid HuBei Electric Power Company Limited Economic Research Institute

Keywords: Transformer substation, Random forest, Project cost.

Abstract. There are many factors affecting the cost of transformer substation project. The project cost difference is large and the data is discrete. It brings great difficulties to the classification analysis, index preparation and control management of project cost. Based on the project division of transformer substation project, this paper analyzes and identifies the main factors affecting the construction cost of transformer substation project, and uses the Random Forest model to measure 25 of the influencing factors, selects the key influencing factors to provide reference for the cost prediction of transformer substation project.

Introduction

In the construction of power transmission and transformation projects, the impact of project cost management and control on the entire project management is very important. It is extremely significant to analyze the level of production cost, structural changes, influencing factors and development trends to achieve effective control of cost, rationally verify transmission costs, improve grid investment efficiency and strengthen project investment management. However, there are many influencing factors of the cost of power transmission and transformation projects, including the complicated construction environment and uneven construction level of the line engineering, the cost of power transmission and transformation projects under the same voltage level can be greatly different. As a result, for a project, the analogous power transmission and transformation project can be referenced has been limited in a small sample size, and the relevant auditors of the cost engineering are difficult to review and estimate the project cost based on experience and conventional statistical estimation models. Therefore, for the actual situation with many influencing factors, the method of attribute screening in data mining to find out the relevant main cost estimation technical indicators of power transmission and transformation engineering has strong reference for the cost management of power transmission and transformation projects.

Research Status of Influencing Factors of Transformer Substation Project Cost

Electric power project cost management is affected by many factors, and reasonable screening of important influencing factors is great significance to engineering investment control [1-3]. Scholars at home and abroad have analyzed and studied the factors influencing the cost of electric power engineering, and have accumulated rich experience.

Liu et al. [4] used the analytic hierarchy process to analyze the key indicators of transformer substation project cost; Hu et al. [5] comprehensively used principal component analysis and multiple regression models to systematically study the influencing factors of investment in electric power project; principal component analysis and factor analysis were also used by many scholars in the analysis of electric power project cost influencing factors [6-7]. An et al. [8] combined random forest algorithms in data mining to screen important factors affecting power grid project, and then established an intelligent budget estimation model suitable for small sample conditions. Finally, the effectiveness of this model was confirmed by a case study in a certain area.
Research Method on Influencing Factors of Transformer Substation Project Cost

Combined with the existing literature on the factors affecting the cost of power transmission and transformation projects, domestic scholars have done a lot of research on related fields with the combination of qualitative and quantitative methods, including analytic hierarchy process, principal component analysis, factor analysis, multiple regression models, support vector machines and random forest models. This report takes the transformer substation project as the research object, uses the random forest model to screen the influencing factors affecting the cost of the transformer substation project, compare the results predicted by the main influencing factors with the actual results.

The Random Forest (RF) algorithm is an algorithm which belongs to machine learning in data mining. It is proposed by Breiman based on the CART classification algorithm. This algorithm draws on the idea of random decision forest, when machine learning generates classification tree, it uses random combination of row variable and column variable data to generate many random classification trees, and then integrates them into random forests. The random forest improves the prediction accuracy under the premise that the computational complexity is not significantly improved. It is an algorithm designed for the classification and estimation of high-dimensional small sample data, can explain the effect of attributes on dependent variables in thousands of attributes. It is widely used in the estimation of high-dimensional small samples. The construction process of random forests is mainly as follows:

Figure 1. The construction process of Random Forest model.

The idea of using a random forest for feature importance assessment is to derive how much contributes that each feature to each tree in the random forest, then calculate the average and finally compare the contribution between the features. The contribution size is usually measured using the Gini index or the out of bag data (OOB) error rate as an indicator.

On the basis of the importance of features, the steps of feature selection are as follows:

1. Calculate the importance of each feature and sort it in descending order
2. Determine the proportion to be eliminated, and remove the corresponding proportion of features according to the importance of the feature to obtain a new feature set.
3. Repeat the above process with a new feature set until m features are left (m is the value set in advance)
4. According to the out-of-bag error rate corresponding to each feature set obtained in the above code, the feature set with the lowest out-of-bag error rate would be selected.

Analysis of Factors that Affecting the Cost of Transformer Substation Projects

Theoretical Analysis of Factors that Affecting the Cost of Transformer Substation Projects

Combined with the “Power System Construction Budgeting and Calculation Regulations (2013 Edition)” issued by the National Energy Administration of China, the division of transformer
substation engineering projects; through reading relevant literature and combining the results of actual case studies, it is found that the main factors affecting the cost of transformer substation projects are mainly including the following points: ground endurance, pollution level, altitude, topography, construction nature, voltage level, transformer substation type, land acquisition area, building area, main transformer, current outlet scale, power distribution unit, secondary equipment, etc.

Take the transformer substation type and power distribution unit as an example.

(1) Transformer substation type. The transformer substation type is mainly divided into outdoor, semi-outdoor and indoor. The indoor transformer substation has a small footprint and high cost; the outdoor transformer substation has a large area and low cost; and the semi-outdoor combines the advantages of both. Different transformer substation types have a greater impact on construction engineering fees.

(2) Power distribution unit. When the transformer substation power distribution unit is different in type, the equipment price level will also change. Take W high-voltage switchgear as an example: air-insulated switchgear (AIS), gas-insulated switchgear (GIS) and multiple technology switchgear (MTS), each of which has its own characteristics, and the unit price is gradually increasing.

Therefore, there are many factors affecting the cost of transformer substation projects, but these factors are related. The main influencing factors of the cost of transformer substation projects are of great significance for predicting the cost of transformer substation projects.

Empirical Analysis of Factors that Affecting the Cost of Transformer Substation Projects

Data Collection and Processing. The original data of this report comes from 53 110kV new or expanded transformer substation projects completed by a research institute in 2016-2018. Data of 26 indicators were collected, including static investment in transformer substation engineering(y), construction nature(x1), total capacity of transformer substation(x2), high\medium\low voltage side outlets(x3\x4\x5), low voltage capacitors(x6), land acquisition area(x7), area of main control building(x8), construction site requisition and cleaning fees(x9), main transformer unit price(x10), high\medium voltage side power distribution unit type(x11\x14), high\medium\low voltage side number of circuit breakers(x12\x15\x17), high\medium\low voltage side unit price of circuit breaker(x13\x16\x18), cost of inbound road(x19), foundation treatment method(x20), foundation treatment cost(x21), field leveling cost(x22), retaining wall and slope protection(x23), power cable(24), control cable(25), etc. Among them, static investment is a dependent variable, and others are independent variables.

Because in the process of identifying the influencing factors, the construction nature, the high-voltage side power distribution device type, the medium-voltage side power distribution device type and the foundation treatment method are descriptive factors, so the quantitative processing is required before the analysis, and the processing method is as follows:

The construction nature of the transformer substation project in the data sample is divided into new construction and expansion project. When the analysis is carried out, the expansion project is defined as 1 and the new construction project is defined as 2.

The foundation treatment methods are mainly divided into five kinds of attributes: non-treatment method, cast-in-place pile method, masonry concrete replacement method, mixing pile method and dynamic compaction method. These five attributes are represented by 0, 1, 2, 3, and 4 respectively.

The specific data is shown in Table 1:
Table 1. Sample data of transformer substation project (1).

| Project Number | Y (10^4 Yuan) | x1 (MVA) | x2 (MVar) | x3 (mu) | x4 (m²) |
|----------------|---------------|----------|-----------|---------|---------|
| Project 1      | 718.16        | 1        | 40        | 0       | 6       | 6.00    | 0.00    | 0      |
| Project 2      | 826.01        | 1        | 50        | 0       | 6       | 8.00    | 0.00    | 0      |
| Project 3      | 2710.12       | 2        | 50        | 2       | 4       | 8.00    | 10.12   | 486    |
| Project 4      | 633.80        | 1        | 50        | 0       | 5       | 3       | 8.00    | 0.00    | 0      |
| Project 5      | 629.77        | 2        | 50        | 3       | 0       | 8       | 8.40    | 8.33    | 390    |
| Project 6      | 3127.07       | 2        | 50        | 2       | 4       | 8       | 8.40    | 8.60    | 390    |
| Project 7      | 2756.41       | 2        | 50        | 3       | 0       | 8       | 8.40    | 8.60    | 390    |
| Project 8      | 1650.29       | 1        | 100       | 0       | 0       | 9       | 9.60    | 0.00    | 0      |
| Project 9      | 2498.13       | 2        | 50        | 1       | 0       | 8       | 8.00    | 8.99    | 390    |

Table 1. Sample data of transformer substation project (2).

| Project Number | x9 (10^4 Yuan) | x10 (10^4 Yuan/eac h) | x11 | x12 | x13 (10^4 Yuan/each) | x14 | x15 | x16 (10^4 Yuan/eac h) |
|----------------|----------------|-----------------------|-----|-----|---------------------|-----|-----|-----------------------|
| Project 1      | 2.65           | 161.13                | 2   | 1   | 33.50               | 0   | 0   | 0.00                  |
| Project 2      | 0.00           | 220.53                | 1   | 2   | 9.40                | 1   | 9   | 6.05                  |
| Project 3      | 225.75         | 298.90                | 1   | 3   | 9.25                | 0   | 0   | 0.00                  |
| Project 4      | 0.00           | 208.90                | 1   | 2   | 9.25                | 1   | 1   | 6.05                  |
| Project 5      | 0.00           | 228.70                | 1   | 2   | 9.25                | 0   | 0   | 0.00                  |
| Project 6      | 335.65         | 267.31                | 1   | 5   | 14.80               | 0   | 0   | 0.00                  |
| Project 7      | 224.50         | 262.51                | 1   | 3   | 15.00               | 0   | 0   | 0.00                  |
| Project 8      | 25.20          | 289.76                | 1   | 4   | 14.83               | 0   | 0   | 0.00                  |
| Project 9      | 258.17         | 0.00                  | 2   | 2   | 14.70               | 0   | 0   | 0.00                  |

Table 1. Sample data of transformer substation project (3).

| Project Number | x17 (10^4 Yuan/eac h) | x18 (10^4 Yuan) | x19 (10^4 Yuan) | x20 | x21 (10^4 Yuan) | x21 (10^4 Yuan) | x23 (10^4 Yuan) | x24 (m) | x25 (m) |
|----------------|-----------------------|----------------|----------------|-----|----------------|----------------|----------------|---------|---------|
| Project 1      | 6                     | 7.13           | 0.00           | 0   | 0.00           | 0.00           | 0.00           | 600     | 7140    |
| Project 2      | 6                     | 22.99          | 0.00           | 1   | 36.30          | 0.00           | 0.00           | 2264    | 8648    |
| Project 3      | 9                     | 4.79           | 8.25           | 1   | 90.42          | 116.43         | 76.88          | 3570    | 8140    |
| Project 4      | 9                     | 8.20           | 0.00           | 2   | 8.19           | 0.00           | 0.00           | 1202    | 4100    |
| Project 5      | 8                     | 4.55           | 0.00           | 2   | 8.07           | 0.00           | 0.00           | 1270    | 6568    |
| Project 6      | 17                    | 5.71           | 10.10          | 1   | 114.55         | 31.53          | 5.32           | 4740    | 13870   |
| Project 7      | 17                    | 5.71           | 7.54           | 1   | 56.80          | 98.19          | 78.86          | 5051    | 10800   |
| Project 8      | 21                    | 6.25           | 0.23           | 0   | 0.00           | 0.21           | 0.00           | 5360    | 17147   |
| Project 9      | 15                    | 5.18           | 8.24           | 2   | 102.78         | 47.57          | 0.00           | 3280    | 14620   |

...
**Random Forest Factor Screening.** After using the random forest model above to filter the above attributes, the static investment is selected as the decision variable, and the rest is used as the attribute variable to form the preliminary data set. Using the RandomForest package in the R language for analysis, Gini index of each attribute would be calculated, as shown in Figure 2.

In Figure 2, IncNodePurity represents the purity of the node, which has the same meaning as the Gini index. The larger the value, the higher the importance of the attribute. Among the attribute factors affecting transformer substation projects, there are 18 indicators with an IncNodePurity index exceeding 106, including construction nature, total capacity of transformer substation, high-voltage side outlets, low-voltage side outlets, low-voltage capacitors, land acquisition area, main control building area, construction site requisition and cleaning fee, main transformer unit price, high voltage side circuit breaker unit, high voltage side circuit breaker unit price, low voltage side circuit breakers number, the cost of the inbound road, the foundation treatment cost, the foundation treatment method, the field leveling cost, the retaining wall and slope protection, the power cable.

**Model Verification.** Using the 18 important indicators filtered above, the static investment of 23 forecasting projects is selected as the decision variable, and the random forest model is used for forecasting. The static investment forecast and error rate are shown in Table 2 and Figure 3.
Table 2. Static investment forecast value and forecast error rate of substation project.

| Project Number | Construction Nature | Total Capacity (MVA) | Land Acquisition Area (mu) | Main Control Building (m²) | Requisition and Cleaning Fees (10^4 Yuan) | Actual Static Investment Value | Forecast Static Investment Value | Error Rate |
|----------------|---------------------|----------------------|-----------------------------|----------------------------|------------------------------------------|-------------------------------|-------------------------------|------------|
| Project 1      | 2                   | 50                   | 9.68                        | 617                        | 191.57                                  | 2821.10                       | 2868.30                      | 1.67%      |
| Project 2      | 2                   | 50                   | 9.64                        | 550                        | 156.89                                  | 2610.99                       | 2520.17                      | -3.48%     |
| Project 3      | 2                   | 50                   | 8.20                        | 390                        | 194.64                                  | 2730.78                       | 2700.19                      | -1.12%     |
| Project 4      | 2                   | 50                   | 13.16                       | 384                        | 110.86                                  | 2522.52                       | 2578.34                      | 2.21%      |
| Project 5      | 1                   | 50                   | 0.00                        | 0                          | 3.42                                    | 827.83                        | 921.87                       | 11.36%     |
| Project 6      | 2                   | 50                   | 7.55                        | 486                        | 189.74                                  | 2367.36                       | 2358.74                      | -0.36%     |
| Project 7      | 1                   | 50                   | 0.00                        | 31                         | 0.00                                    | 876.43                        | 890.57                       | 1.61%      |
| Project 8      | 2                   | 31.5                 | 5.73                        | 347                        | 161.10                                  | 2207.52                       | 2440.01                      | 10.53%     |
| Project 9      | 1                   | 20                   | 1.24                        | 0                          | 19.57                                   | 1047.72                       | 995.89                       | -4.95%     |
| Project 10     | 2                   | 50                   | 11.26                       | 409                        | 457.50                                  | 2673.57                       | 2634.03                      | -1.48%     |
| Project 11     | 2                   | 50                   | 9.66                        | 467                        | 330.70                                  | 2651.00                       | 2606.80                      | -1.67%     |
| Project 12     | 1                   | 50                   | 0.00                        | 367                        | 20.28                                   | 1030.57                       | 978.80                       | -5.02%     |
| Project 13     | 2                   | 50                   | 0.00                        | 0                          | 0.00                                    | 2852.81                       | 2505.59                      | -12.17%    |
| Project 14     | 2                   | 50                   | 8.42                        | 556                        | 144.07                                  | 2983.50                       | 2784.36                      | -6.67%     |
| Project 15     | 2                   | 50                   | 9.06                        | 303                        | 137.36                                  | 2508.52                       | 2558.84                      | 2.01%      |
| Project 16     | 1                   | 50                   | 0.00                        | 0                          | 0.00                                    | 916.00                        | 904.14                       | -1.29%     |
| Project 17     | 2                   | 100                  | 14.17                       | 620                        | 266.40                                  | 3336.11                       | 3316.15                      | -0.60%     |
| Project 18     | 2                   | 50                   | 8.25                        | 517                        | 148.50                                  | 2279.52                       | 2368.02                      | 3.88%      |
| Project 19     | 2                   | 100                  | 6.79                        | 2445                       | 65.89                                   | 3719.50                       | 3463.59                      | -6.88%     |
| Project 20     | 2                   | 150                  | 12.43                       | 620                        | 215.33                                  | 2544.06                       | 2677.81                      | 5.26%      |
| Project 21     | 2                   | 100                  | 12.81                       | 423                        | 158.80                                  | 3152.05                       | 3132.29                      | -0.63%     |
| Project 22     | 2                   | 50                   | 7.60                        | 487                        | 85.27                                   | 2378.21                       | 2437.13                      | 2.48%      |
| Project 23     | 2                   | 100                  | 6.87                        | 1005                       | 188.69                                  | 3759.05                       | 3458.83                      | -7.99%     |

Figure 3. Comparison of static investment forecast value and actual value of transformer substation project.

It can be seen from Figure 3 that the static investment forecast value curve obtained by using 18 main indicators is very close to the static investment actual value curve corresponding to all indicators, indicating that the 18 major indicators can be used to accurately predict the static investment value of transformer substation project. Figure 4 is a summary of the interval division of the prediction error rate. It can be seen that the error of all test data is within 15%, and the test data with the prediction error within 5% reaches 65%, which further proves that the above 18 main indicator variables are representative of the information contained in all indicators.
Conclusions and Prospects

This report studies the selection of affecting factors for transformer substation projects in the case of high-dimensional small samples by random forest model. For the multi-factor and high-dimensional situation, the random forest method is used to screen out the main factors affecting the transformer substation project, after that, the indicator factors need to be reduced in dimension. Then the 18 main factors selected as the input variables of the model are used to predict and verify the static investment of the transformer substation project. The obtained prediction results are very close to the actual values, which can provide an effective reference for transformer substation project cost control. And as to enrich and improve the transformer substation project data, the analysis of the affecting factors of the cost of the transformer substation project and the accuracy of the cost prediction model will be further improved.

Reference

[1] Mashayekhi, A N. Project cost cost dynamics for development policy-making[J]. Journal of the Operational Research Society, 51(3):301-310.

[2] Kim, Byung-Cheol. Integrating Risk Assessment and Actual Performance for Probabilistic Project Cost Forecasting: A Second Moment Bayesian Model[J]. IEEE Transactions on Engineering Management, 62(2):158-170.

[3] A. Robertson. Project Cost Control Using Networks[J]. Journal of the Operational Research Society, 21(1):143.

[4] Liu Shili, Zhu Xiaohu, Chen Fulei, etc. Critical Factors for Cost Management of Power Grid Projects Based on Analytic Hierarchy Process [J]. Journal of Engineering Management, 2016, 30 (6): 129-133.

[5] Hu Jinlan, Gao Xiaobin. Identification and Data Modeling of Transmission and Transformation Projects’ Factors [J]. Journal of North China Electric Power University (Social Sciences), 2016 (5): 55-59.

[6] Xu Li, Li Zhuoran. Research on Engineering Cost Forecasting Model of UHV Power Grid Based on Factor Analysis and BP Neural Network [J]. Journal of Industrial Technological Economics, 2017, 36 (7): 18-26.

[7] Xiao Junyi, Hao Xiaoling, Tang Xiaoqian, etc. Influencing Factors of Electric Power Project Cost Based On Principal Component Analysis [J]. Management and Administration, 2014 (3): 128-133.

[8] An Lei, Zhang Jie, Qi Xia, et al. Transmission Line Project Cost Forecasting Based on Random Forest [J]. Control Engineering of China, 2016 (11): 1841-1844.